JOURNAL OF VIROLOGY, Dec. 1986, p. 1085–1097
0022-538X/86/121085-13$02.00/0

Vol. 60, No. 3

# Complete Nucleotide Sequence and Genome Organization of Bovine Parvovirus

KATHERINE C. CHEN,* BRUCE C. SHULL, ELIZABETH A. MOSES, MURIEL LEDERMAN, ERNEST R. STOUT, AND ROBERT C. BATES

*Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*

Received 16 June 1986/Accepted 8 September 1986

**We determined the complete nucleotide sequence of bovine parvovirus (BPV), an autonomous parvovirus. The sequence is 5,491 nucleotides long. The terminal regions contain nonidentical imperfect palindromic sequences of 150 and 121 nucleotides. In the plus strand, there are three large open reading frames (left ORF, mid ORF, and right ORF) with coding capacities of 729, 255, and 685 amino acids, respectively. As with all parvoviruses studied to date, the left ORF of BPV codes for the nonstructural protein NS-1 and the right ORF codes for the major parts of the three capsid proteins. The mid ORF probably encodes the major part of the nonstructural protein NP-1. There are promoterlike sequences at map units 4.5, 12.8, and 38.7 and polyadenylation signals at map units 61.6, 64.6, and 98.5. BPV has little DNA homology with the defective parvovirus AAV, with the human autonomous parvovirus B19, or with the other autonomous parvoviruses sequenced (canine parvovirus, feline panleukopenia virus, H-1, and minute virus of mice). Even though the overall DNA homology of BPV with other parvoviruses is low, several small regions of high homology are observed when the amino acid sequences encoded by the left and right ORFs are compared. From these comparisons, it can be shown that the evolutionary relationship among the parvoviruses is B19↔AAV↔BPV↔MVM. The highly conserved amino acid sequences observed among all parvoviruses may be useful in the identification and detection of parvoviruses and in the design of a general parvovirus vaccine.**

The mammalian parvoviruses are divided into two genera on the basis of the requirement for a helper virus for productive infection. The defective parvovirus, adeno-associated virus (AAV), requires coinfection with adenovirus or herpesvirus. The autonomous parvoviruses, which include bovine parvovirus (BPV), canine parvovirus (CPV), feline panleukopenia virus (FPV), H-1, Kilham rat virus (KRV), LuIII, and minute virus of mice (MVM), do not require a helper virus for infection (51).

Parvoviruses are DNA viruses with a linear single-stranded genome of ca. 5,000 nucleotides (nt). The complete nt sequences of AAV2 (serotype 2 of AAV) (50), MVMp (prototype of MVM) (4), MVMi (immunosuppressive variant of MVM) (3, 45), H-1 (43), and the partial nt sequences of CPV2 (type 2 of CPV) (42) and FPV (15) have been determined. Also, transcription has been studied in detail for AAV2, H-1, and MVMp (16). From these studies, several common features of the parvovirus genomes are apparent. (i) There are two large open reading frames (ORFs) within the plus strand (the strand with the same polarity as the mRNAs). (ii) The mRNAs from both ORFs are polyadenylated and 3′ coterminated at about map unit (m.u.) 95. (iii) The left ORF encodes one or more noncapsid proteins which are necessary for viral DNA replication (17, 23, 43, 52). (iv) The right ORF encodes the major capsid proteins of the virus as a nested set (4, 26, 43).

BPV is an autonomous parvovirus. Four virus-specific proteins have been reported: three capsid proteins (VP1, VP2, and VP3) of $M_r$ 80,000, 72,000 and 62,000 and one noncapsid protein (NP-1; $M_r$ 28,000) (30, 32). However, BPV shows very little homology with other autonomous parvoviruses (FPV, H-1, KRV, LuIII, and MVM) either by heteroduplex mapping or by immunological cross-reactivity of the capsid proteins, whereas the others as a group have significant homology among themselves (5, 48).

In this paper, the complete nt sequence of BPV is presented. Genome organization, possible coding regions for the viral proteins, and comparison with other parvoviruses are discussed. The sequence of BPV has several small regions of conservation with other parvoviruses which appear to be useful for understanding the evolutionary relationship of parvoviruses, in the design of vaccines, and for the clinical detection of parvoviruses.

Recently the nucleotide sequence of B19, a human autonomous parvovirus, was published (47). The relationship of B19 with other parvoviruses is presented in the section "Comparison with B19" at the end of the Discussion.

## MATERIALS AND METHODS

**Materials.** Restriction enzymes were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), and New England BioLabs, Inc. (Beverly, Mass.). *Escherichia coli* DNA polymerase I (Klenow fragment) and universal sequencing primer were obtained from Bethesda Research Laboratories. Deoxynucleotides and dideoxynucleotides were purchased from Pharmacia P-L Biochemicals, Inc. (Piscataway, N.J.). Terminal deoxynucleotide transferase, [α-$^{35}$S]dATP and [α-$^{35}$S]ddATP were obtained from New England Nuclear Corp. (Boston, Mass.). Various synthetic oligonucleotide primers for DNA-sequencing reactions were the kind gift of R. Foote, Oak Ridge National Laboratory.

**Cell culture and virus propagation.** BPV was propagated in bovine fetal lung cells grown in monolayer culture and maintained in Eagle minimal essential medium supplemented with 10% fetal calf serum as described by Parris and Bates (39).

