JOURNAL OF VIROLOGY, Mar. 1994, p. 1382–1396
0022-538X/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 68, No. 3

Sequence and Genome Organization of Borna Disease Virus†

BEATRICE CUBITT, CHRISTOPHER OLDSTONE, AND JUAN CARLOS DE LA TORRE*

Department of Neuropharmacology, The Scripps Research Institute, La Jolla, California 92037

Received 16 September 1993/Accepted 13 November 1993

We have previously demonstrated that Borna disease virus (BDV) has a negative nonsegmented single-stranded (NNS) RNA genome that replicates in the nucleus of infected cells. Here we report for the first time the cloning and complete sequence of the BDV genome. Our results revealed that BDV has a genomic organization similar to that of other members of the *Mononegavirales* order. We have identified five main open reading frames (ORFs). The largest ORF, V, is located closest to the 5′ end in the BDV genome and, on the basis of strong homology with other NNS-RNA virus polymerases, is a member of the L-protein family. The intercistronic regions vary in length and nucleotide composition and contain putative transcriptional start and stop signals. BDV untranslated 3′ and 5′ RNA sequences resemble those of other NNS-RNA viruses. Using a set of overlapping probes across the BDV genome, we identified nine in vivo synthesized species of polyadenylated subgenomic RNAs complementary to the negative-strand RNA genome, including monocistronic transcripts corresponding to ORFs I, II, and IV, as well as six polycistronic polyadenylated BDV RNAs. Interestingly, although ORFs III and V were detected within polycistronic transcripts, their corresponding monocistronic transcripts were not detected. Our data indicate that BDV is a member of the *Mononegavirales*, specially related to the family *Rhabdoviridae*. However, in contrast to the rest of the NNS-RNA animal viruses, BDV replication and transcription occur in the nucleus of infected cells. These findings suggest a possible relationship between BDV and the plant rhabdoviruses, which also replicate and transcribe in the nucleus.

Borna disease (BD) is an infectious disease of the central nervous system characterized by profound behavioral abnormalities, inflammatory cell infiltrates, and the accumulation of disease-specific antigens in limbic system neurons (for recent reviews, see references 45, 47, and 59). Naturally occurring infections with Borna disease virus (BDV), the etiological agent of BD, have been confirmed only in horses and sheep. However, the disease can be experimentally transmitted to a wide range of animal species including rodents and nonhuman primates, with variable clinical and pathological manifestations (reviewed in references 47 and 59). In addition, recent epidemiological data suggest that BD may be more widespread in a subclinical form with very long incubation periods (40). Moreover, serological data indicate a possible involvement of BDV, or a related virus, in human mental disorders (1, 6, 7, 31, 62, 70), providing further impetus for the study of this virus, so far unclassified and little understood. Early studies indicated that the infectious agent is a highly neurotropic enveloped virus characterized by noncytolytic replication (23, 27, 28, 37, 47). Using BDV-specific cDNA probes (46, 70), we have demonstrated that BDV has an RNA genome which is associated with the nuclear fraction of the BDV-infected cells (26). Subgenomic poly(A)$^+$ RNAs, complementary to the BDV genome, are also found in the nucleus of the infected cells, from where they are transported to the cytoplasmic compartment (8), suggesting a nuclear localization for the transcription of the BDV genome. Recently, we obtained conclusive evidence that BDV has a nonsegmented, negative, single-stranded (NNS) RNA genome, whose replication and transcription take place in the nucleus, associated with infectious BDV-ribonucleoprotein complexes (BDV-RNPs) (21).

The lack of cell-free virus associated with BDV infection has hampered the cloning and molecular characterization of the BDV genome. However, our finding that relatively high levels of infectious BDV-RNPs are present in the nucleus of infected cells provided us with an alternative source of BDV genomic RNA. Here, we report for the first time the cloning and complete sequence of a full-length BDV genome, as well as a preliminary analysis of the transcriptional capability of the BDV genome. Five major open reading frames (ORFs), designated I, II, III, IV, and V, are predicted in the 8,903-nucleotide BDV genome sequence. ORFs I and II correspond to the previously identified 24- and 40-kDa BDV proteins, respectively (49, 58, 66). ORF III probably encodes the 14.5-kDa protein associated with BDV infection (64). The polypeptides predicted from both ORFs IV and V have not yet been identified in protein samples from BDV-infected material. ORF IV is capable of encoding a predicted polypeptide of approximately 40 kDa, possibly a viral glycoprotein on the basis of sequence features. The largest ORF identified, V, is capable of encoding a predicted protein of approximately 170 kDa, whose deduced amino acid sequence displays strong homology to other NNS-RNA virus polymerases (L-protein family).

We have identified nine species of polyadenylated subgenomic RNAs complementary to the negative-sense RNA genome, including six polycistronic poly(A)$^+$ RNAs and monocistronic mRNAs corresponding to ORFs I, II, and IV, whereas monocistronic poly(A)$^+$ for ORFs III and V were not detected.

Our data indicate that BDV has a genomic organization characteristic of members of the *Mononegavirales* order. Moreover, on the basis of the homology of the BDV L protein to the other NNS-RNA virus polymerases, BDV seems especially closely related to the *Rhabdoviridae* family. However, we have documented that BDV replicates and transcribes in the nucleus of infected cells, suggesting that BDV appears to have an unique molecular biology among NNS-RNA animal viruses.

* Corresponding author. Mailing address: Department of Neuropharmacology, The Scripps Research Institute, 10666 N. Torrey Pines Rd., La Jolla, CA 92037. Phone: (619) 554-9492. Fax: (619) 554-9981.

† Publication no. 8219-NP from the Department of Neuropharmacology, The Scripps Research Institute.

MATERIALS AND METHODS

Cells and virus. The C6 cells persistently infected with BDV (C6BV cells) were generated and grown as described previously (12, 21). Procedures for the preparation of BDV stocks were as described previously (21).

Animal samples. BD rat brain samples were obtained 3 weeks after inoculation of 4- to 6-week-old Lewis rats intracerebrally with 30 μl of BD brain homogenate (3×10^4 50% tissue culture infective doses). (ii) PTI-NB rat brain samples were obtained from BDV-infected newborn rats, which develop a persistent tolerant infection (11).

Detection of viral antigens and quantitation of BDV infectivity. Viral infectivity was determined by using the immunofocus assay (21, 26, 37), and viral antigens were detected by indirect immunofluorescence (21, 26).

Northern blot analysis. RNA from BD and PTI-NB rat brain, as well as from C6 and C6BV cells, was purified by the guanidinium isothiocyanate-acid phenol method (17). Poly(A)$^+$ RNA was purified by two cycles of oligo(dT)-cellulose chromatography as described previously (3). RNA samples were analyzed by Northern (RNA) blot hybridization as described previously (21). cDNA probes were generated by random primer labeling of the DNA (29), and specific-strand riboprobes were prepared as described previously (50). Hybridizations were routinely conducted with probes of specific activity 5×10^8 to 10×10^8 cpm/μg at a concentration of 2 to 5 ng/ml. Washes were done first at low stringency (2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 60°C) followed by high stringency (0.2× SSC at 65°C). The probes were stripped from the blot by boiling three times (15 min each) in 10 mM Tris HCl (pH 8.5)–5 mM EDTA–0.2% sodium dodecyl sulfate. When indicated, RNA hybridization signals were quantitated by densitometric analysis of the Northern blot autoradiograms with an LKB Ultrascan XL laser densitometer.

Isolation of a BDV genome-enriched RNA population. BDV-RNPs were prepared from nuclear fractions of C6BV cells as described previously (15, 21, 55). RNA was extracted from BDV-RNPs and fractionated by oligo(dT)-cellulose chromatography (3, 21), and the poly(A)$^-$ was recovered by ethanol precipitation. Northern blot analysis with BDV-specific probes for the 24- and 40-kDa ORFs (49, 66) revealed a single RNA species of ~9 kb.

Construction of BDV genomic cDNA libraries. First-strand cDNA synthesis was done with 2.5 μg of RNA enriched for BDV genomic RNA as described above and primed with BDV-specific oligonucleotides of antigenomic polarity. Superscript reverse transcriptase (RT) was used to improve the average size of the cDNA synthesis. The final composition of the first-strand reaction mixture was 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, 500 μM each dATP, dCTP, dGTP, and dTTP, 4 μM specific primer, 2.5 μg of RNA, and 5×10^4 U of superscript RT per ml, and the reaction took place at 42°C for 1 h. To generate directional libraries, the BDV-specific oligonucleotides used for first-strand synthesis reactions contained a *Not*I adapter on the 5′ end.

Second-strand synthesis was performed as described by Gubler and Hoffman (34). The products of the second-strand reactions were purified by phenol-chloroform extraction and ethanol precipitation. The ligation reaction mixture for addition of *Sal*I adapters to the cDNA contained 50 mM Tris HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 50% (vol/vol) polyethylene glycol 8000, 1 mM dithiothreitol, 200 μg of *Sal*I adapters per ml, and 100 U of T4 DNA ligase per ml. The reaction was incubated at 16°C for 16 h. After digestion with *Not*I, the cDNA was size fractionated by chromatography through a Sephacryl S-500 HR column. The cDNA was ligated into plasmid pSPORT-1 (GIBCO BRL) previously digested with *Not*I and *Sal*I. The ligated cDNA was used to transform *Escherichia coli* DH10B.

We generated three independent BDV genomic cDNA libraries. The first library was generated by priming the first-strand synthesis with an oligonucleotide corresponding to nucleotides (nt) 489 to 506 of the BDV 24-kDa ORF (66). BDV-specific clones were identified by hybridization with a BDV-specific oligonucleotide sequence located 3′ from the oligonucleotide used to prime the first-strand reaction. Sequence information obtained from the 3′ end of the largest BDV-specific genomic cDNA isolated from the first library was used to design a BDV-specific oligonucleotide with antigenomic polarity that was used to generate the second library and proceed with the BDV genomic walking. A similar strategy was used for the generation of the third library. The sequences of the three different oligonucleotides used to generate the three independent libraries were as follows: first library, 5′ GCATCAACCGCCGTTGGG 3′; second library, 5′ CTC CCTACTCCCAAGGGC 3′; third library, 5′ GGAGCAGCG GACATG 3′. These sequences do not include the *Not*I adapter. BDV-specific clones were identified as described for the first library. Procedures for radiolabeling of oligonucleotides with [^{32}P]ATP and polynucleotide kinase and hybridization conditions were as described previously (21, 63).

Cloning of the 3′ and 5′ ends of the BDV genome. To determine the 3′ end of the BDV genome, purified RNA enriched for BDV genomic RNA was tailed with A residues by using poly(A) polymerase as described previously (33, 65). The 3′ end of the BDV genome was cloned essentially by the procedures described in the 3′ RACE system for rapid amplification of cDNA ends (no. 8373SA; GIBCO BRL). PCR amplification of the BDV 3′ end was performed by using the *Not*I-$(T)_n$ adapter primer and a BDV-specific oligonucleotide with genomic polarity (5′ CGATCCCTCCGTGAACAAACG CAG) complementary to nucleotides 222 to 243 of the BDV 40-kDa ORF (49). The sequence CGAT was incorporated in the 5′ end of the BDV oligonucleotide to facilitate the cloning of the amplified product by generating an *Acc*I and *Not*I cohesive terminus by treatment with T4 DNA polymerase in the presence of dATP and dTTP only. The cDNA was ligated into plasmid pSPORT-1 previously digested with *Acc*I and *Not*I. BDV-specific clones were identified by hybridization with a probe specific for the BDV 40-kDa ORF (49). The 5′ end of the BDV genome was cloned by RNA ligase-mediated self-ligation of the BDV genomic RNA under experimental conditions that promoted intramolecular rather than intermolecular self-ligation, followed by RT-PCR procedures with specific primers designed to amplify the region across the joint of the 5′ and 3′ ends of the BDV genome. Briefly, purified RNA (1 μg) enriched for BDV genomic RNA was self-ligated in a final volume of 10 μl of ligation buffer (75 mM Tris HCl [pH 7.5], 0.1 mM ATP, 10 mM $MgCl_2$, 5 mM dithiothreitol, 10% [vol/vol] dimethyl sulfoxide, 50 U of RNasin) with 50 U of RNase-free RNA ligase (Pharmacia LKB Biotechnology, Piscataway, N.J.). The reaction was incubated at 25°C for 6 h, and the product was purified by phenol-chloroform extraction and ethanol precipitation. First-strand cDNA synthesis was primed with a BDV-specific oligonucleotide, 5′ CGATCGGTCAGC CCGTCCTAGTGAG 3′, which was designed on the basis of sequence information obtained from the 3′ end of clone A isolated from the third BDV genomic library. PCR amplification was performed by using as a second BDV-specific primer

the same one described for the 3′ RACE procedure, complementary to nt 222 to 243 of the BDV 40-kDa ORF (49). As described above, the sequence CGAT on the 5′ end of the oligonucleotides was added to facilitate cloning of the amplified cDNA. Procedures for cDNA synthesis, RT-PCR amplification, and cloning were as described in Materials and Methods.

cDNA sequencing. DNA sequence determination was performed by using the Sequenase version 2.0 cDNA sequencing kit (United States Biochemicals, Cleveland, Ohio), as well as the PRISM Ready Reaction Dyedeoxy Terminator Cycle sequencing kit with the Applied Biosystems model 373A DNA sequencing system. The entire genome was sequenced in both directions, and each nucleotide position was sequenced at least twice in each strand.

Analysis of nucleotide and amino acid sequences. Screening of the Swiss-Prot 26, GenBank, and EMBL protein data bases for similarities with the predicted amino acid sequences from BDV ORFs was conducted by using the FASTDB (fast pairwise comparison of sequences) program (10) (IntelliGenetics, Inc.). Search of motifs present in the BDV ORFs was conducted by using the QUEST (quick user-directed expression search tool) program (IntelliGenetics, Inc.).

Alignment of the L proteins was generated by the program MULTALIN (19) by using the cluster principle described previously (56).

Nucleotide sequence accession number. The sequence of the BDV genome has been deposited with GenBank under accession number L27077.

