Volume 12 Number 18 1984 Nucleic Acids Research

Primary and secondary structure of black beetle virus RNA2, the genomic messenger for BBV coat protein precursor

Ranjit Dasgupta, Amit Ghosh*, Bimalendu Dasmahapatra, Linda A. Guarino+ and Paul Kaesberg§

Biophysics Laboratory and Biochemistry Department, University of Wisconsin-Madison, Madison, WI 53706, USA

Received 20 July 1984; Accepted 20 August 1984

ABSTRACT

The nucleotide sequence of black beetle virion (BBV) RNA2 has been determined. RNA2 is 1399 b long. Its 5' terminus is capped. Its 3' terminus has an unidentified moiety that renders the RNA resistent to polyadenylation and ligation. The first AUG codon at base 23 is followed by an open reading frame for a protein 407 amino acids long, the predicted size of coat protein precursor. A second open reading frame for a putative protein 72 amino acid residues long begins at base 1110. No other large open reading frames exist. The 5' half of the RNA can be folded into a long, imperfect hairpin of high predicted stability. The 3' half of the RNA can fold into a complex set of multiply bifurcated stem and loop regions.

INTRODUCTION

BBV (black beetle virus) is a member of the family Nodaviridae (1). Its genome consists of two messenger-sense RNAs (2). Virion RNA1 (22S) serves as the genome and the messenger for protein A (m.w. 104 kD), involved in viral RNA replication (3). The 3' terminal region of RNA1 also encodes the sequence of RNA3 (9S) found in infected cells but not in virions. RNA3 is the messenger for protein B (m.w. 10 kD) involved in the regulation of translation of protein A (4). Virion RNA2 (15S) is the genome and the messenger for protein alpha (m.w. 47 kD) which is proteolytically processed to give viral coat protein. RNA2 also exerts a regulatory function on the production of RNA3 (3). RNA1 can replicate independently of RNA2 and its gene product protein B (3). Crystals of BBV have been obtained and their analysis is underway (5).

We undertook nucleotide sequence analysis of the BBV RNAs to search for signals for RNA encapsidation, RNA replication and other functions crucial for viral infectivity. In this report, we present the sequence of BBV RNA2 and thereby the primary structure of the encoded virion coat protein and examine other features suggested by the sequence.

MATERIALS AND METHODS

Enzymes and isotopes required for direct RNA sequencing were as described (6-8). Other sources were as follows: reverse transcriptase (Life Sciences, Inc.); tobacco acid pyrophosphatase, E. coli poly A polymerase and vaccinia guanylyl transferase (B.R.L.); restriction enzymes (New England Biolabs, Biotec or B.R.L.); oligodeoxynucleotides (P-L Biochemicals); (gamma-32P)ATP, (5000 Ci/mmole), (alpha-32P)ATP, (alpha-32P)GTP and (alpha-32P)dATP (400-800 Ci/mmole) and (5'-32P)pCp (3000 Ci/m mole), (Amersham). Low melting point agarose was from B.R.L.

Virion RNA2 was isolated as described (9). dsRNAs were isolated from BBV infected Drosophila cells by the LiCl precipitation method (10) and were purified in either 1.5% low melting point agarose gels or 6% polyacrylamide gels in Tris-EDTA-Borate buffer, pH 8.3.

Intact RNA2 was labeled at its 5' end with guanylyl transferase (11) after chemical decapping (12) and with polynucleotide kinase (6) after chemical or enzymatic decapping and phosphatase treatment (13).

Purified dsRNA2 was labeled at its 3' end with 32pCp and T4 RNA ligase (14). Positive and negative strands were separated on 5% polyacrylamide gels with low cross-linking (8). RNA fragments, obtained after complete RNase A digestion (15) or partial T1 ribonuclease digestion (6), were labeled at their 3' ends with pCp and T4 RNA ligase and were fractionated on 10% or 20% urea-polyacrylamide gels (8).

End labeled RNAs were sequenced by enzymatic cleavage methods (6,16) and also by the chemical method (17).