**Isolation of viral DNA.** Single-stranded viral DNA was purified from virions isolated from BPV-infected cells. In-

* Corresponding author.

fected cells were frozen, thawed, and centrifuged at 15,000 × *g* for 15 min. The supernatant fraction was then centrifuged at 25,000 rpm in a Beckman SW27 rotor for 3.5 h at 4°C to pellet the virus. The pellet was resuspended in 50 mM Tris hydrochloride, pH 8.0, and loaded on a CsCl step gradient (CsCl [2 ml at 40%, 4 ml at 35%, and 2 ml at 30%] overlaid with 1 M sucrose in the same buffer), and the gradient was centrifuged at 34,000 rpm for 18 h at 15°C in a Beckman SW41 rotor. Virus particles at a density of 1.39 to 1.41 g/cm$^3$ (full particles) were visualized by light scattering, collected, and dialyzed against 50 mM Tris hydrochloride, pH 8.0. Viral DNA was isolated by proteinase K digestion (100 μg/ml, 55°C, 8 to 12 h) and phenol-chloroform extraction and ethanol precipitated. Since BPV encapsidates ca. 10% plus strands in bovine fetal lung cells (6) and the plus and minus strands readily form hybrids under the isolation conditions, the DNA thus prepared contained ca. 20% double-stranded DNA and ca. 80% single-stranded DNA of the minus polarity. Single-stranded DNA was separated from double-stranded DNA by centrifugation in an SW27.1 rotor (25,000 rpm, 15°C for 14 h) on a high-salt sucrose gradient (5 to 20% sucrose in 1 M NaCl, 10 mM Tris hydrochloride, pH 8.0, 1 mM EDTA, and 0.15% Sarkosyl). Gradient fractions were analyzed by electrophoresis on 1.0% agarose gels, and fractions containing single-stranded DNA and double-stranded DNA of monomer length were pooled separately.

The minus single-stranded DNA was replicated in vitro to double-stranded form by *E. coli* DNA polymerase I (Klenow fragment) as described previously (14), by using the 3′ palindrome of virion DNA as a self-primer. This is referred to as the in vitro replicative form (RF), and it was the source for cloning and sequencing of the DNA from the m.u. 5 to 100 region.

In vivo RF DNA which contains monomer RF with both fully extended and joined termini, as well as dimeric RF (2), was obtained from BPV-infected cells by the method of Hirt (24). This is referred to as the in vivo RF, and it was the source of DNA for cloning of the m.u. 0 to 18 region.

**Cloning of BPV DNA.** *Eco*RI fragment A (m.u. 18 to 92) and *Pst*I fragment A (m.u. 5 to 76) were obtained from in vitro RF and first cloned into pAT153. *Eco*RI fragment B (m.u. 0 to 18, from in vivo RF) and *Eco*RI fragment C (m.u. 92 to 100, from in vitro RF) were cloned into pUC8 after the attachment of *Sal*I linkers to the ends. *Sal*I linkers were chosen because there is no *Sal*I site within the BPV genome. The cloned fragments were then subcloned into M13mp18 and M13mp19 vectors for sequencing (55).

S1 nuclease-resistant fragments of single-stranded virion DNA were cloned into *Sma*I-digested M13mp18 vectors to permit direct sequencing of the terminal duplex regions in the virion DNA.

All recombinant pUC plasmids and M13 phages were propagated in *E. coli* JM107, whereas pAT153 recombinant plasmids were propagated in *E. coli* HB101. Transformation was performed as described by Hanahan (21).

**Sequencing of BPV DNA.** The chain termination method of Sanger et al. (46) was the primary method used for sequencing. A 15-mer universal primer was used for most runs. Several synthetic oligonucleotides were used to sequence overlapping clones and to obtain extended sequence data from larger fragments. Buffer gradient gels and [α-$^{35}$S]dATP were used according to the method of Biggin et al. (12).

The 5′ end (m.u. 100) region was also sequenced by the method of Maxam and Gilbert (34). The pUC8 clones containing the *Eco*RI C fragment (m.u. 92 to 100) were linearized at the *Sal*I site, end labeled with [α-$^{35}$S]ddATP and terminal deoxynucleotide transferase (44), and sequenced by the chemical degradation method.

**Computer analysis.** DNA sequences were analyzed by the Pustell and Kafatos (41) sequencing programs available commercially from International Biotechnologies Inc. (New Haven, Conn.). Locally homologous protein sequences were detected with the program written by Goad and Kanehisa (19) which was implemented in the ICR SEQuence program package obtained from Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

## RESULTS

In this paper, the convention of Armentrout et al. (1) is used: the 3′ terminus of the minus-strand DNA is referred to as the 3′ end or the map origin, and the map is drawn with the 3′ end at the left. Additionally, the term virion DNA will be used to refer to the minus strand, since the method of purification yielded minus-strand DNA from total virion DNA.

**Sequence determination.** We determined the complete nt sequence of BPV to be 5,491 bases long. The sequencing strategy used is shown in Fig. 1. As described in Materials and Methods, the internal fragments of BPV (i.e., *Pst*I fragment A, m.u. 5 to 18; *Eco*RI fragment A, m.u. 18 to 92) were cloned from in vitro RF into pAT153. The 5′ end (*Eco*RI fragment C, m.u. 92 to 100) was also cloned from in vitro RF, whereas the 3′ end (*Eco*RI fragment B, m.u. 0 to 18) was cloned from in vivo RF into pUC8 after the addition of *Sal*I linkers. Each was then subcloned into M13mp18 and M13mp19 vectors for shotgun sequencing. However, the 5′ end clones were found to have deletions in the terminal sequences nt 5484 to 5491 were deleted). To obtain sequence data for the extreme ends, we took advantage of the fact that the terminal sequences of the single-stranded virion DNA form hairpin loop structures and that the stems of the hairpins are resistant to S1 nuclease digestion (11, 22). Cloning and sequencing of the S1-resistant fragments of single-stranded virion DNA gave the sequences at nt 11 to 44, 5371 to 5422, and 5440 to 5491. The latter two sequences correspond to the stems of the 5′ terminal hairpins (see Fig. 3B). The reason that the first 10 bases (nt 1 to 10) were not found in the S1-resistant fractions as would be expected (see Fig. 3A) is probably due to the high A-T content and hence low melting temperature ($T_m$) in this region.