RESULTS

Cloning and sequencing of the BDV genome. The strategy followed to clone the BDV genome is illustrated in Fig. 1. RNA enriched for BDV genomic RNA was obtained from infectious BDV-RNPs present in the nucleus of C6BV cells (21) and used to generate BDV cDNA genomic libraries, which allowed us to conduct a BDV genomic walking. The first library was generated by using an oligonucleotide whose sequence corresponded to nt 489 to 506 of the previously described BDV 24-kDa ORF (66). BDV-specific clones were identified as described in Materials and Methods, and those bearing cDNA inserts of 2.5 kb or more were further analyzed by standard procedures, including sequencing of their 5′ and 3′ ends. In addition to isolating clones in which the 5′ end corresponded to the expected nt 489 to 506 of the BDV 24-kDa ORF (66), we recovered BDV-specific clones that initiated at nt 69 of the BDV 40-kDa ORF (49). Representative examples of each type of clone, C90 and C38, respectively, are shown in Fig. 1. The similarity between the sequences 5′ CGCCGTTGGG 3′ and 5′ CACCGTTGGG 3′, present in BDV 24- and 40-kDa ORFs, respectively, facilitated the ability of the oligonucleotide 5′ GCATCAACCGCCGTTGGG 3′, used to prime first-strand cDNA synthesis, to cross-hybridize with both sequences, accounting for the isolation of both types of BDV genomic clones. A second BDV cDNA genomic library was generated by priming first-strand cDNA synthesis with a BDV-specific oligonucleotide designed on the basis of the sequence information obtained from the 3′ end of clone C90. BDV-specific clones bearing cDNA inserts of 2.5 kb or more were identified and analyzed as described above. Clone C8 (Fig. 1) is a representative clone of this library. A similar approach was followed to generate a third library. In this third library, we did not detect BDV-specific clones bearing inserts bigger than 1.5 kb, which suggested that we had reached the 5′ end of the BDV genome. Clone CA (Fig. 1) is a representative clone of the third library.

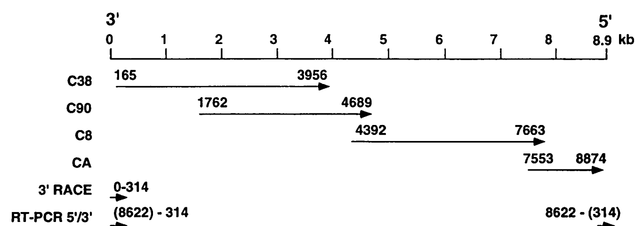

FIG. 1. Strategy for cloning the BDV genome. The BDV genome, 8,903 nt long, is presented on top (negative sense). Representative clones from the three BDV genomic libraries generated (see Materials and Methods) are indicated: clones C38 and C90 are representatives from the first library; clone C8 is from the second; and clone CA is from the third. 3′ RACE is a representative clone of those obtained by 3′ RACE procedures with poly(A) tailed BDV genomic RNA to determine the 3′ end of the BDV genome (see Materials and Methods). Clone RT-PCR 5′/3′ contains the last 298 nt and the first 335 nt of the 5′ and 3′ ends, respectively, of the BDV genome, being a representative clone of those obtained by RT-PCR procedures with self-ligated BDV genomic RNA (see Materials and Methods). Numbers indicate the nucleotide positions in the BDV genome (Fig. 2), delimiting the area covered by each clone.

The 3′ end of the BDV genome was cloned by 3′ RACE procedures (see Materials and Methods). Six independent clones were sequenced to determine the sequence between nt 222 of the BDV 40-kDa ORF (49) and the 3′ end of the BDV genome. Finally, the 5′ end of the BDV genome and its precise 3′-end sequence (without the ambiguity of extra adenosine residues at the 3′ end introduced during the 3′ RACE procedures) were determined by sequencing across the 5′-3′ junction of the BDV genome created by RNA ligase-mediated self-ligation of the 5′ and 3′ ends of the BDV genomic RNA. The use of experimental conditions that promoted intramolecular rather than intermolecular self-ligation was followed by RT-PCR procedures to amplify the region in the circularized BDV genomic RNA flanked by sequences corresponding to BDV 40-kDa ORF and clone CA. PCR amplification did not require treatment of the RNA with tobacco acid pyrophosphatase prior to RNA-ligase-mediated RNA circularization.

