Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

J. Mol. Biol. (1985) **182**, 183–189

Structure of the Black Beetle Virus Genome and its Functional Implications

Bimalendu Dasmahapatra, Ranjit Dasgupta, Amit Ghosh† and Paul Kaesberg

Biophysics Laboratory and Biochemistry Department
University of Wisconsin, Madison, Wisc. 53706, U.S.A.

(Received 4 September 1984)

The black beetle virus (BBV) is an isometric insect virus whose genome consists of two messenger-active RNA molecules encapsidated in a single virion. The nucleotide sequence of BBV RNA1 (3105 bases) has been determined, and this, together with the sequence of BBV RNA2 (1399 bases) provides the complete primary structure of the BBV genome. The RNA1 sequence encompasses a 5′ non-coding region of 38 nucleotides, a coding region for a protein of predicted molecular weight 101,873 (protein A, implicated in viral RNA synthesis) and a 3′ proximal region encoding RNA3 (389 bases), a subgenomic messenger RNA made in infected cells but not encapsidated into virions. The RNA3 sequence starts 16 bases inside the coding region of protein A and contains two overlapping open reading frames for proteins of molecular weight 10,760 and 11,633, one of which is believed to be protein B, made in BBV-infected cells. A limited homology exists between the sequences of RNA1 and RNA2. Sequence regions have been identified that provide energetically favorable bonding between RNA2 and RNA1 possibly to facilitate their common encapsidation, and between RNA2 and negative strand RNA1 possibly to regulate the production of RNA3.

1. Introduction

The black beetle virus is an isometric insect virus, named after the black beetle (*Heteronychus arator*), the host from which it was first isolated (Longworth & Archibald, 1975). It is a member of the Nodaviridae family. Its genome consists of two messenger-active RNA molecules, RNA1 and RNA2, with sedimentation coefficients of 22 S and 15 S, respectively (Longworth & Carey, 1976). RNA1 serves as messenger for 104,000 M_r protein A (Friesen & Rueckert, 1981). RNA2 directs the synthesis of the 47,000 M_r virion capsid protein precursor (protein alpha, from which the 43,000 M_r coat protein is processed; Friesen & Rueckert, 1981). Cells infected with BBV‡ produce an additional messenger, RNA3, coding for a protein of molecular weight 10,000, designated protein B (Friesen & Rueckert, 1982). Synthesis of RNA3 and of protein B is also induced in cells transfected with RNA1 alone at a level of synthesis higher than when RNA2 is present (Gallagher *et al.*, 1983). Analyses in cell-free protein synthesizing systems show that the protein B cistron is silent on RNA1, suggesting that its expression in cells requires synthesis and subsequent translation of RNA3.

Here we report the nucleotide sequence of RNA1 and demonstrate that the RNA3 sequence is contained in the 3′ proximal region of RNA1. We describe computer analyses of the secondary structure of RNA1 and of sequence relationships to RNA2 in terms of the functions required for virus synthesis.

2. Materials and Methods

(a) *Materials*

Enzymes were obtained from following sources: AMV reverse transcriptase, Life Sciences, Inc.; restriction enzymes, New England Biolabs, B.R.L. or Biotec; nuclease S_1, DNA polymerase 1, and terminal deoxynucleotidyl transferase, P-L Biochemicals; polynucleotide kinase and poly(A) polymerase, B.R.L. Nucleotides and synthetic oligodeoxynucleotides were obtained from P-L Biochemicals. An oligodeoxynucleotide synthesis kit was obtained from New England

† Present address: Indian Institute of Chemical Biology, Jadavpore, Calcutta 700 032, India.

‡ Abbreviation used: BBV, black beetle virus.