Single stranded BBV2 cDNA was prepared by random priming with partially digested calf thymus DNA as described by Taylor et al. (18). Actinomycin D (40 ug/ml) was used to prevent double stranded DNA formation. The cDNAs were fragmented with restriction enzymes HaeIII, AluI, TaqI, HpaII and HhaI all of which recognize ss DNA (19,20).

Double stranded cDNA was prepared from ss cDNA and cloned into pBR322 as described (21). The cDNA inserts were separated from pBR322 sequences by PstI digestion and gel electrophoresis. Fragments were prepared by digestion with suitable restriction enzymes, phosphatase treatment, polynucleotide kinase labeling, and strand separation (8,22).

Such DNA, above, was sequenced by the chemical method (8).

RESULTS AND DISCUSSION

The sequence of BBV RNA2 is given in Fig. 1.

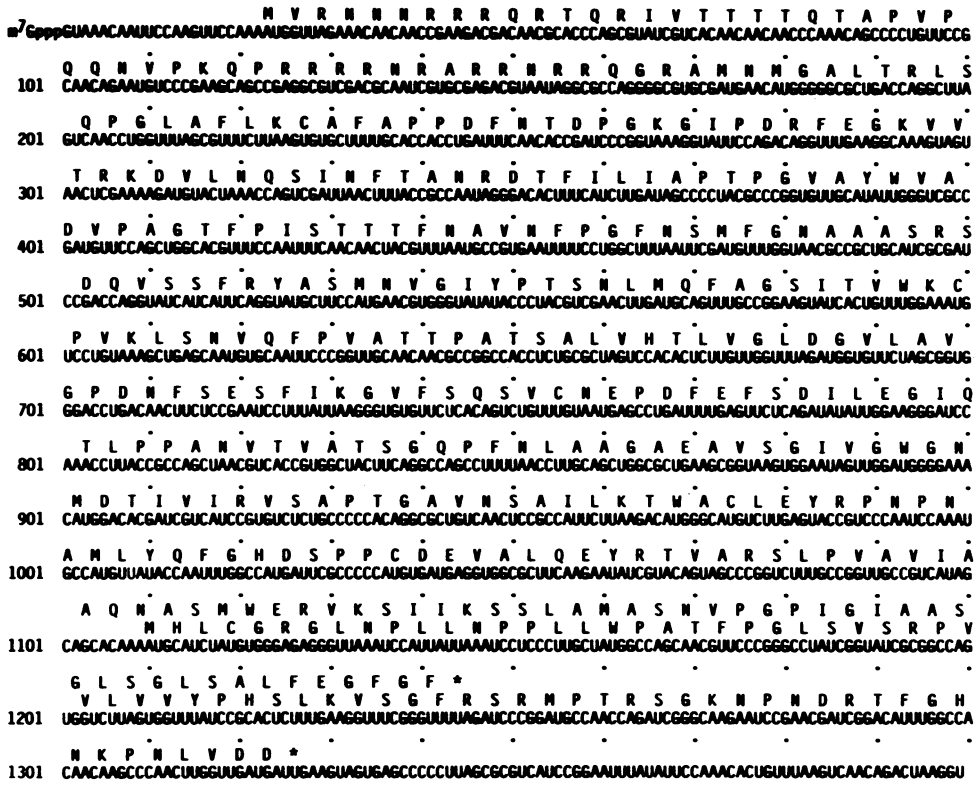

Fig. 1. Nucleotide sequence of BBV RNA2 and the encoded amino acid sequence of the viral coat protein precursor. Asterisks show the termination codons for the two open reading frames.

5' Terminal Structure and Direct RNA sequencing

BBV RNA2 could be labeled at the 5' end only after chemical or enzymatic treatment appropriate for removing a capped structure. Attempts to label after alkaline phosphatase treatment alone were not successful. Analysis of uniformly labeled RNA2 by enzymatic digestion and paper electrophoresis showed the capped structure to be m7GpppGp.

In preliminary studies of 5' labeled RNA the sequence of the first 10 5'-proximal bases was determined by mobility shift analysis with two dimensional paper electrophoresis and homochromatography (23).