FIG. 2. Sequence of the BDV genome. The complete sequence of the BDV genome is presented in the antigenomic polarity (positive sense). Numbers at left represent nucleotide positions in the BDV genome starting at the 5′ end of the anti-genomic BDV RNA. Roman numerals at the right (I, II, III, IV, and V), as well as the symbols x1 and x2, refer to the BDV ORFs. Arabic numerals at the right indicate amino acid positions in the corresponding predicted protein sequence for each BDV ORF. ATGs that are most likely candidates to be utilized as the start codon for the different BDV ORFs are typed in boldface and underlined. The potential transcription start and termination-polyadenylation signals flanking each ORF have been underlined: single line for the start signals, and double line for the stop signals. Additional putative termination-polyadenylation signals located within ORF V are also indicated. The amino acids in BDV ORF V participating in the four conserved motifs within the proposed catalytic domain of the L-polymerases are typed in boldface. Amino acids 326 to 338 in ORF V (underlined) correspond to the putative template-binding site described in the L-polymerases (56). The carboxy half of ORF V contains three RGD motifs (underlined) clustered in a region 320 amino acids long, whose biological significance is presently unknown.

```
   1 GTTGCGTTAACAACCAACCAATCATTATCCTTCTAATAAAATGAACACACGCAATGCCACCCAAGAGACGCCTGGTTGATGACGCCGATG
                                                           M  P  P  K  R  R  L  V  D  D  A  D  12-I
  91 CCATGGAGGACCAAGATTTATATGAACCCCCAGCGAGCCTCCCCAAGCTCCCCGGAAAATTCCTACAATACACCGTTGGGGGGTCTGACC
      A  M  E  D  Q  D  L  Y  E  P  P  A  S  L  P  K  L  P  G  K  F  L  Q  Y  T  V  G  G  S  D  42-I
 181 CGCATCCGGGTATAGGGCATGAGAAGGATATCAGGCAGAACGCAGTGGCATTGTTAGACCAGTCACGGCGCGATATGTTTCATACAGTAA
      P  H  P  G  I  G  H  E  K  D  I  R  Q  N  A  V  A  L  L  D  Q  S  R  R  D  M  F  H  T  V  72-I
 271 CGCCCAGCCTTGTGTTTCTATGTTTGCTAATCCCAGGACTGCACGCTGCGTTTGTTCACGGAGGGGTGCCTCGTGAATCTTACCTGTCGA
      T  P  S  L  V  F  L  C  L  L  I  P  G  L  H  A  A  F  V  H  G  G  V  P  R  E  S  Y  L  S  102-I
 361 CGCCTGTTACGCGTGGGGAACAGACTGTCGTTAAGACTGCAAAGTTTTACGGGGAAAAGACAACACAGCGTGATCTCACCGAGCTGGAGA
      T  P  V  T  R  G  E  Q  T  V  V  K  T  A  K  F  Y  G  E  K  T  T  Q  R  D  L  T  E  L  E  132-I
 451 TCTCCTCTATATTCAGCCATTGTTGCTCATTACTAATTGGGGTTGTGATAGGATCGTCATCTAAGATTAAAGCAGGAGCCGAGCAGATCA
      I  S  S  I  F  S  H  C  C  S  L  L  I  G  V  V  I  G  S  S  S  K  I  K  A  G  A  E  Q  I  162-I
 541 AGAAAAGGTTTAAAACTATGATGGCAGCCTTAAACCGGCCATCCCATGGTGAGACTGCTACACTACTTCAGATGTTTAATCCACATGAGG
      K  K  R  F  K  T  M  M  A  A  L  N  R  P  S  H  G  E  T  A  T  L  L  Q  M  F  N  P  H  E  192-I
 631 CTATAGATTGGATTAACGGCCAGCCCTGGGTAGGCTCCTTTGTGTTGTCTCTACTAACTACAGACTTTGAGTCCCCAGGTAAAGAATTCA
      A  I  D  W  I  N  G  Q  P  W  V  G  S  F  V  L  S  L  L  T  T  D  F  E  S  P  G  K  E  F  222-I
 721 TGGATCAGATTAAACTTGTCGCAAGTTATGCGCAGATGACTACGTACACTACTATAAAGGAGTACCTCGCAGAATGTATGGATGCTACCC
      M  D  Q  I  K  L  V  A  S  Y  A  Q  M  T  T  Y  T  T  I  K  E  Y  L  A  E  C  M  D  A  T  252-I
 811 TTACAATCCCTGTAGTTGCATATGAGATTCGTGACTTTTTAGAAGTTTCAGCAAAGCTTAAAGAGGAACATGCTGACCTGTTTCCGTTCC
      L  T  I  P  V  V  A  Y  E  I  R  D  F  L  E  V  S  A  K  L  K  E  E  H  A  D  L  F  P  F  282-I
 901 TGGGGGCTATTCGGCACCCCGACGCTATCAAGCTTGCGCCACGGAGCTTTCCCAATCTGGCTTCTGCAGCGTTTTACTGGAGTAAGAAGG
      L  G  A  I  R  H  P  D  A  I  K  L  A  P  R  S  F  P  N  L  A  S  A  A  F  Y  W  S  K  K  312-I
 991 AGAATCCCACAATGGCGGGCTACCGGGCCTCCACCATCCAGCCGGGCGCGAGTGTCAAGGAGACCCAGCTTGCCCGGTATAGGCGCCGCG
      E  N  P  T  M  A  G  Y  R  A  S  T  I  Q  P  G  A  S  V  K  E  T  Q  L  A  R  Y  R  R  R  342-I
1081 AGATATCTCGCGGGGAAGACGGGGCAGAGCTCTCAGGTGAGATCTCTGCCATAATGAGAATGATAGGTGTGACTGGTCTAAACTAGAAAA
      E  I  S  R  G  E  D  G  A  E  L  S  G  E  I  S  A  I  M  R  M  I  G  V  T  G  L  N  *     370-I
1171 CAATGAACAAACCAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTGATGAGTTCCGACCTCCGGCTGACATTGCTTGAATTAGT
                                                          M  S  S  D  L  R  L  T  L  L  E  L  V 13-x1
1261 CAGGAGGCTCAATGGCAACGGGACCATCGAGTCTGGTCGACTCCCTGGAGGACGAAGAAGATCCCCAGACACTACGACGGGAACGATCGG
       R  R  L  N  G  N  G  T  I  E  S  G  R  L  P  G  G  R  R  R  S  P  D  T  T  T  G  T  I  G 43-x1
                 M  A  T  G  P  S  S  L  V  D  S  L  E  D  E  E  D  P  Q  T  L  R  R  E  R  S  26-II
1351 GGTCACCAAGACCACGGAAGATCCCAAGGAATGCATTGACCCAACCGGTAGACCAGCTCCTGAAGGACCTCAGGAAGAACCCCTCCATGA
       V  T  K  T  T  E  D  P  K  E  C  I  D  P  T  G  R  P  A  P  E  G  P  Q  E  E  P  L  H  D 73-x1
      G  S  P  R  P  R  K  I  P  R  N  A  L  T  Q  P  V  D  Q  L  L  K  D  L  R  K  N  P  S  M  56-II
1441 TCTCAGACCCAGACCAGCGAACCGGAAGGGAGCAGCTATCGAATGATGAGCTTATCAAGAAGCTAGTGACGGAGCTGGCCGAGAATAGCA
       L  R  P  R  P  A  N  R  K  G  A  A  I  E  *                                              87-x1
      I  S  D  P  D  Q  R  T  G  R  E  Q  L  S  N  D  E  L  I  K  K  L  V  T  E  L  A  E  N  S  86-II
1531 TGATCGAGGCTGAGGAGGTGCGGGGCACTCTTGGGGACATCTCGGCTCGCATCGAGGCAGGGTTTGAGTCCCTGTCCGCCCTCCAAGTGG
      M  I  E  A  E  E  V  R  G  T  L  G  D  I  S  A  R  I  E  A  G  F  E  S  L  S  A  L  Q  V  116-II
1621 AAACCATCCAGACAGCTCAGCGGTGCGACCACTCCGATAGCATCAGAATCCTTGGCGAGAACATCAAGATACTGGATCGCTCCATGAAGA
      E  T  I  Q  T  A  Q  R  C  D  H  S  D  S  I  R  I  L  G  E  N  I  K  I  L  D  R  S  M  K  146-II
1711 CAATGATGGAGACAATGAAGCTCATGATGGAGAAGGTGGACCTCCTCTACGCATCAACCGCCGTTGGGACCTCTGCACCCATGTTGCCCT
      T  M  M  E  T  M  K  L  M  M  E  K  V  D  L  L  Y  A  S  T  A  V  G  T  S  A  P  M  L  P  176-II
1801 CCCATCCTGCACCTCCGCGCATTTATCCCCAGCTCCCAAGTGCCCCGACAGCGGATGAGTGGGACATCATACCATAAAAAAATCGAATCA
      S  H  P  A  P  P  R  I  Y  P  Q  L  P  S  A  P  T  A  D  E  W  D  I  I  P  *              201-II
1891 CCATGAATTCAAAGCATTCCTATGTGGAGCTCAAGGACAAGGTAATCGTCCCTGGATGGCCCACACTGATGCTTGAGATAGACTTTGTAG
        M  N  S  K  H  S  Y  V  E  L  K  D  K  V  I  V  P  G  W  P  T  L  M  L  E  I  D  F  V  29-III
1981 GAGGGACTTCACGGAACCAGTTCCTTAACATCCCATTTCTTTCAGTGAAAGAGCCTCTGCAGCTTCCACGCGAGAAGAAGTTGACCGACT
      G  G  T  S  R  N  Q  F  L  N  I  P  F  L  S  V  K  E  P  L  Q  L  P  R  E  K  K  L  T  D  59-III
2017 ACTTCACCATTGACGTAGAGCCAGCAGGTCATTCCCTGGTCAACATATACTTCCAGATTGACGACTTCTTGCTCCTAACACTCAACTCAC
      Y  F  T  I  D  V  E  P  A  G  H  S  L  V  N  I  Y  F  Q  I  D  D  F  L  L  L  T  L  N  S  89-III
2161 TGTCCGTATACAAGGACCCGATTAGGAAATACATGTTCCTACGCCTCAACAAGGAACAGAGCAAGCACGCAATTAATGCAGCTTTCAATG
      L  S  V  Y  K  D  P  I  R  K  Y  M  F  L  R  L  N  K  E  Q  S  K  H  A  I  N  A  A  F  N  119-III
                                                                                 M  Q  L  S  M  5-x2
2251 TCTTCTCTTATCGGCTTCGGAACATTGGTGTTGGCCCTCTCGGCCCAGACATTCGATCTTCAGGGCCTTAGTTGCAATACTGACTCCACT
      V  F  S  Y  R  L  R  N  I  G  V  G  P  L  G  P  D  I  R  S  S  G  P  *                    142-III
       S  S  L  I  G  F  G  T  L  V  L  A  L  S  A  Q  T  F  D  L  Q  G  L  S  C  N  T  D  S  T 35-x2
2341 CCTGGATTAATCGATCTGGAGATAAGCGACTTTGCCACACCCCAACGGAAAATGTCATTTCATGCGAGGTTAGTTATCTTAACCACACGA
       P  G  L  I  D  L  E  I  S  D  F  A  T  P  Q  R  K  M  S  F  H  A  R  L  V  I  L  T  T  R 65-x2
2431 CTATTAGCCTCCCGGCAGTCCACACGTCATGCCTCAAGTACCACTGCAAAACCTATTGGGGATTCTTTGGTAGCTACAGCGCTGACCGAA
       L  L  A  S  R  Q  S  T  R  H  A  S  S  T  T  A  K  P  I  G  D  S  L  V  A  T  A  L  T  E 95-x2
2521 TCATCAATCGGTACACTGGTACTGTTAAGGGTTGTTTAAACAACTCAGCGCCAGAGGATCCCTTCGAGTGCAACTGGTTCTACTGCTGCT
       S  S  I  G  T  L  V  L  L  R  V  V  *                                                    107-x2
2611 CGGCGATTACAACAGAGATCTGCCGATGCTCTATTACAAATGTCACGGTGGCTGTACAGACATTCCCACCGTTCATGTACTGCAGTTTCG
                                                                                M  Y  C  S  F   5-IV
2701 CGGACTGTAGTACTGTGAGTCAGCAGGAGCTAGAGAGTGGCAAGGCAATGCTGAGCGATGGCAGTACCTTAACTTATACCCCGTATATCT
      A  D  C  S  T  V  S  Q  Q  E  L  E  S  G  K  A  M  L  S  D  G  S  T  L  T  Y  T  P  Y  I  35-IV
2791 TACAATCAGAAGTCGTGAACAAAACCCTTAATGGGACTATACTCTGCAACTCATCCTCCAAGATAGTTTCCTTCGATGAATTTAGGCGTT
      L  Q  S  E  V  V  N  K  T  L  N  G  T  I  L  C  N  S  S  S  K  I  V  S  F  D  E  F  R  R  65-IV
```

```
2881 CATACTCCCTAGCGAATGGTAGTTACCAGAGCTCATCAATCAATGTGACGTGTGTAAACTACACGTCGTCCTGCCGGTCCAAGTTGAGAA
      S  Y  S  L  A  N  G  S  Y  Q  S  S  S  I  N  V  T  C  V  N  Y  T  S  S  C  R  S  K  L  R   95-IV 2971
GGCGGCGTAGGGATACTCAACAGATTGAGTACCTAGTTCACAAGCTTAGGCCTACACTGAAAGATGCGTGGGAGGACTGTGAGATCCTCC
      R  R  R  R  D  T  Q  Q  I  E  Y  L  V  H  K  L  R  P  T  L  K  D  A  W  E  D  C  E  I  L   125-IV
3061 AGTCTCTGCTCCTAGGGATGTTTGGTACTGGGATTGCAAGTGCTTCGCAATTCTTGAGGGGCTGGCTCAACCACCCTGATATCATCGGGT
      Q  S  L  L  L  G  M  F  G  T  G  I  A  S  A  S  Q  F  L  R  G  W  L  N  H  P  D  I  I  G   155-IV
3151 ATATAGTTAATGGAGTTGGGGTAGTCTGGCAATGCCATCGTGTTAATGTCACGTTCATGGCGTGGAATGAGTCCACATATTACCCTCCAG
      Y  I  V  N  G  V  G  V  V  W  Q  C  H  R  V  N  V  T  F  M  A  W  N  E  S  T  Y  Y  P  P   185-IV
3241 TAGATTACAATGGACGGAAGTACTTTCTGAATGATGAGGGGAGGCTACAAACAAACACCCCCGAGGCAAGGCCAGGGCTTAAGCGGGTCA
      V  D  Y  N  G  R  K  Y  F  L  N  D  E  G  R  L  Q  T  N  T  P  E  A  R  P  G  L  K  R  V   215-IV
3331 TGTGGTTCGGCAGGTACTTCCTAGGGACAGTAGGGTCTGGGGTGAAACCGAGGAGGATTCGGTACAATAAGACCTCACATGATTACCATC
      M  W  F  G  R  Y  F  L  G  T  V  G  S  G  V  K  P  R  R  I  R  Y  N  K  T  S  H  D  Y  H   245-IV
3421 TAGAGGAGTTTGAGGCAAGTCTCAACATGACCCCCCAGACCAGTATCGCCTCGGGTCATGAGACAGACCCCATAAATCATGCCTACGGAA
      L  E  E  F  E  A  S  L  N  M  T  P  Q  T  S  I  A  S  G  H  E  T  D  P  I  N  H  A  Y  G   275-IV
3511 CGCAGGCTGACCTCCTTCCATACACCAGGTCTAGTAATATAACGTCTACAGATACAGGCTCAGGCTGGGTGCACATCGGCCTACCCTCAT
      T  Q  A  D  L  L  P  Y  T  R  S  S  N  I  T  S  T  D  T  G  S  G  W  V  H  I  G  L  P  S   305-IV
3601 TTGCTTTCCTCAATCCTCTCGGGTGGCTTAGGGACCTACTTGCGTGGGCGGCCTGGTTGGGTGGGGTTCTATACTTAATAAGTCTTTGTG
      F  A  F  L  N  P  L  G  W  L  R  D  L  L  A  W  A  A  W  L  G  G  V  L  Y  L  I  S  L  C   335-IV
3691 TTTCCTTACCAGCCTCCTTCGCGAGGAGGAGACGCCTCGGCCGGTGGCAGGAATAAACCGTACCGACCAAACTCTTAAAAACCCTCTTCT
      V  S  L  P  A  S  F  A  R  R  R  R  L  G  R  W  Q  E  *                                    353-IV
3781 CGGGACAGAGGTCTCTTTCTGCCTTAAATCGAGTTCACTCCCCCATCACGTACGAGCATTGGGCCAGATTAAAGCAAAGAACCTGGCATC
3871 CTGTGACTATTACTTGCTATTCCGCCAAGTTGTATTGCCCCCTGAAGTATATCCCATTGGTGTCTTAATAAGAGCTGCGGAGGCCATACT
3961 AACAGTTATAGTATCAGCTTGGAAGCTGGATCACATGACAAAGACCCTATACTCCTCTGTGAGATATGCACTCACCAATCCCCGGGTCCG
4051 GGCCCAACTTGAGCTCCACATTGCCTACCAGCGCATAGTGGTCAGTCTCGTATAGCCGGGAAGCAGATATAGGGCCAAAAAGGCTTGGGA
4141 ATATGTCATTGCAATTCATCCAATCCCTCGTTATTGCCACCATAGACACAACGAGCTGCCTAATGACCTACAACCACTTTCTTGCTGCAG
      M  S  L  Q  F  I  Q  S  L  V  I  A  T  I  D  T  T  S  C  L  M  T  Y  N  H  F  L  A  A   29-V
4231 CAGACACAGCCAAGAGCAGATGCCACCTCCTAATCGCCTCAGTGGTCCAAGGAGCCCTTTGGGAGCAAGGGTCATTTCTTGATCATATAA
      A  D  T  A  K  S  R  C  H  L  L  I  A  S  V  V  Q  G  A  L  W  E  Q  G  S  F  L  D  H  I   59-V
4321 TCAACATGATCGACACAATTGACTCAATCAACCTCCCCCATGATGATTACTTCACAATTATTAAGTCTATCTCTCCCTACTCCCAAGGGC
      I  N  M  I  D  T  I  D  S  I  N  L  P  H  D  D  Y  F  T  I  I  K  S  I  S  P  Y  S  Q  G   89-V
4411 TTGTTATGGGGAGGCACAATGTGTCAGTCTCCTCTGATTTTGCGTCCGTATTTACTATTCCTGAATCATGCCCACAACTAGACAGCTTAC
      L  V  M  G  R  H  N  V  S  V  S  S  D  F  A  S  V  F  T  I  P  E  S  C  P  Q  L  D  S  L   119-V
4501 TAAAAAAACTGCTTCAACTTGACCCTGTTCTCCTCCTCATGGTCTCTTCGGTGCAGAAGTCATGGTACTTCCCTGAGATCCGAATGGTTG
      L  K  K  L  L  Q  L  D  P  V  L  L  L  M  V  S  S  V  Q  K  S  W  Y  F  P  E  I  R  M  V   149-V
4591 ACGGGTCACGGGAGCAGCTCCACAAGATGCGTGTCGAGCTGGAGACGCCCCAAGCCCTGCTGTCATACGGCCATACCCTCCTGTCAATAT
      D  G  S  R  E  Q  L  H  K  M  R  V  E  L  E  T  P  Q  A  L  L  S  Y  G  H  T  L  L  S  I   179-V
4681 TTCGAGCAGAGTTTATCAAAGGCTATGTCTCAAAGAATGCGAAGTGGCCGCCTGTACACCTGCTCCCAGGCTGTGACAAATCCATAAAGA
      F  R  A  E  F  I  K  G  Y  V  S  K  N  A  K  W  P  P  V  H  L  L  P  G  C  D  K  S  I  K   209-V
4771 ATGCGAGAGAGCTGGGCCGCTGGAGCCCGGTGTTTGACCGACGATGGCAGCTCTTCGCGAAGGTTGTCATTCTAAGAATTGCTGACCTAG
      N  A  R  E  L  G  R  W  S  P  V  F  D  R  R  W  Q  L  F  A  K  V  V  I  L  R  I  A  D  L   239-V
4861 ATATGGATCCCGACTTCAACGATATTGTTAGCGACAAGGCGATAATCAGCTCAAGAAGGGACTGGGTATTTGAGTACAATGCAGCAGCCT
      D  M  D  P  D  F  N  D  I  V  S  D  K  A  I  I  S  S  R  R  D  W  V  F  E  Y  N  A  A  A   269-V
4951 TTTGGAAGAAATACAGTGAGCGGTTGGAGAGGCCCCCTGCCAGATCGGGACCATCACGGCTTGTGAATGCTCTGATCGATGGACGCTTAG
      F  W  K  K  Y  S  E  R  L  E  R  P  P  A  R  S  G  P  S  R  L  V  N  A  L  I  D  G  R  L   299-V
5041 ATAATATCCCAGCCCTGCTAGAGCCATTTTACAGGGGAGCGGTTGAGTTTGAGGATCGGCTGACTGTGCTCGTGCCTAAGGAGAAGGAGT
      D  N  I  P  A  L  L  E  P  F  Y  R  G  A  V  E  F  E  D  R  L  T  V  L  V  P  K  E  K  E   329-V
5131 TGAAGGTAAAGGGAAGGTTCTTCTCGAAGCAAACATTGGCAATCAGGATATATCAGGTTGTTGCTGAAGCTGCACTTAAGAACGAGGTTA
      L  K  V  K  G  R  F  F  S  K  Q  T  L  A  I  R  I  Y  Q  V  V  A  E  A  A  L  K  N  E  V   359-V
5221 TGCCATACTTAAAAACACATTCAATGACCATGAGCTCAACGGCCCTAACCCATCTTCTTAACCGGCTATCACATACTATCACTAAGGGTG
      M  P  Y  L  K  T  H  S  M  T  M  S  S  T  A  L  T  H  L  L  N  R  L  S  H  T  I  T  K  G   389-V
5311 ACTCCTTTGTTATTAACTTAGATTATAGCTCCTGGTGCAACGGTTTCCGACCAGAACTACAAGCCCCACTCTGTCGTCAGTTGGATCAGA
      D  S  F  V  I  N  L  D  Y  S  S  W  C  N  G  F  R  P  E  L  Q  A  P  L  C  R  Q  L  D  Q   419-V
5401 TGTTCAATTGCGGGTACTTCTTCAGGACTGGGTGCACACTGCCATGCTTTACCACGTTTATTATTCAGGACAGATTCAACCCGCCCTATT
      M  F  N  C  G  Y  F  F  R  T  G  C  T  L  P  C  F  T  T  F  I  I  Q  D  R  F  N  P  P  Y   449-V
5491 CCTTCCGTGGTGAGCCCGTTGAAGACGGTGTCACATGCGCGGTTGGGACTAAGACAATGGGAGAGGGTATGAGGCAGAAACTATGGACAA
      S  F  R  G  E  P  V  E  D  G  V  T  C  A  V  G  T  K  T  M  G  E  G  M  R  Q  K  L  W  T   479-V
5581 TTCTTACGAGCTGCTGGGAGATAATTGCTCTTCGGGAAATTAACGTGACGTTTAATATACTAGGCCAGGGTGATAATCAGACAATCATTG
      I  L  T  S  C  W  E  I  I  A  L  R  E  I  N  V  T  F  N  I  L  G  Q  G  D  N  Q  T  I  I   509-V
5671 TACATAAATCTGCAAGCCAAAATAATCAGCTATTAGCGGAGCGAGCATTGGGAGCTTTGTACAAGCATGCTAGATTAGCTGGCCATAACC
      V  H  K  S  A  S  Q  N  N  Q  L  L  A  E  R  A  L  G  A  L  Y  K  H  A  R  L  A  G  H  N   539-V
5761 TTAAGGTAGAAGAATGTTGGGTGTCAGATTGTCTGTATGAGTATGGAAAGAAGCTCTTCTTCCGTGGTGTACCTGTCCCAGGCTGTTTGA
      L  K  V  E  E  C  W  V  S  D  C  L  Y  E  Y  G  K  K  L  F  F  R  G  V  P  V  P  G  C  L   569-V
5851 AGCAGCTCTCGCGGGTGACGGACTCCACTGGGGAGTTATTCCCAAACCTATACTCAAAGTTAGCCTGCTTAACATCATCATGCTTAAGCG
      K  Q  L  S  R  V  T  D  S  T  G  E  L  F  P  N  L  Y  S  K  L  A  C  L  T  S  S  C  L  S   599-V
5941 CAGCGATGGCAGACACATCCCCATGGGTGGCACTCGCGACAGGTGTCTGTCTGTATCTTATCGAGTTGTATGTTGAGCTGCCTCCGGCAA
      A  A  M  A  D  T  S  P  W  V  A  L  A  T  G  V  C  L  Y  L  I  E  L  Y  V  E  L  P  P  A   629-V
6031 TCATGCAGGACGAGTCGCTGTTAACGACCCTCTGTCTCGTAGGTCCATCCATTGGTGGGCTTCCAACTCCTGCAACCCTGCCCAGTGTCT
      I  M  Q  D  E  S  L  L  T  T  L  C  L  V  G  P  S  I  G  G  L  P  T  P  A  T  L  P  S  V   659-V
```