The complete sequence of the genome is given in Fig. 2. The sequence shown is that for the plus strand, which has the same polarity as BPV mRNAs. DNA sequence was obtained from both strands and across all restriction sites. The 5′ end sequence was also confirmed by the Maxam and Gilbert method (34). The sequence data agree with the published restriction map of BPV (14).

**Terminal structures.** The terminal hairpin structures of BPV virion DNA (minus strand) are shown in Fig. 3. In agreement with other autonomous parvoviruses (H-1, MVMp, MVMi) (3, 22, 45), the 3′ terminus can form a T-shaped configuration; the 5′ terminus can form a U-shaped configuration. However, the 3′ terminus is longer (150 nt versus 115 and 116 nt) and the 5′ terminus is shorter (121 nt versus 207 and 242 nt) than those of the rodent parvoviruses. It was reported (2) that the 3′ terminal sequence of the rodent parvovirus DNA is unique, whereas the 5′ terminal sequence is heterogeneous. The 5′ end possesses two mutually exclusive sequences that are inverted complements of each other (denoted as the flip and the flop orientations). We analyzed four independent 3′ end clones of BPV and ob-

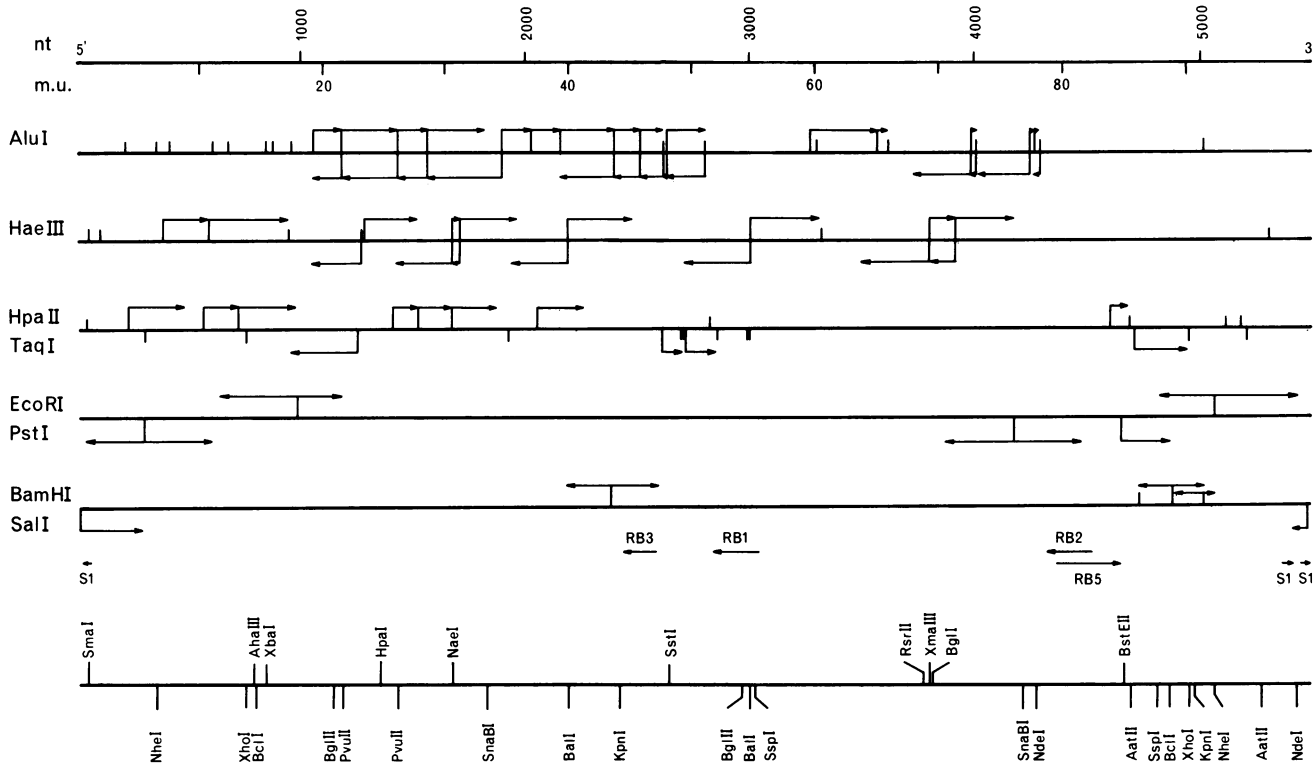


FIG. 1. Sequencing strategy and restriction map of BPV. The sequences were determined in the direction of the arrows. 0 m.u. corresponds to the 5' end of the plus strand (3' end of the minus strand). The RB series are synthetic oligonucleotide primers used in some sequencing reactions. The S1 series are a set of clones obtained by S1 nuclease treatment of virion DNA (minus strand) and cloned into M13mp18 vectors. The restriction enzymes indicated by the bars above the last line are those with single cleavage sites, and those below the line have two cleavage sites in BPV.

tained identical sequences. For the 5' end, we analyzed only one clone, the sequence of which is given in Fig. 2. However, *Nde*I digestion (CA/TATG) of the end-labeled in vitro RF (which should cut at nt 5437 in the flip orientation, and at nt 5422 in the flop orientation) yielded two fragments of equal intensity of the expected sizes (55 and 70 bases), thus demonstrating the existence of two sequence orientations in the 5' end of BPV (data not shown). Recently, we constructed several full-length genomic clones of BPV by cloning reannealed plus-strand and minus-strand DNAs from purified viruses into pUC8 plasmids. These recombinant clones were infectious when transfected into bovine fetal lung cells, and the 5' end sequence of one (pGC17) is in the flop orientation (B. C. Shull, unpublished results).