The first 400 bases were determined by enzymatic sequencing on gels (16).

cDNA Sequencing

All but the 67 3' terminal bases were determined by chemical sequencing of

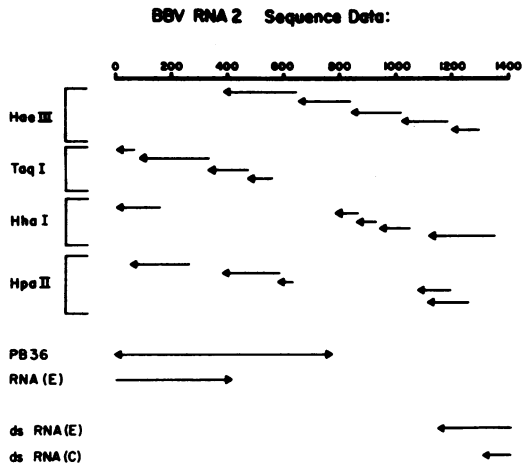

Fig. 2. Schematic diagram summarizing the BBV RNA2 sequence data. The scale at the top indicates the nucleotide position. The arrows, which are aligned with the scale, show the direction and extent of the sequence determined in each experiment. HaeIII, TaqI, HhaI and HpaII indicate ss cDNA fragments generated by the respective restriction enzymes. PB36 is a cDNA clone used to determine sequences in both directions. All DNA sequencing was done by the chemical method of Maxam and Gilbert. RNA(E) indicates 5'-labeled RNA2 sequenced by the enzymatic method. dsRNA(E) and dsRNA(C) indicate the positive strand of 3'-labeled dsRNA2 sequenced by enzymatic and chemical methods respectively.

restriction fragments of cDNA (See Fig. 2) that had been synthesized with reverse transciptase with partially digested calf thymus DNA serving as a primer (18). We avoided cloning procedures which might arbitrarily select a particular sequence in the event of heterogeneity.

We were unsuccessful in polyadenylating the 3' terminus of virion-derived RNA2 which would have allowed primed cDNA synthesis to begin there. We used standard (24) as well as denaturing conditions e.g. prior heating to 100°C, prior treatment with phosphatase, reaction at low ionic strength, and addition of DMSO. Such virion-derived RNA2 also resisted enzymatic addition of pCp which would have permitted use of direct RNA sequencing procedures.

Sequencing of the 3' terminal Regions of dsRNA2

Double stranded RNA2, isolated from infected cells, accepted both adenylate and pCp under standard conditions with poly A polermerase and T4 RNA ligase, respectively. Thus dsRNA2 was labeled with pCp at its 3' termini, the strands were separated on a 5% nondenaturing gel (8). and the sequence of the 200 3' terminal bases of each were determined. The sequence at the 3' end of positive strand agreed with the sequence of ssRNA2 already determined and

extended it to its terminus. The sequence at the 3' end of the negative strand was identical to the compliment of the sequence at the 5' end of ssRNA, indicating that dsRNA2 indeed contains a full length copy of virion RNA.

In order to obtain additional structural information about the RNA2 3' terminus, a complete RNase A digest of the plus strand of dsRNA2 was treated with ligase in the presence of p32pCp. Only the original terminus should contribute an hydroxyl group available for labeling. Subsequent fractionation yielded a labeled hexamer whose sequence we determined to be ApApGpGpUpCp in agreement with the determined 3' terminal sequence. Single stranded RNA2, derived from virions, was treated similarly. There was no significant labeling of any of the products of the RNase A digestion. Because short RNA fragments are unable to make base pairs within themselves these results suggest that virion RNA2 has a 3' terminal structural feature other than secondary structure that prevents pCp or adenylate addition.