FIG. 2—*Continued.*

```
6121 TTTTCAGAGGAATGTCCGACCCATTGCCCTTTCAGCTAGCACTCTTGCAGACCCTCATTAAAACGACAGGGGTGACTTGTAGCTTGGTGA
      F  F  R  G  M  S  D  P  L  P  F  Q  L  A  L  L  Q  T  L  I  K  T  T  G  V  T  C  S  L  V    689-V
6211 ATCGTGTGGTTAAGTTACGGATAGCACCCTATCCAGACTGGCTCTCCCTAGTGACTGACCCGACTTCACTCAACATTGCTCAGGTGTACC
      N  R  V  V  K  L  R  I  A  P  Y  P  D  W  L  S  L  V  T  D  P  T  S  L  N  I  A  Q  V  Y    719-V
6301 GGCCAGAACGTCAAATCAGGAGGTGGATTGAGGAGGCAATAGCAACAAGCTCACACTCGTCACGCATAGCAACTTTTTTCCAGCAGGCCC
      R  P  E  R  Q  I  R  R  W  I  E  E  A  I  A  T  S  S  H  S  S  R  I  A  T  F  F  Q  Q  A    749-V
6391 TCACGGAGATGGCCCAGCTGCTTGCGAGGGACCTCTCAACAATGATGCCTCTTCGGCCCCGGGATATGTCGGCCTTATTCGCATTATCAA
      L  T  E  M  A  Q  L  L  A  R  D  L  S  T  M  M  P  L  R  P  R  D  M  S  A  L  F  A  L  S    779-V
6481 ATGTCGCATATGGTCTAAGCATTATAGATCTATTTCAAAAGTCCTCTACCGTTGTCTCTGCAAGTCAAGCTGTCCATATCGAAGATGTTG
      N  V  A  Y  G  L  S  I  I  D  L  F  Q  K  S  S  T  V  V  S  A  S  Q  A  V  H  I  E  D  V    809-V
6571 CCCTAGAGAGTGTAAGGTATAAGGAATCTATCATTCAGGGTCTGTTAGACACTACTGAGGGGTACAACATGCAACCTTATTTGGAAGGTT
      A  L  E  S  V  R  Y  K  E  S  I  I  Q  G  L  L  D  T  T  E  G  Y  N  M  Q  P  Y  L  E  G    839-V
6661 GCACTTACCTTGCAGCCAAGCAGCTACGGAGGTTGACGTGGGGTCGAGACCTAGTTGGAGTTACAATGCCGTTTGTTGCCGAGCAATTCC
      C  T  Y  L  A  A  K  Q  L  R  R  L  T  W  G  R  D  L  V  G  V  T  M  P  F  V  A  E  Q  F    869-V
6751 ATCCCCATAGTTCTGTCGGTGCAAAAGCAGAACTCTACCTCGATGCTATCATATACTGCCCACAAGAGACGTTGCGGTCACACCATCTGA
      H  P  H  S  S  V  G  A  K  A  E  L  Y  L  D  A  I  I  Y  C  P  Q  E  T  L  R  S  H  H  L    899-V
6841 CTACCAGGGGGGACCAGCCGCTTTACCTTGGATCTAATACGGCTGTCACGGTTCAGCGAGGTGAGATCACAGGCCTAACAAAGTCAAGGG
      T  T  R  G  D  Q  P  L  Y  L  G  S  N  T  A  V  T  V  Q  R  G  E  I  T  G  L  T  K  S  R    929-V
6931 CTGCAAATCTAGTCAAGGACACTCTCGTTCTCCACCAGTGGTACAAGGTCCGTAAGGTTACCGATCCACACTTGAACACTCTCATGGCGC
      A  A  N  L  V  K  D  T  L  V  L  H  Q  W  Y  K  V  R  K  V  T  D  P  H  L  N  T  L  M  A    959-V
7021 GCTTCTTGCTTGAGAAGGGATACACATCTGACGCTCGGCCTAGCATTCAGGGTGGGACCCTCACACATCGTCTCCCATCCCGTGGAGACT
      R  F  L  L  E  K  G  Y  T  S  D  A  R  P  S  I  Q  G  G  T  L  T  H  R  L  P  S  R  G  D    989-V
7111 CACGCCAAGGGCTCACTGGGTATGTGAATATACTCAGCACGTGGCTCCGGTTCTCAAGTGATTATCTTCACTCTTTCTCGAAATCATCAG
      S  R  Q  G  L  T  G  Y  V  N  I  L  S  T  W  L  R  F  S  S  D  Y  L  H  S  F  S  K  S  S    1091-V
7201 ATGACTACACAATCCACTTCCAGCATGTATTCACATACGGTTGCCTCTATGCTGATTCGGTGATTAGATCGGGCGGTGTTATTTCCACTC
      D  D  Y  T  I  H  F  Q  H  V  F  T  Y  G  C  L  Y  A  D  S  V  I  R  S  G  G  V  I  S  T    1049-V
7291 CTTACCTTTTGAGTGCAAGTTGTAAAACATGCTTTGAGAAGATAGACTCAGAGGAGGTCGTCCTGGCATGCGAACCTCAATATAGGGGTG
      P  Y  L  L  S  A  S  C  K  T  C  F  E  K  I  D  S  E  E  V  V  L  A  C  E  P  Q  Y  R  G    1079-V
7381 CTGAGTGGCTGATATCAAAGCCAGTTACTGTCCCTGAGCAGATAATTGACGCTGAAGTCGAGTTTGACCCCTGTGTGAGTGCGAGTTATT
      A  E  W  L  I  S  K  P  V  T  V  P  E  Q  I  I  D  A  E  V  E  F  D  P  C  V  S  A  S  Y    1109-V
7471 GTCTCGGGATTCTCATTGGCAAGTCATTCTTGGTTGACATAAGGGCAAGTGGGCATGATATTATGGAGCAGCGGACATGGGCTAACTTGG
      C  L  G  I  L  I  G  K  S  F  L  V  D  I  R  A  S  G  H  D  I  M  E  Q  R  T  W  A  N  L    1139-V
7561 AGAGGTTTTCTGTGTCGGACATGCAGAAACTTCCATGGAGTATTGTAATTCGGTCTCTCTGGAGATTCCTTATTGGCGCACGACTCCTCC
      E  R  F  S  V  S  D  M  Q  K  L  P  W  S  I  V  I  R  S  L  W  R  F  L  I  G  A  R  L  L    1169-V
7651 AGTTTGAGAAGGCTGGCCTTATTAGGATGCTGTATGCTGCAACAGGTCCAACCTTTAGCTTCCTAATGAAAGTCTTTCAAGACTCAGCCC
      Q  F  E  K  A  G  L  I  R  M  L  Y  A  A  T  G  P  T  F  S  F  L  M  K  V  F  Q  D  S  A    1199-V
7741 TACTTATGGACTGCGCACCTCTTGATCGGCTGTACCCTAGGATCAACTTTCATAGTCGGGGAGACCTCGTCGCCAAGCTCGTTTTATTAC
      L  L  M  D  C  A  P  L  D  R  L  Y  P  R  I  N  F  H  S  R  G  D  L  V  A  K  L  V  L  L    1229-V
7831 CCTTCATCAACCCGGGTATAGTGGAGATTGAAGTGTCTAGAATTAATAGCAAGTATCATGCAGTATCGGAGGCTAATATGGATCTGTACA
      P  F  I  N  P  G  I  V  E  I  E  V  S  R  I  N  S  K  Y  H  A  V  S  E  A  N  M  D  L  Y    1259-V
7921 TCGCTGCTGCAAAATCTGTGGGCGTAAAGCCCACACAGTTTGTTGAGGAAACAAACGACTTTACGGCCCGCGGCCACCACCATGGTTGTT
      I  A  A  A  K  S  V  G  V  K  P  T  Q  F  V  E  E  T  N  D  F  T  A  R  G  H  H  H  G  C    1289-V
8011 ATTCCCTTTCTTGGTCTAAGTCACGCAATCAATCACAGGTCCTAAAGATGGTAGTGCGGAAGCTGAAGCTATGTGTCCTGTATATATACC
      Y  S  L  S  W  S  K  S  R  N  Q  S  Q  V  L  K  M  V  V  R  K  L  K  L  C  V  L  Y  I  Y    1319-V
8101 CCACAGTCGATCCCGCCGTTGCTCTCGACCTGTGCCACCTGCCAGCACTAACTATAATCCTAGTGCTCGGCGGTGACCCAGCGTACTACG
      P  T  V  D  P  A  V  A  L  D  L  C  H  L  P  A  L  T  I  I  L  V  L  G  G  D  P  A  Y  Y    1349-V
8191 AGCGATTACTTGAGATGGACCTATGCGGGGCTGTGTCAAGTCGCGTTGATATCCCCCATTCCCTAGCTGCCAGAACGCACAGGGGGTTCA
      E  R  L  L  E  M  D  L  C  G  A  V  S  S  R  V  D  I  P  H  S  L  A  A  R  T  H  R  G  F    1379-V
8281 CAATAGGCCCAGACGCTGGTCCAGGTGTGATTAGACTTGACAAGTTAGAGTCGGTTTGTTACGCCCACCCCTGTTTGGAGGAGCTAGAGT
      T  I  G  P  D  A  G  P  G  V  I  R  L  D  K  L  E  S  V  C  Y  A  H  P  C  L  E  E  L  E    1409-V
8371 TTAATGCGTACCTAGACTCTGAGTTAGTTGATATTAGTGATATGTGCTGCCTCCCCCTAGCGACACCCTGTAAGGCCCTATTCAGGCCAG
      F  N  A  Y  L  D  S  E  L  V  D  I  S  D  M  C  C  L  P  L  A  T  P  C  K  A  L  F  R  P    1439-V
8461 TGTATCGGAGCTTACAGTCGTTCAGGTTAGCCTTAATGGACAACTATAGTTTTGTAATGGACCTCATTACGATCCGGGGGGTGGACATCA
      V  Y  R  S  L  Q  S  F  R  L  A  L  M  D  N  Y  S  F  V  M  D  L  I  T  I  R  G  V  D  I    1469-V
8551 GGCCTCACCTTGAGGAGTTTGATGAACTGCTTGTGGTGGGGCAGCATATCCTCGGTCAGCCCGTCCTAGTGGAGGTTGTTTACTACGTTG
      R  P  H  L  E  E  F  D  E  L  L  V  V  G  Q  H  I  L  G  Q  P  V  L  V  E  V  V  Y  Y  V    1499-V
8641 GAGTTGTTGGGAAGCGTCCTGTGTTAGCGAGGCATCCCTGGTCAGCAGATCTTAAGCGAATCACTGTAGGGGCGAGCGCCCTGCCCTTCT
      G  V  V  G  K  R  P  V  L  A  R  H  P  W  S  A  D  L  K  R  I  T  V  G  A  S  A  L  P  F    1529-V
8731 GCTGCTGGACTGCGTGATGAGGATTGTCGGGGGTCTCTGCTGGTTGGGCTTCCCGCTGGATTGACGCAGTTGTTGGTGGTTGATTGAGGT
      C  C  W  T  A  *                                                                             1534-V
8821 TGAGCCATCTACTGCCCTATTCTTAAAAAACCATACGTCAGTGGTGCAGTGCTTGGGTTTGGTGATTGGTTGGTTGTAGCGCT
```

FIG. 2—*Continued.*

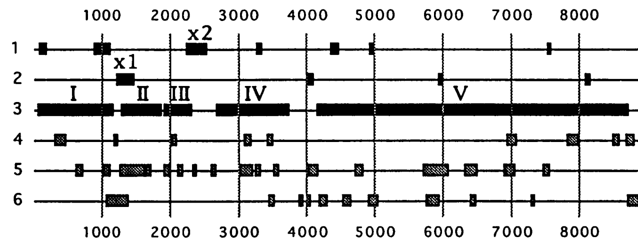

FIG. 3. Coding capability of the BDV genome. All possible ORFs larger than 25 amino acids in the BDV genome are indicated. Lanes 1, 2, and 3 correspond to ORFs found in the mRNA polarity, and lanes 4, 5, and 6 correspond to those found in the genomic RNA. The location of ORFs described in Fig. 2 are indicated as I, II, III, IV, V, x1, and x2. ORFs I and II correspond to the 40- and 24-kDa BDV proteins, respectively (49, 66). ORF III probably encodes the 14.5-kDa protein associated with BDV infection (64). We have identified ORF V as the BDV L polymerase.

Six independent clones were sequenced to determine the exact 5′ and 3′ ends of the BDV genome. A representative of these clones, RT-PCR 5′/3′, is presented in Fig. 1.

The sequence of the BDV genome, 5′-3′ positive sense (antigenic polarity), is presented in Fig. 2. Five major ORFs that we have designated as ORFs I, II, III, IV, and V are predicted in the BDV genome sequence (Fig. 3). ORFs I (nt 54 to 1163) and II (nt 1272 to 1874) corresponded to the previously described 40- and 24-kDa BDV proteins, respectively (49, 58, 66). ORF III (nt 1893 to 2318) is capable of encoding a predicted protein of approximately 16 kDa, which could account for the 14.5-kDa protein found associated with BDV infection (64). However, it has been proposed that the BD-specific 14.5-kDa protein may be encoded by a different ORF (58, 70), corresponding to ORF x1 (nt 1223 to 1483) in our sequence, that overlaps in a different frame with the ORF II (see Discussion). We also identified ORF x2 (nt 2236 to 2556), which overlaps in a different frame with the -COOH terminal of ORF III and is capable of encoding a predicted protein of approximately 12 kDa. Both ORFs x1 and III have good translational start (TS) signals, RNNATG, whereas ORF x2 has a very poor TS signal, YNNATGY (14).

ORFs IV (nt 2685 to 3743) and V (nt 4143 to 8744) are capable of encoding predicted proteins of approximately 40 and 170 kDa, respectively, which have not yet been identified in protein samples from BDV-infected cells or animals. The first ATG of ORF IV is flanked by a poor TS signal (YN NATGY) (14), whereas 21 amino acids downstream in the same frame there is a second ATG flanked by a good TS signal (RNNATG). In ORF V, the first ATG is flanked by a favorable TS signal (RNNATG) (14).