0022-2836/85/060183-07 $03.00/0

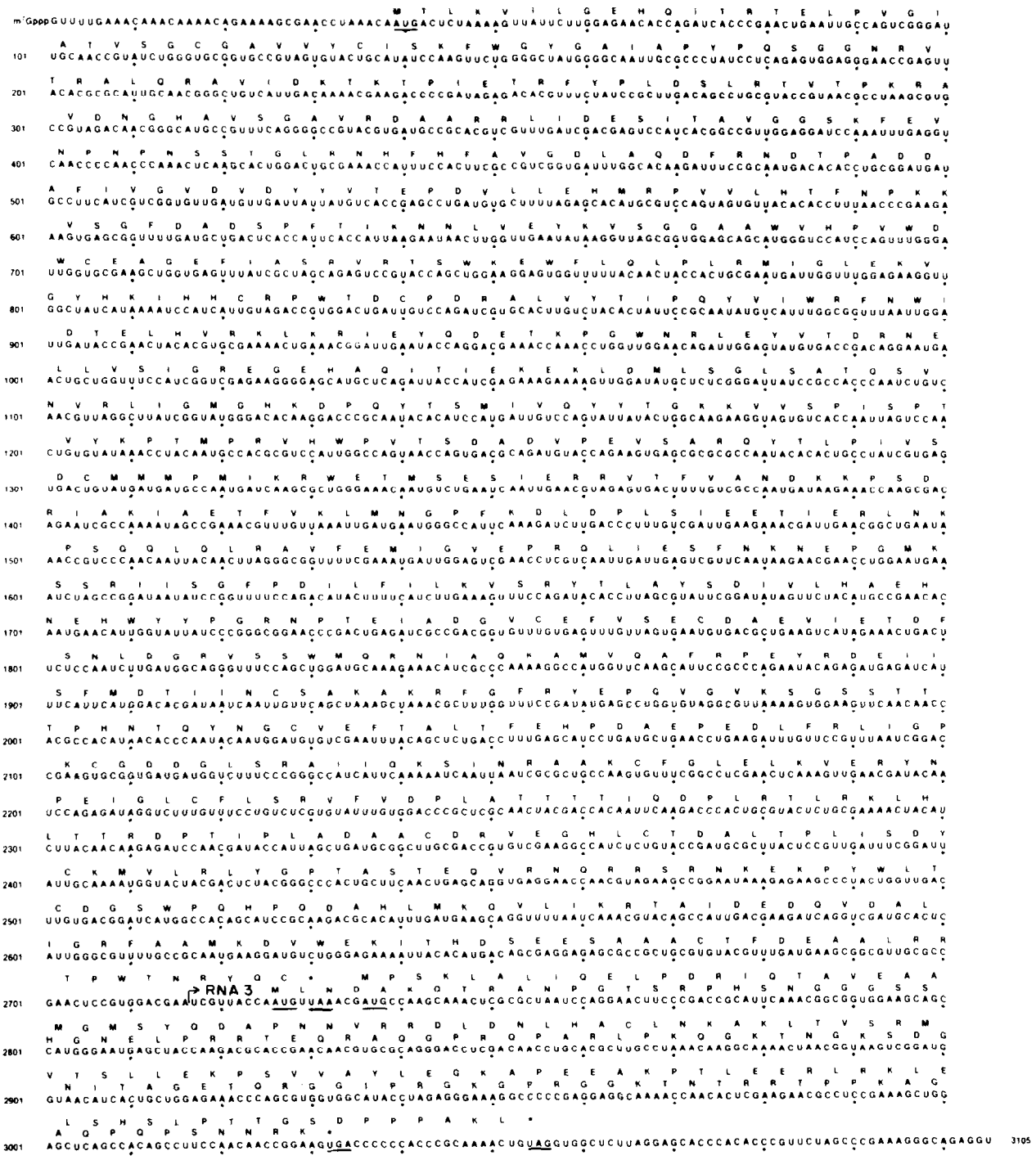

Figure 1. Nucleotide sequence of BBV RNA1. Initiation and termination codons for major open reading frames are underlined and the encoded amino acids are given over the first base of each codon, asterisks denoting termination codons.

BioLabs. [α-^{32}P]dATP, [α-^{32}]ATP (410 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol) were obtained from Amersham.

(b) *Preparation of BBV RNA1 and double-stranded RNAs*

BBV RNA1 was isolated and purified as described (Guarino *et al.*, 1981). Double-stranded RNAs were isolated from *Drosophila* cells 8 h after their infection with BBV as described (Guarino *et al.*, 1984).

(c) *Cap structure and direct RNA sequencing*

The 5′ end cap structure of BBV RNA1 was determined as described (Dasgupta *et al.*, 1976).

RNA1 was decapped enzymatically (Efstratiadis *et al.*, 1977), dephosphorylated with alkaline phosphatase and was treated with kinase by [γ-^{32}P]ATP and phage T4 polynucleotide kinase (Dasgupta *et al.*, 1980). Such labeled RNA1 was purified by polyacrylamide gel electrophoresis and sequenced by mobility shift analysis

(Dasgupta & Kaesberg, 1977) and by enzymatic RNA sequencing (Donis-Keller, 1980).