**Genomic organization.** A diagram of the stop codons and ATG codon in each of the three reference frames of BPV plus-strand DNA is shown in Fig. 4. There are three large and five small ORFs. The location, the coding capacity, and the putative amino acid sequence encoded by each of the three large ORFs (i.e., left, mid, and right ORF) are summarized in Fig. 5. When the minus-strand DNA was analyzed, like all parvoviruses studied, no ORFs of significant sizes were found. Indeed, RNA from BPV-infected cells is complementary only to the minus strand (P. R. Burd, Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, 1982).

A search for the consensus promoter sequence (TATAA) in the BPV sequence revealed three possible promoters at nt 251 (m.u. 4.5, TATAAA), nt 705 (m.u. 12.8, TATAAT), and nt 2128 (m.u. 38.7, TATTAA). In contrast, only two promoters (at m.u. 4 and 38) have been demonstrated for the rodent autonomous parvoviruses H-1 and MVM (8, 40), whereas three promoters were identified (at m.u. 5, 18, and 39) for the defective parvovirus AAV (50).

Bensimhon et al. (9), after comparing the promoter sequences of simian virus 40 early genes and those of herpes simplex virus *tk* genes, proposed that the promoter recognition and initiation by eucaryotic RNA polymerases do not depend on the TATA consensus sequence alone. They suggested that the eucaryotic promoters should have three basic components: (i) an enhancer region of moderately high local stability, roughly 100 base pairs (bp) upstream of the cap site; (ii) an activation region (a G-C-rich, high stability domain), 50 to 70 bp upstream of the cap site; and (iii) an A-T-rich domain (TATA box), 30 ± 5 bp upstream of the cap site, that traps and positions the RNA polymerase for initiation of transcription.

We analyzed the promoters of AAV2, MVMp, and H-1 which have been demonstrated unambiguously to be functional by primer extension studies (16) for the presence of the above three components (9). All the known promoters of AAV2, MVMp, and H-1 satisfy these criteria (Table 1). When this analysis is extended to the possible promoters of BPV, all three satisfy the criteria (Table 1) and hence may all be functional. In addition to the three possible promoters mentioned above, there is a TATAAAA sequence at nt 3922 (m.u. 71.4). This one, however, does not have all three components described by Bensimhon et al. (9). Experiments to map the BPV promoters are under way.

The most characteristic feature of the 3' ends of eucaryotic mRNAs is the presence of the sequence AAUAAA at about 10 to 30 nt before the poly(A) tract (53). This sequence


is an essential but not sufficient signal for polyadenylation. Several additional signals have been suggested, such as downstream G/T clusters (13), secondary structure in the vicinity of AAUAAA (35, 54), and a CAYUG sequence upstream or downstream from the poly(A) site (10), but no consistent picture of the absolute requirements has emerged. There are six AATAAA sequences in the BPV genome at m.u. 2.5, 60.0, 61.6, 64.6, 74.8, and 98.5. The one at m.u. 2.5 is clearly not functional, as it precedes any possible promoters. The one at m.u. 74.8 also may not be functional because it is in the middle of the right ORF and is about 1 kilobase (kb) away from the ends of the left and the mid ORFs. The two poly(A) signals at m.u. 61.6 and 64.6 may be functional, as one has G/T clusters (m.u. 61.6), the other has a CAYUG sequence (m.u. 64.6), and both have secondary structures (m.u. 61.6 and 64.6). Our preliminary RNA mapping data


FIG. 2. Complete nt sequence of BPV (plus strand).

indicate that the 1.1-kb RNA coding for NP-1 terminates in this region (Burd, Ph.D. thesis). The poly(A) signal at m.u. 98.5 must be functional because it is the only one available for the RNA transcribed from the right ORF, and it does have secondary structure and a CAYUG sequence around the AATAAA site. The poly(A) signal at m.u. 60.0 may not be functional because it does not have G/T clusters, a CAYUG sequence, or secondary structure.

**Assignment of coding domains.** By analogy with the genomic organizations reported for other parvoviruses (AAV, H-1, MVM), and also by the amino acid homologies observed for the putative proteins (see section on Homologies), we assign the left ORF of BPV to code for the major part of the nonstructural protein NS-1, the right ORF to code for the major parts of the three capsid proteins, and the mid ORF to code for the major part of NP-1.

NS-1 is coded by the left half of the genome in AAV, H-1, and MVM (17, 43) and is necessary for viral DNA replication (23, 52). It is detected by immunoprecipitation with homologous antiserum from in vitro translation reactions of RNAs from MVM-infected (17) and H-1-infected cells (37). Recently, we observed two proteins of $M_r$ 75,000 and 82,000, one or both of which may be the BPV equivalent of NS-1 (M. Lederman and R. C. Bates, unpublished results). These proteins are detected as immunoprecipitation products of BPV-infected cell lysates and of in vitro translations of RNA from BPV-infected cells by using antiserum from a calf experimentally infected with BPV. These proteins are synthesized early in infection, and their rate of synthesis declines as the capsid proteins accumulate.