Sequencing of Cloned cDNA

Although the sequence determinations described above were completely unambiguous, it was possible that a sequencing band might be obscured by the existence of a strong secondary structure feature. To minimize this chance for error we examined sequences of the negative strands of some cDNA clones that had been prepared for use as probes in experiments unrelated to this report. In particular a clone comprised of the first 800 bases from the 5' end BB2 DNA was sequenced in both directions after fragmentation with restriction enzymes HaeIII, DDEI, Sau3A and Sau96. This, and all of our other sequence determinations of cDNA clones, completely confirmed our original analyses.

Features of the Sequence

The first AUG codon for BBV RNA2 is 23 bases from the 5' end. Consistent with the eukaryotic ribosome binding model of Kozak (25), the third prior base is A and a G immediately follows the AUG. The leader sequence is markedly rich (45%) in A. The open reading frame following the initial AUG extends to base 1243 (i.e. 156 bases from the 3' terminus, Fig. 3) encoding for 407 amino acid residues in good agreementh with electrophoretic determinations of the size of coat protein precursor. A second open reading frame extends from base 1110 to 1325 (74 bases from the 3' terminus) coding for a putative protein 72 amino acid residues in length. There is a long open reading frame, corresponding to a protein of 267 amino acid residues in the negative strand (Fig. 3). The codon usage in BBV coat protein shows a slight preference for U in the third position. There are five cysteine residues, all coded by UGU.

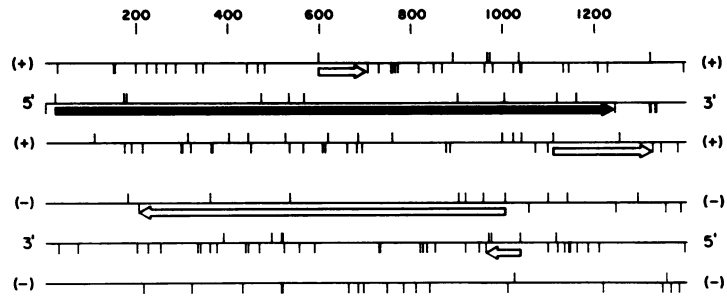

Fig. 3. Map of BBV RNA2 showing open reading frames. The top three lines show the possible reading frames in the virion RNA and the bottom three lines show the same in the complementary strand. The marks above the lines indicate positions of AUG codons, the marks below the lines indicated positions of UGA, UAG or UAA codons. The cistron for coat protein precursor alpha is shown by the solid arrow. Others are shown by open arrows.

The amino acid sequence of coat protein precursor (Fig. 1) shows that the N-terminal structure is very rich in basic amino acids. Sixteen amino acids out of first 50 are arginine. There are 27 arginines in the coat protein. Also, asparagines, threonines and arginine appear in a row towards the N-terminus. This gives this end of coat protein a distinctive structure similar to that existing with plant viruses (26-29) where it is thought to be involved in the protein-RNA interaction necessary for encapsidation.

Identical stretches of 6 bases (bases 9-14 and bases 16-21) and 8 bases (bases 33-40 and bases 71-78) exist near the 5' end of RNA2.

We observed some evidence of sequence heterogeneity. Position 627 showed a minor band due to A in chemical cDNA sequencing. Also, in enzymatic RNA sequencing, position 1394 showed a minor band due to A and positions 1395 and 1396 showed minor bands due to G.

Secondary Structure

Various computer programs have been devised for predicting RNA secondary structure given the base sequence and assuming particular values for the strength of hydrogen bonding, stacking and other interactions. Among these the method of Zuker and Stiegler (30) has reproduced for viroid RNA and ribosomal RNA, the secondary structures that had already been derived on the basis of a variety of physical, enzymological and geneological evidence. Application of the Zuker and Stiegler program to RNA2 gave a complicated melange of stems and loops that suggest that BBV RNA2 has a remarkably stable and distinctive secondary stucture in solution. As shown in Fig. 4, the 3'