(d) *Complementary DNA synthesis, cloning and sequencing*

DNA complementary to BBV RNA1 was synthesized with reverse transcriptase, with $p(dT)_{12}$ or with partially digested calf thymus DNA as primers. Single-stranded cDNAs were converted to the double-stranded form which, after oligo(dC)-tailing with terminal transferase, were annealed to pBR322 that had been linearized with *Pst*1 and tailed with oligo(dG). These were used to transform *Escherichia coli* Mm-294 (Daghert & Ehrlich, 1979; Ahlquist *et al.*, 1981*a*). Clones containing recombinant plasmids were selected on the basis of their resistance to tetracycline and sensitivity to ampicillin (Villa-Komaroff *et al.*, 1978) and then were screened for the presence of BBV RNA1 sequences by the *in situ* colony hybridization technique of Grunstein & Hogness (1975). Randomly primed cDNAs of RNA1 and subsequently inserts, nick-translated by the method of Maniatis *et al.* (1975), were used as hybridization probes. Recombinant plasmids carrying BBV1 cDNA inserts were isolated by the alkaline lysis procedure (Ish-Horowicz & Burke, 1981; Brinboim & Doly, 1979). Plasmid DNAs were cleaved by *Pst*I. Their cDNA inserts were purified by polyacrylamide gel electrophoresis and were sequenced by the chemical method of Maxam & Gilbert (1980). Experiments involving recombinant plasmids were carried out under PI/EKI containment as prescribed by the National Institutes of Health guidelines.

(e) *Computer analyses*

Sequences were assembled with the computer program of Staden (1980) and were further analyzed with software from the UW Genetics computer group (Devereaux *et al.*, 1984). Secondary structure was analyzed by the methods of Zuker & Stiegler (1981).

3. Results

(a) *The BBV RNA1 sequence*

The sequence of BBV RNA1 (3105 bases) is presented in Figure 1. The 5′ end has the cap structure 7mGpppGp.

(b) *Determination of the sequence*

The sequence of the first ten bases from the 5′ end of decapped RNA1 was determined by mobility shift analysis and the sequence of 234 nucleotides following the capping group was obtained by enzymatic RNA sequencing.

It had been our intention to polyadenylate the 3′ terminus of RNA1 and to use oligo(dT) as a primer to synthesize full-length DNA copies to be used for direct sequencing and for cloning and subsequent sequencing. Polyadenylation of the 3′ terminus proved to be difficult (see below), however, we found that oligo(dT) served as efficient internal primer on RNA1 and we used this to our advantage for obtaining most of the RNA1 sequence. Such cDNAs, obtained by internal priming, were cloned and several of the recombinant plasmids carrying

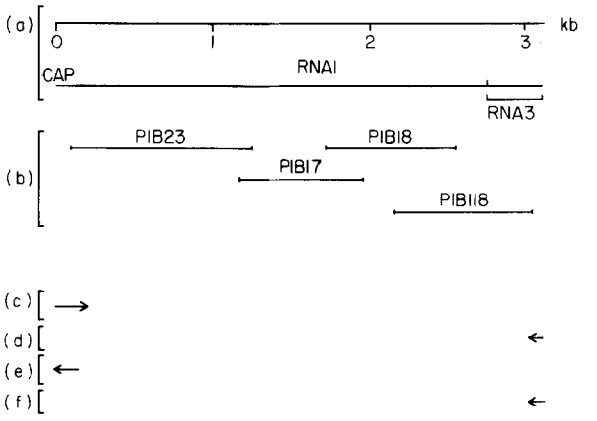

Figure 2. Summary of sequence data for BBV RNA1. (a) Schematic of BBV RNA1 showing the location of RNA3 on RNA1. (b) Plasmid cDNA clones of BBV RNA1 sequenced. (c) Direct RNA sequencing results from the 5′ end-labeled RNA1. (d) RNA sequencing result obtained from the 3′ end-labeled positive strand of double-stranded RNA3. (e) DNA sequencing result of the *Hha*I fragment of random primed cDNA of RNA1. (f) Sequencing result of cDNA made with the synthesized deoxyoligonucleotide as primer on the 3′ end of RNA1. kb = 10^3 bases.

RNA1 inserts were isolated. The largest insert, designated PIB23 (Fig. 2(b)), was sequenced and was found to overlap the 5′ end region of RNA1 (bases 89 to 1244) that had been sequenced enzymatically.

Most of the other clones were able to hybridize to PIB23, but restriction enzyme mapping showed that none extended the sequence beyond that of PIB23 in either direction. A non-hybridizable insert, designated PIB18, was sequenced; it was 790 bases long and, as we will demonstrate below, mapped 3′ to PIB23.

To prepare clones covering other regions, cDNAs synthesized with calf thymus DNA fragments as primers, were used to generate a plasmid cDNA clone bank. A total of 300 colonies were examined by hybridization with ^{32}P-labeled PIB23 and PIB18 DNA inserts. Three groups of clones were selected for analysis: group 1 clones hybridized to both PIB23 and PIB18; group 2 clones hybridized to PIB18 only; and group 3 clones hybridized to neither. The recombinant plasmid in group 1 containing the largest cDNA insert was selected (PIB17) and its insert was sequenced. This sequence of 969 bases joined PIB23 and PIB18 and produced a contiguous sequence of 2533 bases from the 5′ end. Group 2 clones were analyzed with restriction enzymes and from these the clone PIB118 was selected and its insert sequenced. It started at base 2153 and extended the sequence to base 3022. All of the group 3 clones were hybridizable to PIB118 but restriction analysis showed that none contained sequences further downstream.