In all the parvoviruses studied, the capsid proteins are encoded by the right ORF as a nested set with the amino acid sequence of the smaller being largely contained in the next larger protein (27). In the autonomous parvoviruses MVM


FIG. 3. Terminal structures of BPV DNA (minus strand).

and H-1, there are two coded capsid proteins VP1 and VP2' ($M_r$ 84,000 and 62,000). The two proteins are translated from two differently spliced mRNAs generated from the promoter at m.u. 38 (3, 29, 38, 40). VP1 initiates at m.u. 44 with the first ATG codon in its mRNA, and it contains the conserved amino acid sequence NPYL region (see section on Homologies). VP2' initiates with its first ATG codon at m.u. 54 (the ATG codon at m.u. 44 for VP1 initiation is spliced out). The codon at m.u. 54 is in the most favorable sequence context for an initiation codon, according to Kozak (AnnATGG) (28), and it is located 75 nt upstream of a glycine-rich region (4, 43). The defective parvovirus AAV has three capsid proteins VP1 to 3 ($M_r$ 92,000, 72,000, and 60,000). They are translated from two differently spliced mRNAs, both generated from the promoter at m.u. 39 (16, 26). VP2 and VP3 are translated from one spliced mRNA of 2.3 kb initiated at different sites. VP2 initiates at m.u. 56 with ACG as the initiation codon; VP3 initiates at m.u. 60 with the first ATG

TABLE 1. Analysis of promoter sequences in parvoviruses

| Virus | Promoter | Enhancer | Activator | TATA box |
|---|---|---|---|---|
| AAV2 | P (5.5)[a] | GGGAGGTC (186, −99)[b] | GCGACACC (231, −54)[b] | TATTTAA (255, −30)[b] |
| | P (18.0) | GCTCCCCA (791, −82) | GGGCGTGG (820, −53) | TATTTAA (843, −30) |
| | P (39.0) | GGTGGAGC (1757, −94) | CCCGCCCC (1803, −48) | ATATAAG (1821, −30) |
| MVMp | P (3.4) | GCGGTTCA (121, −84) | GGCGCGAA (144, −61) | TATATAA (175, −30) |
| | P (38.4) | GGGGCAAA (1921, −85) | GGGCGGAG (1951, −55) | TATAAAT (1976, −30) |
| H-1 | P (3.5) | GCGGTTCA (123, −85) | GGCGGGAA (147, −61) | TATATAA (178,−30) |
| | P (38.2) | GGGGCAAA (1924, −85) | GGGCGGAG (1954, −55) | TATAAAT (1979, −30) |
| BPV | P (4.5) | GGCGAGGA (179, −100) | CCGGGGGG (227, −52) | TATATAA (249, −30) |
| | P (12.8) | CGAGCCGA (645, −89) | GGGGGAGA (681, −53) | ATATAAT (704, −30) |
| | P (38.7) | CCGGGCAT (2054, −104) | GCGCGCAC (2085, −73) | TATTAAA (2128, −30) |

[a] The number in parentheses after the promoter P indicates its map position.

[b] The first number in the parentheses after the sequence indicates the position of the sequence in the genome; the second number indicates the position of the sequence relative to the cap site. Since many RNA cap sites are unknown, we assume that they are 30 bp downstream from the TATA box.

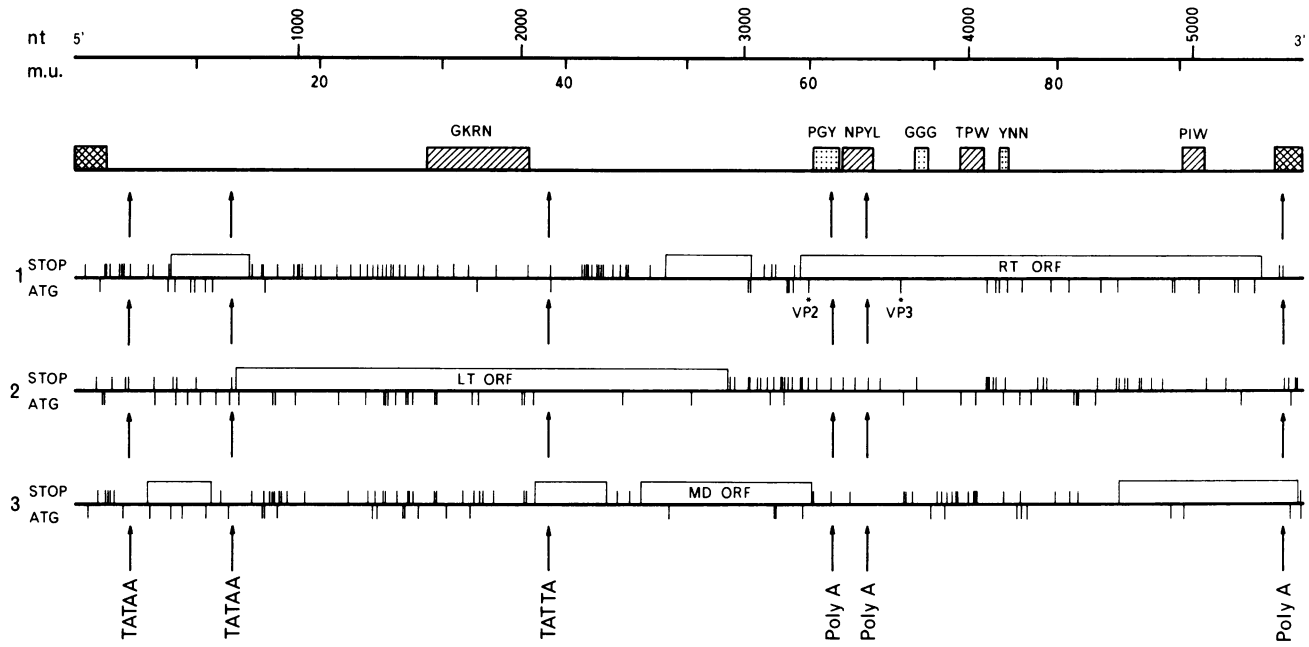


FIG. 4. ORFs and conserved regions in the BPV genome. The lower three lines correspond to the three possible reading frames (1, 2, and 3) of the plus-strand sequence, with the stop codons and the ATG codons indicated by short bars. Three large ORFs (LT ORF, MD ORF, and RT ORF) and five small ORFs are shown by open blocks. The second line from the top represents regions of highly conserved amino acid sequences in BPV as compared with other parvoviruses. ▨, Regions of high amino acid homology among all parvoviruses (AAV2, BPV, MVMp); ▩, regions of high amino acid homology between the autonomous parvoviruses (BPV, MVMp) only; ▩, regions of terminal hairpin structure of BPV.