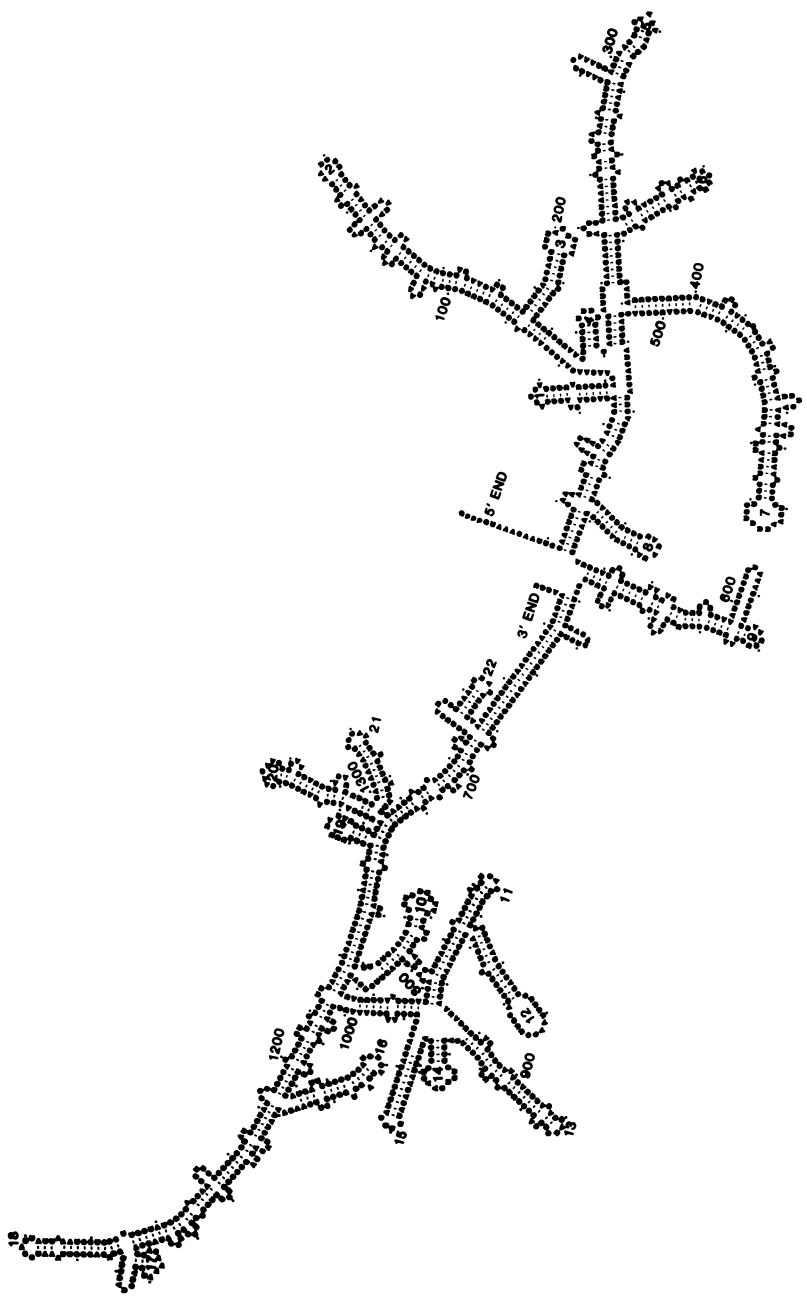

Fig. 4. Computer generated secondary structure of BBV RNA2. Stem-loop structures are numbered consecutively from the 5' end (1 through 22). Bases are numbered in multiples of 100.

half of the RNA can fold into a single, long, imperfect hairpin that has several short stem and loop regions and a large (200 bases long) multiply bifurcated set of stems and loops projecting from its side, the base of the hairpin being stabilized by an exceptionally long (38 bases) helix. The 5' half can fold into an intricate, multiply bifurcated structure also of high predicted stablility.

We have systematically analyzed other possible folding by examination of so-called dot patterns which match all G-C, G-U and A-U base pairs and graphically display those contiguous pairs of a specified minimum length or pairing strength. We conclude that almost all of the strongly base paired regions of Fig. 4 that isolate small loop regions, persist over a range of choices of energy parameters (such parameters as energy gained for particular base pairs and energy lost by looping out of particular bases) and are likely to exist in neutral solutions (e.g. bases 900-924). However base pairings in Fig. 4 between regions whose sequnce separation is large (even e.g. such sets as bases 709-727 vs. 1390-1372) are not strongly favored over markedly different base pairing choices.