Various procedures were tried to modify the primary or secondary structure of the 3′ terminus of RNA1 none of which facilitated its polyadenylation. (1) the RNA was heated for five minutes at

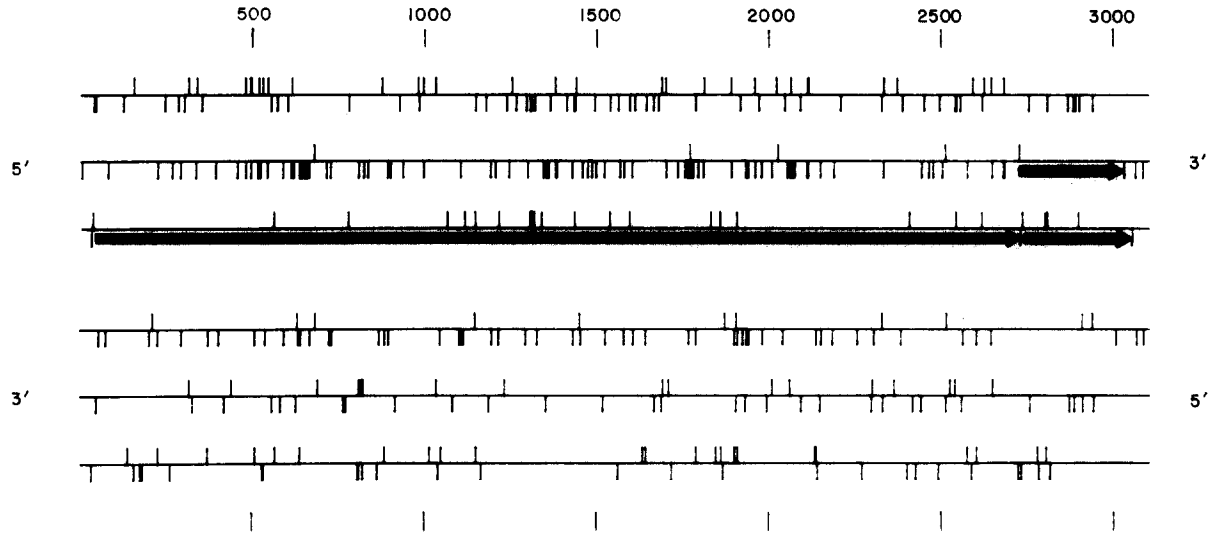

Figure 3. Translational reading frames encoded in the RNA1 sequence showing the 3 possible open reading frames as filled arrows. Vertical bars above and below the horizontal lines denote AUG codons and termination codons, respectively. The first 3 lines are all 3 reading frames of positive strand and the last 3 lines are those of negative strand.

70°C followed by quick chilling at 0°C just prior to attempting polyadenylation; (2) polyadenylation was tried in the presence of CH_3HgOH; (3) the RNA was treated with alkaline phosphatase to remove a possible phosphate group. We were also unsuccessful in attempts to label the RNA1 3′ terminus with [^{32}P]pCp and phage T4 ligase (England & Uhlenbeck, 1978) as a preliminary to enzymatic sequencing. Additional experiments (to be described elsewhere) now indicate that a protein is bound to the 3′ terminus of RNA1.

Previously we had determined the sequence of BBV RNA3 (389 bases) by direct enzymatic RNA sequencing methods and by chemical sequencing of its cDNA (Guarino *et al.*, 1984) and this proved to be helpful in completing the RNA1 sequence. Since RNA3 is produced in RNA1-transfected cells but is not needed for infectivity (Gallagher *et al.*, 1983) we assumed that its entire sequence is encoded in RNA1, and indeed inspection showed that the sequence of the 306 bases at the 5′ end of the RNA3 sequence were found to be identical to the sequence of the 306 bases at the 3′ end of the partial RNA1 sequence, above. Thus, we synthesized a DNA oligonucleotide complementary to the 15 bases at the 3′ terminus of RNA3 sequence and used it for RNA1 priming. This oligonucleotide served as an efficient primer for cDNA sequencing of RNA1 and, as expected, yielded an RNA1 sequence identical to the remaining 83 bases of RNA3. We cannot unequivocally rule out the possibility that bases exist to the right of position 3105. However, this is unlikely inasmuch as we were able to label the 3′ termini of both double-stranded RNA1 and RNA3 (obtained from infected cells) with [^{32}P]pCp and T4 ligase, and showed that the 3′ termini of their positive strands have the sequence we reported immediately to the left of base 3105, and moreover that these sequences terminate with a 3′ hydroxyl group.