codon in this mRNA (7, 26). The initiation codon for VP3 also conforms to the Kozak consensus sequence for initiation. The exact splicing pattern of the mRNA for VP1 is not certain, but VP1 contains the NPYL region (see section on Homologies).

BPV has three coded capsid proteins of $M_r$ 80,000, 72,000 and 62,000 (30). An ATG codon located similarly to the initiation codon of VP2′ in MVM and H-1 (i.e., with a Kozak consensus sequence and 75 nt upstream of a glycine-rich region) is found at nt 3697 (m.u. 67.3), which presumably is used for the initiation of VP3 in BPV. Nt 3697 is the second ATG codon in the right ORF. When we searched for the initiation of VP2 in the region downstream from the NPYL region to nt 3697, no ACG codon analogous to the initiation codon for VP2 in AAV or other ATG codon was found. Therefore, we assign tentatively the initiation site for VP2 at nt 3286 (m.u. 59.8), the first ATG codon in the right ORF. The initiator for VP2 (TTGATGC) is not in an as favorable a context as the initiator for VP3 (AACATGG) for protein initiation, which may explain the relative abundance of the two proteins. The predicted molecular weights for VP2 and VP3 are 75,000 and 60,000, whereas the observed values computed from electrophoretic mobilities on sodium dodecyl sulfate-polyacrylamide gels are 72,000 and 62,000. If the assignment of nt 3286 for VP2 initiation is correct, then VP2 will contain the NPYL region. In other parvoviruses, the largest capsid protein is the one that contains the NPYL region. This will make BPV different from the others by having one extra larger capsid protein VP1.

A computer search for the possible splice donor and acceptor sites (36) for VP2-VP3 mRNA revealed two possible donor sites at nt 2336 (ACA/GTGAGT) and 2441 (CCA/GTGAGT) and one possible acceptor site at nt 3241 (TTCGTTGCAG/A) just before the start of the right ORF. In either case, VP2 will initiate with the second ATG (nt 3286) and VP3 will initiate with the third ATG (nt 3697) in the spliced mRNA (the first ATG at nt 3261 is soon followed by in frame termination codon at nt 3300).

Previous studies by partial proteolysis indicate that the amino acid sequence in VP2 is largely contained in that of VP1 (30). With the sequence data alone, we cannot assign the coding region for the N-terminal amino acid sequence specific for VP1, because there are many small ORFs and possible splice donor and acceptor sites. Wherever that coding region may be, it must overlap partially with the coding region for NP-1 (see below).

NP-1 is a highly phosphorylated noncapsid protein of apparent $M_r$ 28,000. It has been detected in the nuclei of BPV-infected cells in association with chromatin and as an in vitro translation product of BPV-specific RNAs. NP-1 shares structural homology with VP1 (but not with VP2 or VP3) as demonstrated by partial proteolysis (30, 32). Polyclonal antibodies to VP1 react with NP-1, but polyclonal antibodies to VP2 or VP3 do not (R. C. Bates, unpublished results). An abundant BPV-specific RNA species of 1.1 kb, when translated in vitro, gave a protein of $M_r$ 28,000 that could be immunoprecipitated with anti-NP-1 antibody (M. Lederman, unpublished results). When this 1.1-kb RNA (supposedly mRNA for NP-1) was analyzed by the Berk and Sharp S1 nuclease mapping method, it was found to be spliced with a 0.3-kb leader and a 0.8-kb main body (Burd, Ph.D. thesis). In DNA-binding assays, NP-1 binds specifically to the 3′ end region (m.u. 0 to 5) of the RF form of BPV DNA (M. Lederman, B. C. Shull, E. R. Stout, and R. C. Bates, J. Gen. Virol., in press).