As shown in Fig. 3, the genomic polarity has a very limited coding capability. Three ORFs, cORF I (nt 6082 to 5699), cORF II (nt 1627 to 1259), and cORF III (nt 1386 to 1048), capable of encoding predicted proteins of approximately 14, 13.5, and 12.5 kDa, respectively, are found in the genomic polarity. However, none of the cORFs has a favorable TS signal (RNNATG) (14), and they are not flanked by polyadenylation or transcription start signals. It is not yet possible to rule out the possibility that these cORFs direct the synthesis of polypeptides during BDV infection; however, it seems very unlikely that BDV has an ambisense coding strategy.

Similar to other NNS-RNA viruses (2, 30, 32, 42, 67), BDV ORFs are flanked by nontranslated boundary sequences where regulatory signals involved in transcription initiation and stop are located. Examination of the BDV boundary sequences

A)

	Transcription Start	Transcription Stop/Polyadenylation
I (41)	3′ UACUUGUGUG 5′(50)	(1181) 3′ UGGUUAUUUUUU 5′(1192)
II (1173)	3′ UACUUGUUG 5′(1181)	(1871) 3′ UGGUAUUUUUUU 5′(1882)
III		
IV (2385)	3′ UGCCUUUUAC 5′(2394)	(3761) 3′ UGAGAAUUUUU 5′(3771)
Vi) (3995)	3′ UACUGUUUCU 5′(4004)	(8839) 3′ UAAGAAUUUUU 5′(8850)
ii) (4121)	3′ UCCCGGUUUU 5′(4130)	
Consensus	3′ UNC^{U}/C^{U}/G^{U}/GUUNN 5′	3′ U^{G}/A^{G}/A^{U}/G^{U}/AUUUUU 5′

B)

PF3	3′ ---UCCUNNUUUC- 5′	3′ UUNAUNNUNNNN-- 5′
NDV	3′ ---UNNNCNNNNNC 5′	3′ ---ANUCUUUUUU- 5′
SV	3′ ---UCCCANUUUN- 5′	3′ ---AUUCUUUUU-- 5′
MV	3′ ---UCCNNNNUNCN 5′	3′ ---NNNNUUUU--- 5′
RSV	3′ -----CCNGUUUN- 5′	3′ UCAAUAANUUUU-- 5′
VSV	3′ ---UUGUCNNUAG- 5′	3′ ---AUACUUUUUUU 5′
RV	3′ ---UUGUNNNGA-- 5′	3′ ----NACUUUUUUU 5′
MBG	3′ NNCUNCNUNUAAUU 5′	3′ -UAAUUCUUUUU-- 5′
BDV	3′ ---UNC^{U}/C^{U}/G^{U}/GUUNN- 5′	3′ -U^{G}/A^{G}/A^{G}/U^{U}/A^{U}/AUUUUU-- 5′

FIG. 4. Regulatory sequences of the BDV genome. (A) Putative transcriptional start and termination-polyadenylation signals located at the boundaries of BDV ORFs. The transcriptional signals are presented in the genomic polarity (negative sense). The different BDV ORFs are indicated on the left by Roman numerals. Numbers in parentheses indicate the nucleotide positions in the BDV genome delimiting the corresponding transcriptional signals. Potential transcription start and stop signals were not found flanking ORF III. Two putative transcription start signals were identified 3′ from the first ATG of ORF V. BDV consensus sequences of transcription start and termination-polyadenylation signals are indicated. (B) Comparison of consensus transcription start and termination-polyadenylation signals (negative sense) between BDV and other NNS-RNA viruses of the three families *Paramyxoviridae*, *Rhabdoviridae*, and *Filoviridae*. Consensus sequences were derived as described previously (30). Symbols: PF3, human parainfluenza 3 virus; NDV, Newcastle disease virus; SV, sendai virus; MV, measles virus; RSV, respiratory syncytial virus; VSV, vesicular stomatitis virus; RV, rabies virus; MBG, Marburg virus.

revealed the potential transcription start (S) and termination-polyadenylation (E) signals presented in Fig. 4A. We could not identify these signals for ORF III. Both the putative start and termination-polyadenylation signals showed minor variations among the different BDV ORFs. The deduced putative BDV start consensus sequence (UNCNNNUUNN) is identical to the one obtained when NNS-RNA viruses of the three *Mononegavirales* families, *Filoviridae*, *Paramyxoviridae*, and *Rhabdoviridae*, are compared (Fig. 4B) (30). In addition, the deduced putative BDV termination-polyadenylation consensus sequence (3′ U G/A G/A U/G AUUUUU 5′) also showed strong similarities with those found in other NNS-RNA viruses (Fig. 4B). The conserved six-mer AUUUUU located at the 5′ end of each termination-polyadenylation signal (genomic polarity) was always followed by the dinucleotide GG, except in ORF II, where CG was found. Determination of the exact 3′ ends of the mRNAs corresponding to ORFs I and II showed that the GG or CG is not present at the 3′ end of these mRNAs (49, 66; data not shown).

On the basis of the putative BDV transcription start and stop signals proposed (Fig. 4), the termination-polyadenylation signal of BDV ORF I is located immediately 3′ (antigenomic polarity, positive sense) from the transcription start signal of ORF II, whereas two termination-polyadenylation signals, 3′ GAAUGAUUUUUUU 5′, and 3′ UAUGAAUUUUU 5′, are found within ORF V at nt 4496 and 5225, respectively, in the BDV genome. Moreover, two potential transcription start signals are located 5′ upstream from the first ATG of ORF V, at 223 and 349 nt, respectively, from the termination-polyade-

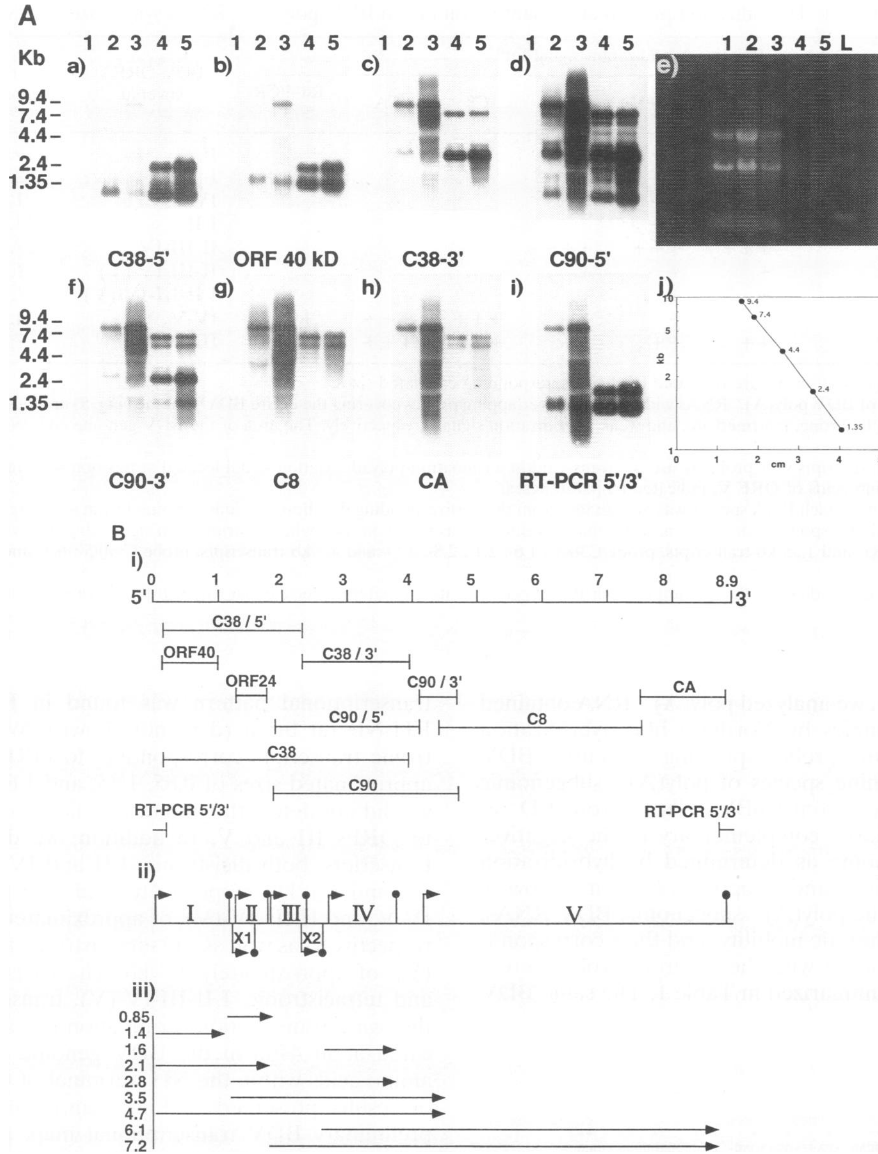

FIG. 5. Analysis by Northern blot hybridization of BDV poly(A)$^+$ RNAs synthesized in vivo. (A) BDV poly(A)$^+$ RNAs detected in RNA obtained from BD rat brain and C6BV cells. Lanes: 1, total RNA (10 μg) from normal rat brain; 2, total RNA (10 μg) from BD rat brain; 3, total poly(A)$^-$ RNA (10 μg) from BD rat brain; 4, cytoplasmic poly(A)$^+$ RNA (1 μg) from C6BV cells; 5, total poly(A)$^+$ RNA (1 μg) from BD rat brain. The same membrane was hybridized to the corresponding probes indicated at the bottom of each blot. Experimental procedures were as described in Materials and Methods. All the probes used had similar specific activities, of about 5×10^8 cpm/μg. The sizes of the BDV poly(A)$^+$ RNAs were estimated by plotting their actual electrophoretic mobilities along the standard curve constructed by plotting the electrophoretic mobilities of RNA species of known size contained in the ladder RNA (panel j). The line drawn through the standard points was obtained from a least-squares regression analysis. Panel e shows the ethidium bromide staining of the gel. L indicates ladder RNA. (B) (i) Location in the BDV genome (antigenomic polarity) of the BDV probes used for Northern blot hybridization experiments. (ii) Location of the BDV ORFs in the BDV genome (positive sense). ORFs are indicated by I, II, III, IV, and V. In addition, we have indicated the locations of ORF x1 (X1) and ORF x2 (X2). Arrows (↑) indicate the initiation of the ORFs (ATG), whereas black dots indicate the corresponding stop codons. (iii) BDV transcriptional map. A preliminary BDV transcriptional map is proposed on the basis of results presented in this figure and Table 1. Numbers on the left indicate the estimated sizes (in kilobases) of the BDV poly(A)$^+$ RNAs detected by Northern blot hybridization (panel A). The sizes of the arrows representing the BDV poly(A)$^+$ transcripts (5'-3', positive sense) do not include the tracks of poly(A) at the 3' end of each transcript, which are probably variable in size.

nylation signal of ORF IV, whereas ORF V is not followed by an S signal.

BDV subgenomic poly(A)$^+$ RNAs. Using BDV-specific probes corresponding to the 24- and 40-kDa BDV ORFs, we and others have described BDV subgenomic poly(A)$^+$ RNAs of approximately 0.85, 1.2, 2.1, and 3.5 kb present in RNA extracted from both BDV-infected cells and BD rat brains (26, 46, 58, 60, 70). However, on the basis of results presented in Fig. 2, the previously described BDV probes (46, 66, 70) represented only the 2-kb-most 3' end of the BDV genome (Fig. 2 and 5) and very probably did not recognize all BDV poly(A)$^+$ transcripts expressed during BDV infection. To

TABLE 1. Hybridization pattern and quantitation of the BDV poly(A)$^+$ RNAs synthesized in vivo

RNA size (kb)[a]	Relative hybridization signal for[b]:								BDV-ORF(s) covered[c]	% Total amount[d]	Molar ratio[e]	% Molar ratio
	C38/5′	40-kD ORF	C38/3′	C90/5′	C90/3′	C8	CA	RT-PCR 5′/3′C				
0.85	+++			++					II	19.10	1	38.40
1.35	+++	+++						+++	I	22.50	0.71	27.30
1.60				++	++				IV	1.80	0.05	1.90
2.10				++					I-II	23.60	0.48	18.50
2.80			+++	+++	+++				II-III-IV	17.20	0.26	10.0
3.50	+		+	+	+				II-III-IV-(V)	1.50	0.02	0.70
4.70	+		+	+	+				I-II-III-IV-(V)	1.20	0.01	0.30
6.10			+	+	++	+++	+++	+	IV-V	4.50	0.03	1.10
7.25			++	++	++	+++	+++	+	III-IV-V	7.30	0.04	1.50

[a] The different BDV poly(A)$^+$ transcripts are indicated by their corresponding estimated sizes.

[b] The hybridization patterns of BDV poly(A)$^+$ RNAs with the set of overlapping probes covering the entire BDV genome (Fig. 5) are indicated by the symbols +++, ++, and +, which correspond to strong, intermediate, and weak hybridization signals, respectively. The area of the BDV genome covered by each probe is indicated in Fig. 5C (i).

[c] The 3.5- and 4.5-kb BDV transcripts very probably use the transcription termination-polyadenylation signal located at position 4496 in the BDV genome (Fig. 2), covering only the first 117 amino acids of ORF V, indicated by parentheses.

[d] The amount of RNA present in each RNA species was quantitated from their corresponding densitometer units obtained from scanning of autoradiogram exposure within the linear range of the film response. For each transcript the quantitation was based on the highest hybridization signal obtained with the different probes that recognized each transcript (0.85- and 1.35-kb transcripts, probe C38/5′; 1.6-, 2.1-, 2.8-, 3.5-, and 4.7-kb transcripts, probe C 90/5′; 6.1- and 7.25-kb transcripts, probe C8).