Figure 2 indicates the methods used to determine various parts of the sequence.

(c) *Sequence characteristics*

The molecular weight of RNA1 calculated from the sequence is $1{\cdot}02 \times 10^6$ and is in good agreement with the estimated molecular weight of RNA1 as obtained by denaturing gel electrophoresis (Longworth & Carey, 1976) and by oligonucleotide fingerprinting (Clewley *et al.*, 1982).

The RNA1 sequence contains several A-rich regions, among them the sequence G-A-A-A-G-A-A-A-A-G (bases 1052 to 1061) which we judge to be the site of the observed strong internal priming with oligo(dT).

The longest open reading frame follows the first AUG codon at bases 39 to 41 and terminates with a UAA codon at bases 2730 to 2732. The 897 amino acid sequence coded by this frame corresponds to a protein of molecular weight 101,873 (Fig. 3), which is in good agreement with a previous estimate of 104,000 for the molecular weight of protein A (Friesen & Rueckert, 1981). The second 5′ proximal initiating codon occurs at bases 564 to 566 and would correspond to a protein too small to be protein A. The other two reading frames are tightly closed in the region 1 to 2700, thus precluding their coding for other proteins of substantial size. The longest open reading frame in the negative strand of RNA1 is only 210 bases long. The RNA3 sequence starts at position 2717. Two open reading frames exist in the RNA3 region, following AUG codons at bases 2736 to 2738 in the protein A phase and bases 2726 to 2728 in a second phase; the third frame is tightly closed. These frames are 318 and 306 bases in length, encoding putative proteins of molecular weight 11,633 (designated protein B2) and 10,760 (designated protein B1), either or both of which could be protein B, found in BBV-infected cells (see the Discussion).

No striking homology was observed among the 5′-terminal nor among the 3′-terminal sequences of RNA1 and RNA2. RNA1 (bases 7 to 12) and RNA2 (bases 3 to 8) both have the sequence A-A-A-C-A-A near their 5′ termini. The sequence A-G-G-U is conserved at the 3′ end of RNAs 1 and 2. Neither were there strong homologies in the coding regions. Nine-base-long homologies exist; bases 3019 to 3027 in RNA1 *versus* bases 34 to 42 in RNA2, and bases 2663 to 2671 in RNA1 *versus* bases 484 to 492 in RNA2.

4. Discussion

This study, together with the sequence determination of BBV RNA2 (Dasgupta *et al.*, 1984), provides the first complete primary structure of the genome of an insect virus, a member of the Nodaviridae family.

With the availability of these sequences it becomes possible to delineate the previously detected proteins more precisely and also to identify other viral proteins. Figure 4 maps the known proteins A and alpha, a BBV replicase component and the virion coat protein precursor, respectively. Mapped also are the two candidates (B1 and B2) for protein B and for a putative 8000 M_r protein on RNA2. No other proteins of substantial size are encoded. Synthesis of BBV must thus be accomplished by means of these proteins and the three BBV RNAs together with constituents provided by the infected cells.

Even though RNAs 1 and 2 exist in roughly equimolar amounts throughout infection the total synthesis of protein A is far less than that of protein alpha, in accordance with their structural and enzymatic functions, respectively. Although it is likely that production of these proteins is regulated at several levels in the course of virus replication, *in vitro* studies suggest that regulation at the level of initiation of translation is very significant. Both proteins are translated well in homologous (*Drosophila* lysates) and in heterologous (rabbit reticulocyte lysates) cell-free systems, synthesis of protein alpha being much greater than that of protein A (Guarino *et al.*, 1981; Friesen & Rueckert, 1984). It is known that initiation of translation of eukaryotic messenger RNA is favored by the existence of an A, three nucleotides before the initiating codon, and a G following the initiation codon (Kozak, 1981). We note that both leader

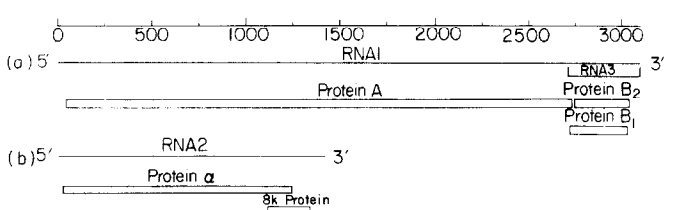

Figure 4. Genetic map of BBV. (a) Schematic of RNA1 showing the coding region of protein A and those of putative proteins B1 and B2 (open bars). (b) Schematic of RNA2 showing the coding regions of protein alpha and the 8000 M_r (8k) protein (open bars).