Several lines of evidence suggest that the mid ORF may be the major coding region for NP-1. (i) RNA size. By using a computer search, a possible splice acceptor site was found at

```
LOCUS   BPVLTAA       729 AA
ORIGIN  TRANSLATION OF BPV LEFT ORF (BP 731-2920; M.U. 13.3-53.1; FRAME 2)
        CALCULATED MOLECULAR WEIGHT =  81523.88      ESTIMATED pI =  5.75

   1 PDKMQPSRGL YSALHSFKHL LITIRQRTHA PDTVLSRDEL NAWNWLQKIA QRVMRGDMDD
  61 IVDILKCRKA NGTLVAQAIN GTEFITRYML PKNRKVADTV LTRHTTPEQS YDSTWGKTYG
 121 FAVCNGETVS EFTRKDLWKV LYNIYTAHPA ENMLNSNPSV WGDLPRVSAN RIDADDAEAR
 181 SRPIKLSRKQ KIMAEVIQRA TDGLLLTYND LVVHLSDLML MLEGMPGGSK TAEQLLTMIH
 241 IKLCAKYNAY EFMLMKTPAT QNMNPGAPHY DCQGNLVFKL LNLQGYNPWQ VGHWLVMMLS
 301 KKTGKRNSTL FYGPASTGKT NLAKAICHAV GLYGCVNHNN KQFPFNDAPN KMILWWEECI
 361 MTTDYVEAAK CVLGGTHVRV DVKHKDSREL PQIPVLLSSN HDVYTVVGGN ATFGVHAAPL
 421 KERITQMNFM KQLPNTFGEI TPGMISNWLS HCAHIHQEHL SLEGFAIKWD VQSVGNSFPL
 481 QTLCPGHSQN WTFSENGVCW HCGGFIQPTP ESDTDSDGDP DPDGAVAGDS DTSANSESTV
 541 SFSSNDSGLG SVTSSAPSVP DRAEEIEEIP SECLEWMREE VDRLSAHDIN SLAHQATGFI
 601 LDPIPEEPEE GERDLAREDA EPEASTSHTP ATKRARVEEG EPWDGTQPIT EGDWIDFESR
 661 QKRRRLEREE KGGEDEDMEV QESDPSAWGE KLGIVEKPGE EPIVLYCFET LPESDEEGDS
 721 DKENKTHTV-


LOCUS   BPVMDAA       255 AA
ORIGIN  TRANSLATION OF BPV MID ORF (BP 2535-3302; M.U. 46.1-60.1; FRAME 3)
        CALCULATED MOLECULAR WEIGHT =  29797.99      ESTIMATED pI = 10.80

   1 IPFQKNQKKG SGIWLERTPS QKHRRVTLQL PKELAWRKVS RGMERSRSPR ETGSTSSRDK
  61 SDADWSERRR EERTRTWKSR SPIRARGERS WGSWRSREKN QSSSTASRPY QKATRKETAT
 121 KKTKHTPFNV FSAHRALSKT DLQFCGFYWH STRLASKGTN EIFNGLKQSF QSKAIDGKLD
 181 WEGVRELLFE QKKCLDTWYR NMMYHFALGG DCEKCNYWDD VYKKHLANVD TYSVAEEITD
 241 SEMLEAAEAV DAANQ-


LOCUS   BPVRTAA       685 AA
ORIGIN  TRANSLATION OF BPV RIGHT ORF (BP 3250-5307; M.U. 59.2-96.6; FRAME 1)
        CALCULATED MOLECULAR WEIGHT =  76622.14      ESTIMATED pI = 10.42

   1 LILKCWKRQK LLMPPTNKAN SKKGLTLPGY NYLGPFNSLF AGAPVNKADA AARKHDFGYS
  61 DLLKEGKNPY LYFNTHDQNL IDELKDDTSF GGKLARGVFQ IKKALAPALP GTSKGGDRAL
 121 KRKLYFARSN KGAKKANREP APSTSNQQNM EVSNDIPNDE AGNQPIELAT RSVVGSGSVG
 181 GGGRGGSGVG YSTGGWTGGT IFSENIVVTK NTRQFICDIK NGHLYKSEVL NTGDTAHRQY
 241 AITTPWSYFN FNQYSSHFSP NDWQHLVNDY ERFRPKAMIV RVYNLQIKQI MTDGAMGTVY
 301 NNDLTAGMHI FCDGDHRYPY VQHPWDDQCM PELPNSIWEL PQYAYIPAPI SVVDNNTTNT
 361 VEEHLLKGVP LYMLENSDHE VLRNGRIYRI YIQLWRLRMD RKQHHIQHAS DDVQSTGQKQ
 421 KNLLIQRTKQ PNKQRFQNAA LRTSNWMSGP GIARGTHNAT LQTQSAGALV TMVTNGADVS
 481 GVRAVRVGYS TDPIYGGQQP ESDLLRLRYS ASAAEGQQNP ILENAARHTF TREARTKLIT
 541 GSNGADGDYK EWWMLPNQMW DSAPISRYNP IWVKVPRVNR KTLLDTQDGS IPMSHPPGTI
 601 FIKLARIPVP GNGDSFLNIY VTGQVSCEVV WEVEKRGTKN WRPEYMHSAT NMSVDAYTIN
 661 NAGVYAGAVQ NADVMQTRFN HHKVL-
```

FIG. 5. Amino acid sequences translated from the three large ORFs in BPV.

the start of the mid ORF (nt 2535, TTCATCCTAG/A). The mRNA for NP-1 may start from one of the three possible promotors so that the N terminus of the protein is coded by the sequence between the first ATG and the splice and then splice to nt 2535 so that the mid ORF becomes the reading frame for the remainder of the protein. It may terminate at the poly(A) signal at nt 3385 (m.u. 61.6), giving it a main body of 850 nt after the splice site. This would be consistent with the previous observation of Burd. (ii) Protein size. The putative protein encoded by the mid ORF has $M_r$ 30,000, which is similar in size to NP-1. (iii) Charge. The pI of the putative protein is estimated to be 10.8; it should be a basic protein (56 Lys + Arg + His versus 33 Asp + Glu), which may explain the observed DNA-binding capability of NP-1. (iv) Serine content. NP-1 is a highly phosphorylated protein, and our preliminary data indicate that the serine residues are phosphorylated (M. Lederman, E. R. Stout, and R. C. Bates, unpublished results). The putative protein is also rich in serine (27 serines in a total of 255 amino acids). RNA mapping studies with labeled restriction fragments to locate the coding region for NP-1 are in progress.

**DNA and amino acid homologies of BPV with other parvoviruses.** The overall DNA homologies of the BPV left and right ORFs with those of the other parvoviruses are not significant (ca. 28%). However, several regions of high local homology are observed when the amino acid sequences encoded by the left and right ORFs are compared. The comparisons among AAV2, BPV, and MVMp are depicted in Fig. 6 and 7. Since CPV, FPV, and H-1 are highly homologous to MVM, their sequences are not shown in the figures.