[e] Molar ratios were determined by dividing the amount of each RNA species by its estimated molecular size and setting the value obtained for the 0.85-kb species as 1.

investigate this question, we analyzed poly(A)$^+$ RNA obtained from BDV-infected samples by Northern blot hybridization with a set of overlapping probes spanning the entire BDV genome. We detected nine species of poly(A)$^+$ subgenomic RNAs in RNA extracted from C6BV cells or from BD rat brain (Fig. 5), with polarity complementary to the negative-strand BDV RNA genome as determined by hybridization experiments with specific-strand riboprobes (data not shown). The estimated sizes of the poly(A)$^+$ subgenomic BDV RNAs, based on their electrophoretic mobility, and their corresponding patterns of hybridization with the different probes across the BDV genome are summarized in Table 1. The same BDV

```
A.

RV  1   UGCGAAUUGUUGGU---CUAGUUUCUUUUUUGUCUGUCGCAGUUA--C-CGUCU-CGUUUUUACA 58
           **********    ****    *    *    *       **   *  ** * ***
BDV 1 CAACGCAAUUGUUGGUUGGUUAGUAA--UA---GGAAGAU--UAUUUUACUUGUGUGCGU      53
             ^  ^      ^  ^                  -----------
                                                   ==========

B.
      1       10            24                   44       53

  3'  CAACGCAAUUGUUGGUUGGUUAGUAAUAGGAAGAUUAUUUUACUUGUGUGCGU
               |||||||||||||||   |  |                  ||  ||| |
  5'  AGCGCUACAACCAACCAAUCACCAAACCCAAGCACUGCACCACUGACGUA

      8903  8897         8883                  8862    8854
```

FIG. 6. BDV 3′ leader RNA. (A) BDV 3′ leader RNA displays significant similarities to rabies virus 3′ leader RNA. Both leader RNAs are presented in genomic polarity. Homology between BDV and rabies virus 3′ leader RNAs was analyzed as described previously (30). Symbols: *, identical nucleotides in both leader RNAs; ^, uridine residues repeated at every second or third position that have been found to be conserved at the 3′ ends among 3′ leader RNAs of members of the *Mononegavirales* (5, 30). A small conserved box of 5 nt (3′ UAUUU 5′, positive sense), which is found in the middle of the 3′ leader RNAs of NNS-RNA viruses (30), is found at the 5′ end of the BDV 3′ leader RNA and is underlined (single solid line). A putative sequence which might operate as a BDV leader-termination signal (3′ AUUAUUUU 5′) is also indicated (single dashed line), as well as a potential transcription start signal for ORF I (3′ UACUUGUGUG 5′) (double dashed line). (B) Complementarity between 3′ and 5′ ends of the BDV genome. Vertical lines indicate complementary nucleotides between both extremities.

transcriptional pattern was found in RNA extracted from PTI-NB rat brain (data not shown). We detected monocistronic transcripts corresponding to ORFs I, II, and IV with approximated sizes of 0.85, 1.35, and 1.6 kb, respectively, but we did not detect the monocistronic transcripts corresponding to ORFs III and V. In addition, we detected polycistronic transcripts, both dicistronic, I-II and IV-V, of approximately 2.1 and 6.1 kb, respectively, and tricistronic, II-III-IV, III-IV-V, and II-III-IV-(V), of approximately 2.8, 7.2, and 3.5 kb, respectively, as well as the tetracistronic transcript, I-II-III-IV-(V), of approximately 4.7 kb. The tricistronic, II-III-IV-(V), and tetracistronic, I-II-III-IV-(V), transcripts apparently use the same transcription termination-polyadenylation signal located at nt 4496 in the BDV genome (positive sense), 117 amino acids within the NH_2 terminal of ORF V. On the basis of results presented in Fig. 5 and Table 1, we propose a preliminary BDV transcriptional map, presented in Fig. 5B (iii).

Leader sequences at the 3′ and 5′ ends of the BDV genome. Untranslated regions known as (+) and (−) leader RNAs [ranging from 46 to 58 nt for the (+) leader RNAs and from 41 to 155 nt for (−) leader RNAs] (2, 30, 32, 42) are found at the 3′ and 5′ ends of the genomes of NNS-RNA viruses. Putative (+) and (−) leader RNAs, both of 53 nt, were found in the 3′ and 5′ ends, respectively, of the BDV genome (Fig. 2 and 6). Similar to other NNS-RNA viruses, the putative 3′ leader RNA of BDV had a high content of adenine and uridine residues (64%), of which uridine residues represented 64% and adenine residues represented 36%. Computer-assisted comparisons revealed a high degree of homology between the 3′-most ends of rabies virus (nt 5 to 19) and BDV (nt 7 to 24) genomes (Fig. 6A). Comparison of the 3′ leaders among the families *Rhabdoviridae*, *Paramyxoviridae*, and *Filoviridae* revealed a conserved motif at the 3′ ends which is characterized by the presence of uridine residues at every second or third position (5, 30). These uridine residues were also found in the 3′ leader of BDV (Fig. 6A). Moreover, the small conserved sequence 3′ UAUUU 5′ found in the middle of the 3′-end leaders of NNS-RNA viruses (30) was also found in the

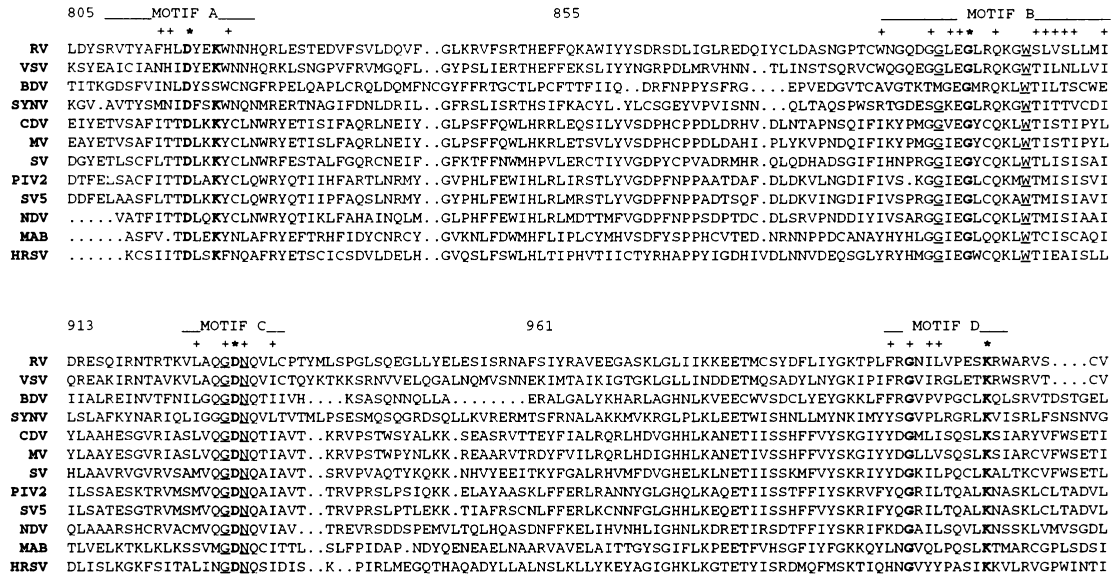

FIG. 7. Alignment of the proposed catalytic domain of the L polymerases of NNS-RNA viruses including the BDV L polymerase. The alignment of the L-protein sequences of the indicated viruses was produced as described in Materials and Methods. Abbreviations of virus names are the same as in Fig. 4, with the following additions: CDV, canine distemper virus; PIV2, parainfluenza virus 2; SV5, simian virus 5; HRSV, human respiratory syncytial virus. The four conserved motifs (A, B, C, and D) within the catalytic domain of the L-proteins are indicated. Strictly invariant residues are shown in boldface type. Plus signs represent identical or chemically similar amino acids conserved in more than 70% of the sequences. Asterisks refer to amino acids conserved in all RNA-dependent polymerases. Amino acid residues characteristic of the partition between nonsegmented and segmented negative-strand RNA viruses are underlined.

putative BDV 3′-end leader (Fig. 6A). However, in contrast to other NNS-RNA viruses, this sequence was found in the 5′ end of the BDV 3′ leader RNA overlapping with a putative transcription termination-polyadenylation signal, which might operate as a leader termination signal (Fig. 6A).

Complementarity at the genome extremities involving 10 to 20 nt is a common feature among NNS-RNA virus genomes (2, 30, 32, 41, 42). Nucleotides 10 to 24 at the 3′ end and 8883 to 8987 at the 5′ end of the BDV genome (negative sense) displayed perfect complementarity (Fig. 6B). Twenty nucleotides removed from this region with 100% complementarity is an additional stretch of 8 nt displaying 75% complementarity, corresponding to nt 44 to 53 at the 3′ end and 8854 to 8862 at the 5′ end (negative sense).

Analysis of the predicted proteins of the BDV genome. The deduced amino acid sequences from the BDV ORFs were used to screen the GenBank protein data base for similarities by using the IntelliGenetics FASTDB program and for functional and structural motifs by using the QUEST program. In agreement with previous reports (58, 66), we did not find significant regions of homology or identity for ORFs I and II, with the exception of the previously mentioned very distant homology of ORF I to the L-gene polymerases (49). BDV ORFs III and IV are capable of encoding polypeptides 142 and 353 amino acids long, respectively, with corresponding calculated molecular masses of 16.2 and 39.9 kDa. ORF III very probably encodes the BD-specific 14.5-kDa protein on the basis of partial amino acid sequence obtained from purified 14.5-kDa protein (9, 64), whereas the predicted polypeptide encoded by ORF IV has not yet been identified in BDV-infected samples. Possible functions associated with these BDV proteins are presently unknown. Both ORFs III and IV have a basic character with pIs of 9.069 and 9.412, respectively, and they both also have a high leucine content, 10.48 and 12.68% for ORFs IV and III, respectively. ORF III showed a very weak homology to the matrix protein of Newcastle disease virus and has six potential phosphorylation sites: a cyclic AMP- and GMP-dependent protein kinase phosphorylation site (RE**KKLTD**), two casein kinase II phosphorylation sites (KH**SYVEL** and FLSVKEP), and three protein kinase C phosphorylation sites (GGTSRN, **FLSVKE**, and VFSTRL) (Fig. 2). In addition, an N-myristoylation site (VG**GTSRNQFL**) is found in ORF III (Fig. 2), which may facilitate the interaction of this protein with membranes. BDV ORF IV displayed a very distant homology to the rat MHC-linked transporter protein and has 10 potential N-glycosylation sites, 6 of which are located between amino acids 41 and 88 in the NH_2 terminal. The presence of seven potential N-myristoylation sites in ORF IV suggested that this polypeptide, similarly to ORF III, may interact with membranes. The interaction with membranes might be further facilitated by the hydrophobic domain localized in the COOH terminal of ORF IV. In addition, ORF IV has eight potential phosphorylation sites: four protein kinase C phosphorylation sites (**LESGKAM**, **NSSSKIV**, **TSSCRSK**, and **RPTLKDA**), and four casein kinase II phosphorylation sites (YC**SFADC**, TV**SQQEL**, IV**SFDE**, and **RPTLKDA**) (Fig. 2).

BDV ORF V (nt 4143 to 8744) is the largest ORF predicted in the BDV genome sequence (Fig. 2 and 3). It is capable of encoding a polypeptide 1,534 amino acids long of approximately 170 kDa, which has not yet been identified in BDV-infected animal samples or BDV-infected cultured cells. The polypeptide encoded by ORF V has 152 acidic and 191 basic amino acids and a high content (18.32%) of the hydrophobic amino acids leucine and isoleucine, about 1.5 times higher than that of an average protein (24). The predicted protein has a net charge of +19.5, calculated on the basis of the units of charges (4). The location of ORF V in the 5′ end of the BDV genome together with its high molecular mass (170 kDa), as well as a large net positive charge and fewer acidic than basic amino acids, which are characteristic features of the NNS-RNA virus

polymerases, suggested that ORF V could encode the BDV L protein. When we compared BDV ORF V with the GenBank protein data base, we found a very high degree of homology to RNA-dependent RNA polymerases, especially to the L proteins of NNS-RNA viruses. The predicted amino acid sequence of BDV ORF V was then compared with the L proteins of NNS-RNA viruses by multiple alignment (data not shown). Significant homology between BDV ORF V and other L proteins was found throughout the amino-terminal half, whereas the carboxy termini displayed a high degree of variability, with BDV containing the less highly conserved sequence among the carboxy termini of the L proteins. The homology of BDV ORF V to L proteins was particularly high for the ORF V region contained between amino acids 360 and 560. This segment corresponded to a conserved domain among the RNA-dependent RNA polymerases, which is most probably responsible for performing the catalytic process (25, 53, 56, 57, 67). This domain consists of 5 invariant and 22 conservatively maintained residues, clustered in four highly conserved motifs, A, B, C, and D (reviewed in reference 67). The alignment along the proposed catalytic domain for the L proteins of NNS-RNA viruses, including the BDV putative polymerase, BDV ORF V, is presented in Fig. 7. The 4 amino acids conserved in all RNA-dependent polymerases were also conserved in the BDV ORF V, as were the 22 conservatively maintained residues (Fig. 7). However, the strictly invariant lysine residue located in motif A of NNS-RNA virus polymerases is replaced by serine in BDV. In addition, the putative template-binding site consisting of positively charged and hydrophobic residues regularly spaced every four amino acids previously identified in the L proteins of NNS-RNA viruses (56), was found in the BDV ORF V (amino acids 326 to 338; Fig. 2).

DISCUSSION

We report here for the first time the cloning and complete genome sequence of BDV. This sequence revealed a BDV genomic organization that is characteristic of the order *Mononegavirales*. We have identified a BDV candidate polymerase gene that belongs to the L-protein family. The genomic organization, together with the L-protein sequence, suggests that BDV is most closely related to the family *Rhabdoviridae* in the order *Mononegavirales*. However, our results indicate that BDV transcriptional mode seems unique for an animal rhabdovirus.

The conclusion that the sequence presented in Fig. 2 indeed constitutes a full-length BDV genome is supported by several experimental findings. First, the determined BDV genome sequence is 8,903 nt long, in excellent agreement with our previously estimated size for the only RNA species found in infectious BDV-RNPs, determined by electrophoretic mobility (21). Second, a high degree of homology was found at the 3′-end untranslated regions between BDV and other NNS-RNA virus genomes (Fig. 6A), whereas the 3′ and 5′ ends of the BDV genome displayed complementarity (Fig. 6B), a common feature among NNS-RNA viruses (2, 30, 32, 41, 42, 67). Moreover, the 5′-most BDV ORF, ORF V, is not followed by an identifiable transcription start signal. Together, these findings indicate that the BDV genome sequence which we determined contains the 5′ and 3′ ends, where important regulatory signals involved in polymerase recognition and encapsidation have been previously described for several NNS-RNA viruses (2, 5, 22, 30, 32, 41, 42, 52, 54). Furthermore, when RT-PCR of the self-ligated BDV genomic RNA (see Materials and Methods; Fig. 1) was performed with the primer complementary to nt 222 to 243 of the 40-kDa BDV ORF (49) to prime first-strand cDNA synthesis, a product of identical size and sequence but reverse orientation to the one described in Fig. 1 was obtained. In addition, about a 20-fold-lower level of amplified product was obtained when priming of the first-strand cDNA synthesis was conducted with the primer of genomic polarity. These results indicated that BDV genomic cRNA is also present in BDV-RNPs, although at a much lower level than the BDV genomic RNA, and that both genomic and antigenomic BDV RNAs have equivalent 5′ and 3′ ends of complementary sequence, which further supports the conclusion that the BDV genome sequence presented here contains the 3′ and 5′ ends of the BDV genome. The biological significance of the complementarity at the ends of NNS-RNA viruses is currently unknown, but it may simply reflect the use of similar sequences at the 3′ ends of genomic and antigenomic RNAs for transcription and replication control or the nucleation site for the binding of viral polypeptides involved in the encapsidation process, or both (2, 32, 41, 67). In contrast to many other NNS-RNA viruses, the complementarity between the 3′ and 5′ ends of the BDV RNA genome does not include the 10 most 3′-end nucleotides (Fig. 6), a characteristic also displayed by the Newcastle disease virus RNA genome (32). Moreover, very limited similarity has been documented between the (−) 3′ and (+) 3′ ends of the respiratory syncytial virus RNA genome. Third, the determined BDV genome sequence has the coding capability to direct the synthesis of at least five independent polypeptides, which could conceivably provide all the required functions for recognition of the target cell, multiplication within the infected cell, and generation of new viral particles.