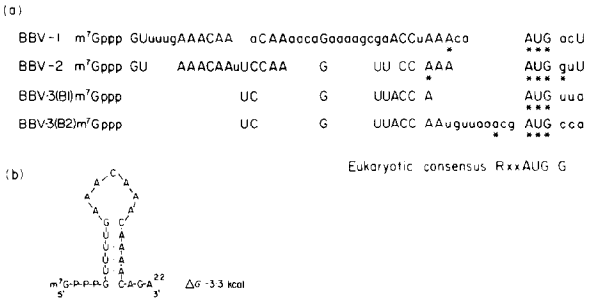

Figure 5. (a) Comparison of the 5′ sequences, translation initiation sites of BBV RNAs. Agreement between 2 sequences as indicated by bold letters. Asterisks represent homology to eukaryotic initiation consensus sequences. (b) Possible stem and loop structure within the 5′ end sequence of RNA1.

sequences are short, and that the leader sequence of RNA1 (38 bases) is 55% A, and the leader sequence of RNA2 (22 bases) contains 45% A. Both RNAs have an A in the −3 position. They differ in that RNA2 has a G in the +4 position as well (Fig. 5(a)). They also differ in that the first 19 bases of the leader sequence of RNA1 (but not of RNA2) can be folded into a stem and loop structure (Fig. 5(b)), which might be expected to impair initiation of translation.

Protein B is the only other new protein detected in BBV-infected *Drosophila* cells. Initially it was discovered as a cell-free translation product of RNA3 but not of RNA1. Now, with the demonstration from the sequence data of the location of the RNA3 sequence in the 3′-terminal region of RNA1, and the existence in that region of two open reading frames, it is unclear which frame represents protein B. Generally, reading frames not encoding proteins are tightly closed, and this suggests the possibility

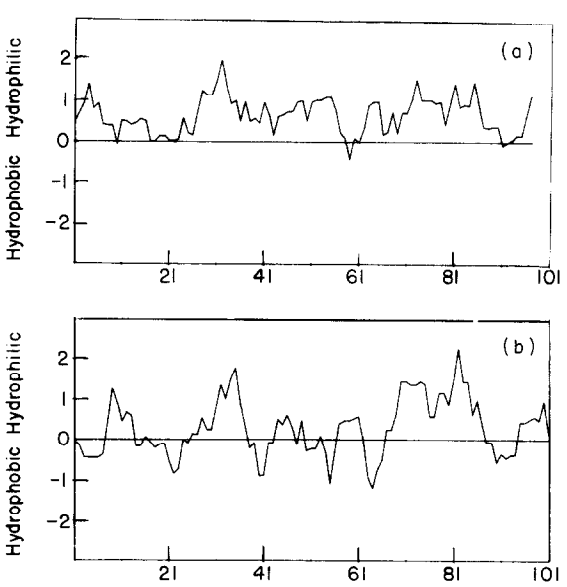

Figure 6. Relative hydrophobicity plots of putative proteins B1 (a) and B2 (b). Hydrophobicity was calculated according to the method of Hopp & Woods (1981).

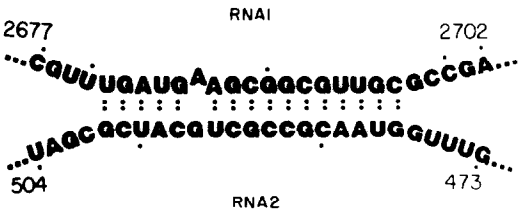

Figure 7. Possible complimentarity between RNA1 and RNA2 sequences as shown by Watson–Crick base-pairing. Numbers indicate the region in RNA sequences. Sequence hyphens have been omitted for clarity.

that both B1 and B2 are functional proteins. The cistron for putative protein B1 would begin just prior to the termination codon for protein A, while that of putative protein B2 would be in the same reading frame as protein A and would begin three bases beyond its termination codon. The leader sequences are only 9 and 19 bases long for B1 and B2, respectively, but therein only B2 has a sequence regarded as desirable for initiation of translation (an A at the −3 position; Fig. 5(a)). Hydrophobicity plots for these putative proteins are shown in Figure 6. B1 would be an exceptionally hydrophilic protein which might be expected to be unfolded in aqueous solution, while protein B2 has a distribution of amino acids more typical of a soluble, folded protein.

RNA2 has a second open reading frame which starts inside the cistron of protein alpha (Fig. 4). No translation product corresponding to the second reading frame in the RNA2 is observed either in cell-free translation systems used under standard conditions or in BBV-infected cells.