ACKNOWLEDGMENTS

We thank Michael Janda for excellent technical assistance. For some of our experiments Tom Gallagher has kindly provided highly purified RNA2. This research was supported by NIH grants AI 1466 and AI 15342 and by NIH Career award AI 21942.

*Present Address: Institute of Chemical Biology, Calcutta, India

+Present Address: Department of Entomology, Texas A and M University, College Station, TX 77843, USA

§To whom reprint requests should be addressed

REFERENCES

1. Longworth, J. (1978) Adv. Virus Res. 23, 103-157.
2. Longworth, J. and Carey, G. (1976) J. Gen. Virol. 33, 31-40.
3. Gallagher, T., Friesen, P., and Rueckert, R. (1983) J. Virol. 46, 481-489.
4. Friesen, P. and Rueckert R. (1984) J. Virol. 49, 116-124.
5. Hosur, M. Schmidt, T. Tucker, T. Johnson, J. Selling, B. and Rueckert, R. (1984) Virology 133, 119-127.
6. Dasgupta, R., Ahlquist, P. and Kaesberg, P. (1980) Virology 104, 339-346.
7. Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1981) Cell 23, 183-189.
8. Maxam, A. and Gilbert, W. (1980) In "Methods in Enzymology" (L. Grossman, ed.), Vol. 65, pp. 499-560. Academic Press, New York.

9. Guarino, L., Hruby, D., Ball, A. and Kaesberg, P (1981) J. Virol. 37, 500-505.
10. Baltimore, D. (1966) J. Mol. Biol. 18, 421-428.
11. Moss, B. (1977) Biochem. Biophys. Res. Commun. 74, 374-383.
12. Kemper, B. (1976) Nature 262, 321-323.
13. Efstratiadis, A., Vournakis, J., Donis-Keller, H. Chaconas, G., Dougal, D. and Kafatos, F. (1977) Nucl. Acids Res. 4, 4165-4174.
14. England, T. and Uhlenbeck, O. (1978) Nature 275, 560-561.
15. Sanger, F., Brownlee, G. and Barrell, B. (1965) J. Mol. Biol. 13, 373-398.
16. Donis-Keller, H. (1980) Nucl. Acids Res. 8, 3133-3142.
17. Peattie, D. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
18. Taylor, J., Illmensee, R. and Summers, J. (1976) Biochim. Biophys. Acta 442, 324-330.
19. Rice, C. and Strauss, J. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2062-2066.
20. Dasgupta, R. and Kaesberg, P. (1982) Nucl. Acids Res. 10, 703-713.
21. Ahlquist, P., Luckow, V. and Kaesberg, P. (1981) J. Mol. Biol. 152, 23-38.
22. Maniatis, T., Fritsch, E. and Sambrook, J. (1982) In "Molecular Cloning--A Laboratory Manual," pp. 98-106. Cold Spring Harbor Laboratory, New York.
23. Dasgupta, R. and Kaesberg, P. (1977) Proc. Natl. Acad. Sci. USA 74, 4900-4904.
24. Devos, R., Van Emmelo, J., Seurinck-Opsomer, C. Gillis, E. and Fiers, W. (1976) Biochim. Biophys. Acta 447, 319-327.
25. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
26. Harrison, S., Olson, A. Schutt, C. and Winkler, F. (1978) Nature 276, 368-373.
27. Moosic, J. McKean, D. Shih, D. and Kaesberg, P. (1983) Virology 129, 517-520.
28. Rossmann, M., Abad-Zapatero, C., Hermodson, M. and Erickson, J. (1983) J. Mol. Biol. 166, 37-83.
29. Tremaine, J., Ronald, W. and Agarwal, H. (1977) Virology 83, 404-412.
30. Zuker, M. and Stiegler, (1981) Nucl. Acids Res. 9, 133-148.