BDV poly(A)$^+$ transcripts synthesized in vivo were analyzed by Northern blot hybridization with a set of overlapping probes across the entire BDV genome. Our results revealed a relatively complex transcriptional pattern (Fig. 5; Table 1). BDV polycistronic poly(A)$^+$ RNAs accounted for about 55% of the total synthesis of BDV poly(A)$^+$ transcripts, representing approximately 32% of the total molar amount of BDV poly(A)$^+$ RNA synthesis. Polycistronic poly(A)$^+$ RNAs have been described previously during infection with the rhabdovirus vesicular stomatitis virus (48) and several paramyxoviruses (reviewed in references 32 and 42); they accounted only for 1% (vesicular stomatitis virus) and 25% (paramyxoviruses) of the total poly(A)$^+$ RNA synthesis (36, 48, 71), considerably lower than the 55% estimated for BDV. Paramyxoviruses and vesicular stomatitis virus in vivo synthesized polycistronic poly(A)$^+$ transcripts are thought to represent aberrant transcription products generated by occasional termination failure and continued transcription into the next gene (35, 72). Therefore, the junctions between gene transcripts are devoid of internal poly(A) sequences and represent accurate copies of the genome sequences at those locations (35, 48, 71, 72). Moreover, it has been documented for paramyxoviruses and vesicular stomatitis virus that changes in the sequences of the transcriptional stop signals, as well as in the intergenic regions, can modulate the incidence of aberrant transcription (35, 48, 51). The mechanisms by which BDV polycistronic poly(A)$^+$ RNAs are derived are presently unknown. Cloning of these BDV polycistronic poly(A)$^+$ RNAs will allow us to determine whether their intercistronic regions represent accurate copies of the genomic sequences at those locations or whether the cistrons within the polycistronic RNAs are connected by sequences containing stretches of poly(A). Similarly, the possible contribution of specific BDV genomic sequences to the regulation of the synthesis of BDV polycistronic poly(A)$^+$ RNAs also remains to be determined.

Presently, with the exception of the 3′ ends of the monocistronic transcripts corresponding to ORFs I and II that have been exactly determined (data not shown), we do not have information about the exact 5′ and 3′ ends of the BDV poly(A)$^+$ transcripts described in Fig. 5 and Table 1. Similarly, cDNAs derived from the BDV mRNAs have been analyzed only for the monocistronic transcripts corresponding to ORFs I and II (46, 49, 66, 70) and the dicistronic transcript ORF I-II (58). Consequently, we cannot rule out the possibility that splicing mechanisms such as those in orthomyxoviruses (41) and addition of nontemplated nucleotides as in paramyxoviruses (32, 42) contribute to the generation of some of the BDV poly(A)$^+$ RNAs.

If biochemically confirmed, the putative transcriptional S signal of BDV ORF II would be located immediately 3′ (genomic polarity) from the transcriptional E signal of BDV ORF I (Fig. 2 and 4), which could contribute to the attenuation of ORF II expression at the level of transcription, similar to the finding described for the L gene of respiratory syncytial virus (18). The absence of consensus transcriptional S and E signals at the 3′ and 5′ ends (genomic polarity), respectively, of ORF III (Fig. 2 and 4) could explain our inability to detect the corresponding monocistronic poly(A)$^+$ transcript for ORF III. However, we detected four polycistronic poly(A)$^+$ RNAs that contained ORF III (Fig. 5; Table 1). The partial peptide amino acid sequence NSKHSYV obtained from purified BDV-specific 14.5-kDa protein (9, 64) exactly corresponds to the NH_2-terminal sequence deduced from ORF III, whereas the indicated peptide sequence is not found in ORF x1, which has been previously proposed to encode the 14.5-kDa BDV protein (58, 70). The relatively high levels of the BDV 14.5-kDa protein found in BDV-infected cells and BD rat brains suggests that polycistronic poly(A)$^+$ containing the ORF III are incorporated into polysomes and are biologically functional as templates for protein synthesis. Polycistronic poly(A)$^+$ transcripts associated with polysomes have been previously described for Newcastle disease virus infection (71). In addition, the matrix gene (M) of the paramyxovirus simian virus 41 is exclusively transcribed as dicistronic transcript (68), whereas studies with measles virus showed that the dicistronic P-M RNA could direct the protein synthesis from the first, but not the second, cistron (72). Similarly, the BD-specific 14.5-kDa protein might be translated from the 7.25-kb BDV polycistronic poly(A)$^+$ transcript containing ORF III at the 5′ end, whereas translation of the 14.5-kDa polypeptide from the other polycistronic poly(A)$^+$ transcripts also containing ORF III would require initiation of translation at an internal cistron or the processing of a polyprotein product. Alternatively, it is possible that a monocistronic mRNA for ORF III is transcribed at levels below the sensitivity of our Northern blot hybridization conditions and that the detection of the 14.5-kDa polypeptide in samples from BDV-infected material reflects the accumulation with time of a very stable protein being synthesized at a very low rate. Whether the polypeptide products from ORFs x1 and x2 are synthesized in BDV-infected cells and whether they might contribute to the 14.5-kDa BDV antigen remain to be determined.

We were also unable to detect the monocistronic mRNA corresponding to ORF V. In contrast to ORF III, for which we could not identify the transcriptional S and E signals, ORF V was flanked by sequences containing putative S and E signals (Fig. 2 and 4). As discussed below, homology to the L-proteins of other NNS-RNA viruses indicated that ORF V is likely to be the BDV polymerase. Thus, translation of ORF V would be required for BDV transcription and replication to take place. Consequently, either the ORF V product is either translated from the second cistron of the BDV polycistronic mRNAs containing ORF V (Fig. 5; Table 1) or, alternatively, there is a monocistronic transcript for the BDV ORF V present at levels below the sensitivity of our Northern blot hybridization assays. Translation of this monocistronic mRNA has to provide the amount of L-protein required to support the relatively high levels of BDV transcription and replication found in BDV-infected cells. The use of hybrid-selected BDV polycistronic poly(A)$^+$ RNAs in in vitro translation assays will allow to determine how the polypeptides predicted by BDV ORFs III and V are synthesized. On the basis of results presented in Fig. 5, the approximately 7.2-kb BDV polycistronic poly(A)$^+$ RNA is initiated at a position 3′ (genomic polarity) from the first ATG of ORF III. We do not yet have information regarding the exact 5′ ends of the BDV polycistronic transcripts, and consequently we cannot exclude the role of splicing mechanisms in their generation. However, if transcription of the BDV 7.25-kb RNA is confirmed to occur around the position suggested in Fig. 5, it would raise the question of why we have not detected the dicistronic RNA ORF III-IV. Moreover, our observation that the monocistronic ORF IV mRNA and the dicistronic ORF IV-V RNA are transcribed without termination of the ORF III transcript, together with a previous observation made with simian virus 41 (68), raises the possibility that in addition to the "start-and-stop" mechanism proposed for transcription of NNS-RNA viruses (reviewed in references 2, 32, and 42), NNS-RNA virus polymerases also enter directly into an upstream noncoding region of a gene without requiring the previous recognition of the genomic 3′ end. The identical pattern of transcription found in BDV-infected cultured cells, BD rat brain, and PTI-NB rat brain supports the conclusion that the results presented in Fig. 5 and Table 1 accurately reflect the transcriptional mode of BDV. However, we cannot rule out the possibility that defective interfering DI particles or defective interfering RNAs can serve as templates for the synthesis of BDV poly(A)$^+$ polycistronic RNAs.

Results presented in Fig. 5 and Table 1, together with the potential transcriptional start and stop signals indicated in Fig. 4, predict BDV intergenic regions of variable size and nucleotide composition. The predicted intergenic region between BDV ORFs IV and V may be variable in size, 221 or 347 nt long, because of the existence of two transcriptional start signals for ORF V. A very large size for the intergenic region located immediately 3′ from the L gene is also found in rabies virus.

Free 3′-end leader RNA species are frequently (2, 32, 42–44) but not always (13, 20, 72) found during infection with NNS-RNA viruses. Experimental evidence indicates that the exact 3′ end of the genome is the entry site for the RNA transcriptional machinery and that 3′-end leader RNA synthesis is mandatory before progression of the transcriptase through the rest of the genome (reviewed in references 2, 32, and 42). However, there is also evidence suggesting alternative patterns of antigenomic RNA synthesis (68, 72). We do not have information regarding whether free 3′-end leader RNA species are present in BDV-infected cells.

The catalytic subunit of the polymerase complex is the only gene whose degree of conservation allows the establishment of evolutionary relationships among negative-strand RNA viruses (56). It has been previously proposed that the 40-kDa BDV polypeptide (BDV p40), corresponding to ORF I in our BDV genome sequence, is the BDV polymerase (49). However, BDV p40 displays only a very distant homology to other L-gene polymerases of NNS-RNA viruses, and it does not contain the proposed catalytic domain conserved among NNS-

RNA virus polymerases (16, 56, 57, 67). In contrast, the predicted polypeptide encoded by ORF V clearly belongs to the L-protein family, as indicated by its high degree of homology to the polymerases of NNS-RNA viruses, especially the *Rhabdoviridae* family (Fig. 7) (data not shown). Moreover, all the conserved motifs previously identified in the proposed catalytic domain of the L proteins (16, 56, 57, 67), as well as the putative template-binding site (16, 56), are also conserved in BDV ORF V (Fig. 2 and 7). Consequently, we propose that ORF V is the BDV polymerase. The presence of the sequence GDN in motif C of the BDV L protein is in agreement with the previously described partition between segmented and nonsegmented viruses based on the L-protein sequences (67). However, the glycine at position 895 (Fig. 7) in motif B, which is characteristic of NNS-RNA viruses, is replaced with methionine in the BDV L protein, more commonly seen in negative-strand RNA viruses with segmented genomes. Interestingly, some additional significant differences are found between the BDV and other L-protein sequences. A multiple alignment of the L proteins (data not shown) placed the NH_2 terminus of BDV ORF V near the first highly conserved region found among the L proteins, which is characterized by the conserved motif DhhhR (where h means a hydrophobic residue, i.e., F, Y, W, I, L, M, or V). This alignment would imply that the motif LXSP (where X is N, S, or D) observed at the NH_2 terminus of the L polymerases (56) is not present in the BDV polymerase, similar to the L protein of the plant rhabdovirus sonchus yellow net virus (SYNV) (16). In addition, several small motifs characteristic of the L proteins of the NNS-RNA viruses are not present in the BDV L protein. Thus, the motifs GHP, CXNIS, PHP, HXH, and GXGXG, located around positions 430, 960, 1270, 1480, and 1950, respectively, in a previously reported alignment (56), are not found in BDV ORF V. The motif GHP is conserved in all L proteins known of the three families of the *Mononegavirales* order: *Rhabdoviridae*, *Paramyxoviridae*, and *Filoviridae*. The motif CXNIS is conserved in paramyxoviruses and rhabdoviruses, with the exception of SYNV. The motif PHP is conserved in the rhabdoviruses including SYNV, whereas in the paramyxoviruses there is an equivalent motif, PDP. The motif HXH is conserved in paramyxoviruses and rhabdoviruses, again with the exception of SYNV. Finally, the motif GXGXG is conserved in filoviruses, paramyxoviruses, and rhabdoviruses, with the exception of SYNV.

We have previously documented that transcription and replication of BDV take place in the nucleus of the infected cells (21). The sequences VSKNAKWPPV and WYKVRKVT at amino acids 189 to 197 and 943 to 950, respectively, in BDV ORF V are putative nuclear localization signals (61) that can mediate the transport of the BDV L protein from its place of synthesis in the cytoplasm to the nucleus.

The complete sequence of the BDV genome presented here has revealed a modular genomic organization similar to that displayed by other members of the *Mononegavirales* order, with three conserved gene blocks (67): block 1 codes for the nucleoproteins and polymerase cofactors, which may correspond to ORFs I and II in the BDV genome; block 2 codes for the matrix and envelope proteins, whose possible counterparts in the BDV genome are ORFs III and IV; and block 3 codes for the viral polymerase, which we have identified as ORF V in the BDV genome. In addition, on the basis of the homology between BDV ORF V and the L proteins of the NNS-RNA viruses, BDV seems to be especially closely related to the rhabdoviruses.

A major difference between BDV and other animal NNS-RNA viruses is that transcription and replication of BDV take place in the nucleus of the infected cells where infectious BDV-RNPs are present (21). Orthomyxoviruses are the only other group of animal negative-strand RNA viruses for which transcription and replication have been established to occur in the nucleus (reviewed in reference 41). Initiation of transcripts by a "cap-snatching" mechanism and the use of host nuclear splicing machinery to generate some of the viral mRNAs are thought to contribute to the nuclear requirement for the replication and transcription of the orthomyxoviruses (reviewed in reference 41). The reasons why BDV needs to conduct replication and transcription in the nucleus, processes that are performed in the cytoplasm by all the other animal NNS-RNA viruses, are presently unknown. It should be mentioned that the subviral agent hepatitis delta virus also has a nonsegmented negative-strand RNA genome that replicates in the nucleus. Further studies will investigate the requirements for BDV transcription and replication. Interestingly, replication and transcription, as well as the morphogenesis of the plant rhabdovirus SYNV, also occur in the nucleus (38, 39, 69). In addition, our results indicated that the L proteins of BDV and SYNV are the most distantly related to the L proteins from animal rhabdoviruses. These findings suggest a possible relationship between BDV and the plant rhabdoviruses. Studies of the molecular biology of BDV may provide valuable information regarding the evolution of rhabdoviruses, especially the relationships between animal and plant rhabdoviruses.