RNAs 1 and 2 are encapsidated into the same virions and we might thus anticipate that they would have similar sequence regions serving as sites for initiation of encapsidation. Our inability to find such common sequences suggests the possibility that RNA1 and RNA2 may interact with each other so that only a single recognition site for coat protein would be needed. We thus looked for sequences on the two RNAs that would have the potential for interaction. Figure 7 shows the region of strongest base-pairing between RNA1 and RNA2 involving a set of 16 out of 17 base-pairs providing a helical interaction of greater than one turn.

RNA viruses that contain multiple genomic RNAs possess sequence homology at the 3′ end of their RNAs as a recognition site for replicase. The homology varies from 200 bases in plant viruses (Ahlquist *et al.*, 1981*b*) to 9 to 20 nucleotides in bunyaviruses (Bishop *et al.*, 1981). With several plant viruses it is the tertiary structure in the RNA that is recognized (Ahlquist *et al.*, 1981*b*). BBV RNAs 1 and 2 have only a limited sequence homology and we have thus far been unable to identify regions of secondary structure homology.

RNA3 is found in cells infected with BBV or with BBV RNA1 plus RNA2 and also in cells transfected with RNA1 alone. It is not encapsidated into black beetle virions. Our current results show that the RNA3 sequence exists uninterrupted on RNA1 in the region of its 3′ terminus indicating that splicing is not required to produce RNA3.

Several mechanisms have been proposed for the production of subgenomic RNAs from the genomes of RNA viruses involving specific nucleolytic cleavage or partial transcription of the genomic RNA plus or minus strands. In corona viruses (Baric *et al.*, 1983; Spaan *et al.*, 1983), fusion of 5′-terminal sequences of genomic RNA to the 5′ ends of coding regions of mRNAs to produce subgenomic RNAs has been reported. A well-documented case is that of Sindbis virus (Ou *et al.*, 1983) where the subgenomic RNA derives from partial transcription of the genomic negative strand. Furthermore, it has been demonstrated recently that the subgenomic RNA of brome mosaic virus (BMV RNA4) is made by partial transcription of BMV RNA3 negative strand (Miller *et al.*, 1985). We have therefore considered the possibility that BBV RNA3 is made by a similar mechanism. It has been shown (Gallagher *et al.*, 1983) that cells transfected with RNA1 alone produce not only RNA1 but also large quantities of RNA3 and protein B, and furthermore that synthesis of RNA3 and protein B is progressively less with increased amounts of RNA2 present in the RNA1 preparation used for transfection. The inhibitory effect of RNA2 on the production of RNA3 is insensitive to cycloheximide, a potent inhibitor of translation (T. Gallagher, personal communication), suggesting the possibility that the inhibition of RNA3 production occurs by a specific base-pairing with RNA2 itself, rather than an interaction with its translation product. We have analyzed base-pairing possibilities between RNA2 and negative strand RNA1. The longest region of Watson–Crick base-pairing is 13 bases and on RNA1 occurs just prior to the start of the RNA3

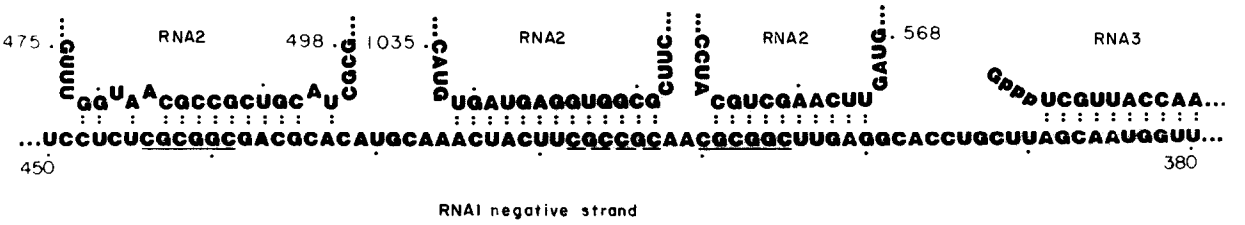

Figure 8. Watson–Crick base-pairing between the RNA2 positive strand and the RNA1 negative strand. Numbers indicate different regions in the RNA sequences. In RNA1 sequences, those above the continuous lines are direct repeats and that above the broken line is complementary to a direct repeat. Sequence hyphens have been omitted for clarity.

sequence; and two other long regions of possible interaction occur just to either side (see Fig. 8). These regions also contain a direct repeat and a sequence complementary to the direct repeat which can form a stable stem and loop structure (ΔG of about -20 kcal (1 kcal = 4·184 kJ) calculated as described by Salser (1977)). It is possible that these sequences act as a recognition site for replicase to initiate RNA3 synthesis, and the base-pairing with the RNA2 sequence impairs the recognition process thus inhibiting RNA3 production.