ACKNOWLEDGMENTS

We thank O. Poch for assistance regarding the analysis of the BDV L gene and K. Carbone for providing the PTI-NB rat brain samples. We also express our deep appreciation to A. Bang for helpful discussion and suggestions, as well as for editorial and graphics assistance.

This work is supported in part by USPHS grant NS-12428.

REFERENCES

1. **Amsterdam, J. D., A. Winokur, W. Dyson, S. Herzog, F. Gonzalez, R. Rott, and H. Koprowski.** 1985. Borna disease virus: a possible etiologic factor in human affective disorders? Arch. Gen. Psychiatry **42:**1093–1096.
2. **Banerjee, A. K.** 1987. Transcription and replication of rhabdoviruses. Microbiol. Rev. **51:**66–87.
3. **Bantle, J. A., I. H. Maxwell, and W. E. Hahn.** 1976. Specificity of oligo(dT)-cellulose chromatography in the isolation of polyadenylated RNA. Anal. Biochem. **72:**413–427.
4. **Barik, S., E. W. Reed, D. Luk, A. K. Banerjee, and C. Yong Kang.** 1990. Nucleotide sequence analysis of L gene of vesicular stomatitis virus (New Jersey serotype): identification of conserved domains in L-proteins of nonsegmented negative-strand viruses. Virology **175:**332–337.
5. **Blumberg, B. M., C. Giorgio, and D. Kolakofsky.** 1983. N protein of vesicular stomatitis virus selectively encapsidates leader RNA in vitro. Cell **32:**559–567.
6. **Bode, L., S. Riegel, W. Lange, and H. Ludwig.** 1992. Human infections with Borna disease virus: seroprevalence in patients with chronic diseases and healthy individuals. J. Med. Virol. **36:**309–315.
7. **Bode, L., S. Riegel, H. Ludwig, J. D. Amsterdam, W. Lange, and H. Koprowski.** 1988. Borna disease virus-specific antibodies in patients with HIV infection and with mental disorders. Lancet **ii:**689.
8. **Briese, T., J. C. de la Torre, A. Lewis, H. Ludwig, and W. I. Lipkin.** 1992. Borna disease virus, a negative-strand RNA virus, transcribes in the nucleus of infected cells. Proc. Natl. Acad. Sci. USA **89:**11486–11489.
9. **Briese, T., W. I. Lipkin, and J. C. de la Torre.** Molecular biology of Borna disease virus. *In* W. I. Lipkin and H. Koprowski (ed.), Borna disease, in press. Springer-Verlag KG, Heidelberg.
10. **Brutlag, D. L., J. P. Dautricourt, S. Maulik, and J. Relph.** 1990.

Improved sensitivity of biological sequence databank searches. Comp. Appl. Biol. Sci. 6:237–245.

11. **Carbone, K. M., S. W. Park, S. A. Rubin, R. W. Waltrip II, and G. B. Vogelsang.** 1991. Borna disease: association with a maturation defect in the cellular immune response. J. Virol. **65:**6154–6164.
12. **Carbone, K. M., S. A. Rubin, A. M. Sierra-Honigmann, and H. M. Lederman.** 1993. Characterization of a glial cell line persistently infected with Borna disease virus (BDV): influence of neurotrophic factors on BDV protein and RNA expression. J. Virol. **67:**1453–1460.
13. **Castaneda, S. J., and T. C. Wong.** 1989. Measles virus synthesizes both leaderless and leader-containing polyadenylated RNAs in vivo. J. Virol. **63:**2977–2986.
14. **Cavener, D. C.** 1987. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acids Res. **15:**1353–1361.
15. **Chayen, J., L. Bitensky, R. Butcher, and W. J. Rutter.** 1969. A guide to practical histochemistry, p. 71–72. Oliver and Boyd, Ltd., Edinburgh.
16. **Choi, T. J., S. Kuwata, E. V. Koonin, L. A. Heaton, and A. O. Jackson.** 1992. Structure of the L (polymerase) protein gene of Sonchus yellow net virus. Virology **189:**31–39.
17. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
18. **Collins, P. L., R. A. Olmsted, M. K. Spriggs, P. R. Johnson, and A. Buckler-White.** 1987. Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. Proc. Natl. Acad. Sci. USA **84:**5134–5138.
19. **Corpet, F.** 1988. Multiple sequence alignment with hierarchial clustering. Nucleic Acids Res. **16:**10881–10890.
20. **Crowley, J. C., P. C. Dowling, J. Menonna, J. I. Silverman, D. Shuback, S. D. Cook, and B. M. Blumberg.** 1988. Sequence variability and function of measles virus 3′ and 5′ ends and intercistronic regions. Virology **164:**498–506.
21. **Cubitt, B., and J. C. de la Torre.** 1994. Borna disease virus (BDV), a nonsegmented RNA virus, replicates in the nucleus of the infected cells where infectious BDV ribonucleoproteins are present. J. Virol. **68:**1371–1381.
22. **Curran, J. A., and D. Kolakofsky.** 1991. Rescue of a Sendai virus DI genome by other parainfluenza viruses: implications for genome replication. Virology **182:**168–176.
23. **Danner, K., and A. Mayr.** 1979. In vitro studies on Borna virus. II. Properties of the virus. Arch. Virol. **61:**261–271.
24. **Dayhoff, M. O., L. T. Hunt, and S. Hurst-Calderone.** 1978. Composition of proteins, p. 363–373. *In* M. O. Dayhoff (ed.), Atlas of protein sequences and structure. vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
25. **Delarue, M., O. Poch, N. Tordo, D. Moras, and P. Argos.** 1990. An attempt to unify the structure of polymerases. Protein Eng. **3:**461–467.
26. **de la Torre, J. C., K. M. Carbone, and W. I. Lipkin.** 1990. Molecular characterization of the Borna disease agent. Virology **179:**853–856.
27. **Duchala, C. S., K. M. Carbone, and O. Narayan.** 1989. Preliminary studies on the biology of Borna disease virus. J. Gen. Virol. **70:**3507–3511.
28. **Elford, W. J., and I. A. Galloway.** 1935. Filtration of the virus of Borna disease through graded collodion membranes. Br. J. Exp. Pathol. **14:**196–205.
29. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
30. **Feldmann, H., E. Mühlberger, A. Randolf, C. Will, M. P. Kiley, A. Sanchez, and H. D. Klenk.** 1992. Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle. Virus Res. **24:**1–19.
31. **Fu, Z. F., J. D. Amsterdam, M. Kao, V. Shankar, H. Koprowski, and B. Dietzschold.** 1993. Detection of Borna disease virus-reactive antibodies from patients with affective disorders by Western immunoblot technique. J. Affective Disord. **27:**61–68.
32. **Galinski, M. S.** 1991. Paramyxoviridae: transcription and replication. Adv. Virus Res. **39:**129–162.
33. **Gething, M.-J., J. Bye, J. Skehel, and M. Waterfield.** 1980. Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature (London) **287:**301–306.
34. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25:**263–269.
35. **Gupta, K. C., and D. W. Kingsbury.** 1985. Polytranscripts of Sendai virus do not contain intervening polyadenylate sequences. Virology **141:**102–109.
36. **Herman, R. C., M. Schubert, J. D. Keene, and R. A. Lazzarini.** 1980. Polycistronic vesicular stomatitis virus RNA transcripts. Proc. Natl. Acad. Sci. USA **77:**4662–4665.
37. **Herzog, S., and R. Rott.** 1980. Replication of Borna disease virus in cell cultures. Med. Microbiol. Immunol. **168:**153–158.
38. **Jackson, A. O., R. I. B. Francki, and D. Zuidema.** 1987. Biology, structure and replication of plant rhabdoviruses, p. 427–508. *In* R. R. Wagner (ed.), The rhabdoviruses. Plenum Publishing Corp., New York.
39. **Jones, R. W., and A. O. Tuckson.** 1990. Replication of sonchus yellow net virus in infected protoplasts. Virology **179:**815–820.
40. **Kao, M., A. N. Hamir, C. E. Rupprech, Z. F. Fu, H. Koprowski, and B. Dietzschold.** 1993. Detection of antibodies against Borna disease virus in sera and cerebrospinal fluid of horses in the USA. Vet. Rec. **132:**241–244.
41. **Kingsbury, D. W.** 1990. Orthomyxoviridae and their replication, p. 1075–1089. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology. Raven Press, New York.
42. **Kingsbury, D. W.** 1990. Paramyxoviridae and their replication, p. 945–962. *In* B. N. Fields and D. M. Knipe (ed.), Virology. Raven Press, New York.
43. **Kurilla, M. G., H. O. Stone, and J. D. Keene.** 1985. RNA sequence and transcriptional properties of the 3′ end of the Newcastle disease virus genome. Virology **145:**203–212.
44. **Leppert, M., L. Rittenhouse, J. Perrault, D. F. Summers, and D. Kolakofsky.** 1979. Plus and minus strand leader RNAs in negative-strand virus-infected cells. Cell **18:**735–747.
45. **Lipkin, W. I., T. Briese, and J. C. de la Torre.** 1992. Borna disease virus: molecular analysis of a neurotropic infectious agent. Microb. Pathog. **13:**167–170.
46. **Lipkin, W. I., G. H. Travis, K. M. Carbone, and M. C. Wilson.** 1990. Isolation and characterization of Borna disease agent cDNA clones. Proc. Natl. Acad. Sci. USA **87:**4184–4188.
47. **Ludwig, H., L. Bode, and G. Gosztonyi.** 1988. Borna disease: a persistent virus infection of the central nervous system. Prog. Med. Virol. **35:**107–151.
48. **Masters, P., and C. E. Samuel.** 1984. Detection of in vivo synthesis of polycistronic mRNAs of vesicular stomatitis virus. Virology **134:**277–286.
49. **McClure, M. A., K. J. Thibault, C. G. Hatalski, and W. I. Lipkin.** 1992. Sequence similarity between Borna disease virus p40 and a duplicated domain within the paramyxovirus and rhabdovirus polymerase proteins. J. Virol. **66:**6572–6577.
50. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
51. **Moscona, A., and M. S. Galinski.** 1990. Characterization of human parainfluenza virus type 3 persistent infection in cell culture. J. Virol. **64:**3212–3218.
52. **Moyer, S. A., S. Smallwood-Kentro, A. Haddad, and L. Prevec.** 1991. Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids. J. Virol. **65:**2170–2178.
53. **Ollis, D. L., P. Brick, R. Hamlin, N. G. Xuong, and T. A. Steitz.** 1985. Structure of the large fragment of Escherichia coli DNA polymerase I complexed with dTPM. Nature (London) **313:**762–766.
54. **Parvin, J. D., P. Palese, A. Honda, A. Ishihama, and M. Krystal.** 1989. Promoter analysis of influenza virus RNA polymerase. J. Virol. **63:**5142–5152.
55. **Penman, S.** 1969. Preparation of purified nuclei and nucleoli from mammalian cells, p. 35–48. *In* K. Habel and P. Salzman (ed.),

Fundamental techniques in virology. Academic Press, Inc., New York.

56. **Poch, O., B. M. Blumberg, L. Bougueferet, and N. Tordo.** 1990. Sequence comparison of five polymerases (L-proteins) of unsegmented negative-strand RNA viruses: theoretical assignments of functional domains. J. Gen. Virol. **71:**1153–1162.
57. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867–3874.
58. **Pyper, J. M., J. A. Richt, L. Brown, R. Rott, O. Narayan, and J. E. Clements.** 1993. Genomic organization of the structural proteins of Borna disease virus revealed by cDNA clone encoding the 38 kDa protein. Virology **195:**229–238.
59. **Richt, J. A., S. VandeWoude, M. C. Zink, J. E. Clements, S. Herzog, L. Stitz, R. Rott, and O. Narayan.** 1992. Infection with Borna disease virus: molecular and immunobiological characterization of the agent. Clin. Infect. Dis. **14:**1240–1250.
60. **Richt, J. A., S. VandeWoude, M. C. Zink, O. Narayan, and J. E. Clements.** 1991. Analysis of Borna disease virus-specific RNAs in infected cells and tissues. J. Gen. Virol. **72:**2251–2255.
61. **Roberts, B.** 1989. Nuclear location signal-mediated protein transport. Biochim. Biophys. Acta **1008:**263–280.
62. **Rott, R., S. Herzog, K. Bechter, and K. Frese.** 1991. Borna disease, a possible hazard for man? Arch. Virol. **118:**143–149.
63. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 10.26–10.37. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
64. **Schädler, R., H. Diringer, and H. Ludwig.** 1985. Isolation and characterization of a 14500 kD molecular weight protein from brains and tissue cultures persistently infected with Borna disease virus. J. Gen. Virol. **66:**2479–2484.
65. **Sippel, A. E.** 1973. Purification and characterization of adenorine triphosphate: ribonucleic acid adenyltransferase from E. coli. Eur. J. Biochem. **37:**31–40.
66. **Thierer, J., H. Riehle, O. Grebenstein, T. Binz, S. Herzog, N. Thiedemann, L. Stitz, R. Rott, F. Lottspeich, and H. Neimann.** 1992. The 24kD protein of Borna disease virus. J. Gen. Virol. **73:**413–416.
67. **Tordo, N., P. DeHaan, R. Goldbach, and O. Poch.** 1992. Evolution of negative-stranded RNA genomes. Semin. Virol. **3:**341–357.
68. **Tsurudome, M., H. Bando, M. Kawano, H. Matsumura, H. Komada, M. Nishio, and Y. Ito.** 1991. Transcripts of simian virus 41 (SV41) matrix gene are exclusively dicistronic with the tricistronic gene, which is also transcribed as a monocistronic. Virology **184:**93–100.
69. **VanBeek, N. A. M., D. Lohuis, J. Dijkstra, and D. Peters.** 1985. Morphogenesis of sonchus yellow net virus in cowpea protoplasts. J. Ultrastruct. Res. **90:**294–303.
70. **VandeWoude, S., J. A. Richt, M. C. Zink, R. Rott, O. Narayan, and J. E. Clements.** 1990. A Borna virus cDNA encoding a protein recognized by antibodies in humans with behavioral diseases. Science **250:**1278–1281.
71. **Wilde, A., and T. Morrison.** 1984. Structural and functional characterization of Newcastle disease virus polycistronic RNA species. J. Virol. **51:**71–76.
72. **Wong, T. C., and A. Hirano.** 1987. Structure and function of bicistronic RNA excoding the phosphoprotein and matrix protein of measles virus. J. Virol. **61:**584–589.