We thank Linda Guarino and Tom Gallagher for giving us BBV RNA in the initial stages of this project, Roland Littlewood for assistance in computer analyses and Keith Saunders for helpful discussions. This research was supported by the National Institutes of Health under Public Health Service grants AI-1466 and AI-15342 and Career Award AI-21942.

References

Ahlquist, P., Luckow, V. & Kaesberg, P. (1981*a*). *J. Mol. Biol.* **153**, 23–28.

Alquist, P., Dasgupta, R. & Kaesberg, P. (1981*b*). *Cell*, **23**, 183–189.

Baric, R. S., Stohlman, S. A. & Lai, M. M. C. (1983). *J. Virol.* **48**, 633–640.

Bishop, D. H. L., Clerx, J. P. M., Clerx-Van Haaster, C. M., Robeson, G., Rozhon, E. J., Ushijima, H. & Veerisetty, V. (1981). In *The Replication of Negative Strand Viruses* (Bishop, D. H. L. & Compans, R. W., eds), pp. 135–146, Elsevier/North-Holland, New York.

Brinboim, H. C. & Doly, O. J. (1979). *Nucl. Acids Res.* **7**, 1513–1523.

Clewley, J. P., Crump, W. A. L., Avery, R. V. & Moore, N. F. (1982). *J. Virol.* **44**, 767–771.

Daghert, M. & Ehrlich, S. D. (1979). *Gene*, **6**, 23–28.

Dasgupta, R. & Kaesberg, P. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4900–4904.

Dasgupta, R., Harada, F. & Kaesberg, P. (1976). *J. Virol.* **18**, 260–267.

Dasgupta, R., Ahlquist, P. & Kaesberg, P. (1980). *Virology*, **104**, 339–346.

Dasgupta, R., Ghosh, A., Dasmahaptra, B., Guarino, L. A. & Kaesberg, P. (1984). *Nucl. Acids Res.* **12**, 7215–7223.

Devereaux, J., Haeberli, P. & Smithies, O. (1984). *Nucl. Acids Res.* **12**, 387–395.

Donis-Keller, H. (1980). *Nucl. Acids Res.* **8**, 3133–3142.

Efstratiadis, A., Vournakis, J. N., Donis-Keller, H., Chaconas, G., Dougal, D. K. & Kafatos, C. F. (1977). *Nucl. Acids Res.* **4**, 4165–4174.

England, T. E. & Uhlenbeck, O. C. (1978). *Nature (London)*, **275**, 560–561.

Friesen, P. D. & Rueckert, R. R. (1981). *J. Virol.* **37**, 876–886.

Friesen, P. D. & Rueckert, R. R. (1982). *J. Virol.* **42**, 986–995.

Friesen, P. D. & Rueckert, R. R. (1984). *J. Virol.* **49**, 116–124.

Gallagher, T. M., Friesen, P. D. & Rueckert, R. R. (1983). *J. Virol.* **46**, 481–489.

Grunstein, M. & Hogness, D. S. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3961–3965.

Guarino, L. A., Hruby, D. E., Ball, L. A. & Kaesberg, P. (1981). *J. Virol.* **37**, 500–505.

Guarino, L. A., Ghosh, A., Dasmahapatra, B., Dasgupta, R. & Kaesberg, P. (1984). *Virology*, **139**, 199–203.

Hopp, T. P. & Woods, K. R. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 3824–3828..

Ish-Horowicz, D. & Burke, J. F. (1981). *Nucl. Acids Res.* **9**, 2989–2998.

Kozak, M. (1981). *Nucl. Acids Res.* **9**, 5233–5252.

Longworth, J. F. & Archibald, R. D. (1975). *New Zealand J. Zool.* **2**, 233–236.

Longworth, J. F. & Carey, G. P. (1976). *J. Gen. Virol.* **33**, 31–40.

Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 1184–1188.

Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.

Miller, W. A., Dreher, T. W. & Hall, T. C. (1985). *Nature (London)*, **313**, 68–70.

Ou, J. H., Strauss, E. G. & Strauss, J. H. (1983). *J. Mol. Biol.* **168**, 1–15.

Salser, W. (1977). *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985–1002.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., vander Zeist, B. A. M. & Siddel, S. G. (1983). *EMBO J.* **2**, 1839–1844.

Staden, R. (1980). *Nucl. Acids Res.* **8**, 3673–3694.

Villa-Komaroff, L., Efstratiadis, A., Broom, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 3727–3731.

Zuker, M. & Stiegler, P. (1981). *Nucl. Acids Res.* **9**, 133–148.

Edited by S. Brenner