In the left ORF, there is a region of 55 amino acids within the region marked as GKRN that shows a 60% homology among AAV, BPV, and MVM (Fig. 4 and 6). The GKRN region (87 amino acids), located at m.u. 29.6 to 34.4, is in the coding region for the nonstructural protein NS-1. The biological significance of this conserved region is not known. It contains a sequence analogous to the nuclear targeting signal for a yeast mating type protein (20), which may act as the signal for the nuclear transport of NS-1 (31).

Six small blocks of strong amino acid homologies are found in the right ORF (Fig. 4 and 7). Three are conserved in both the defective and autonomous parvoviruses (marked as NPYL, TPW, and PIW in Fig. 4), whereas the other three are conserved among the autonomous parvoviruses (marked as PGY, GGG, and YNN; for the PGY region, see the Discussion [The right ORF] also).

The mid ORF of BPV is located at m.u. 46 to 60. There are several smaller ORFs (besides the major left and right ORFs) in the corresponding locations in the genomes of AAV (m.u.

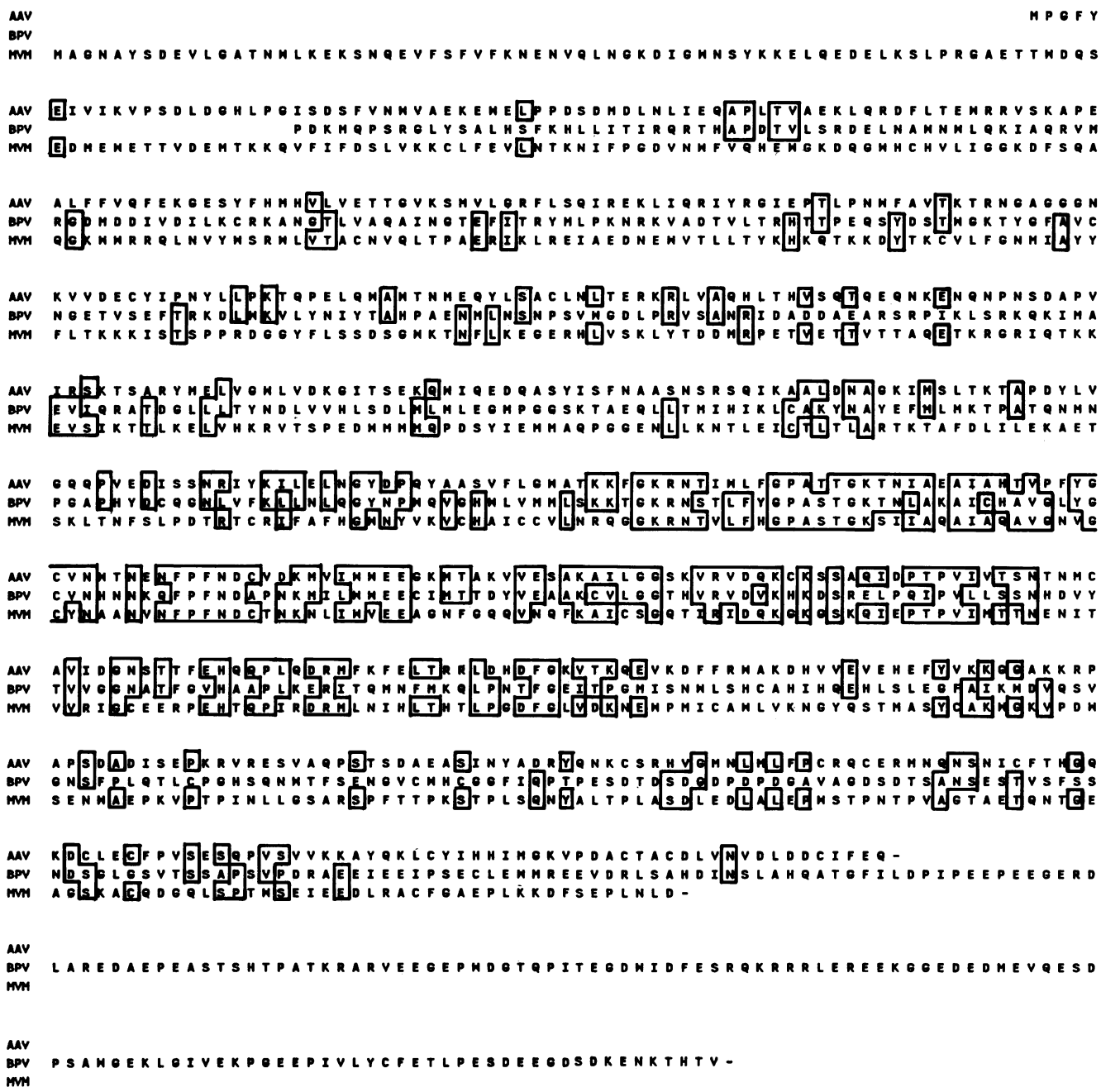

FIG. 6. Homology of the translated left ORFs among parvoviruses AAV2, BPV, and MVMp. Homologous regions are enclosed by boxes.

39.2 to 46.3, 46.8 to 53.8, and 49.6 to 56.5) and MVM (m.u. 37.6 to 44.7 and 46.6 to 52.2) also. These smaller ORFs, when combined, would have coding capacity similar to that of the mid ORF of BPV. The small ORF of MVM at m.u. 37.6 to 44.7 (denoted as mid ORF A) is known to contain sequences coding for the C-terminal region of NS-2, a nonstructural protein of $M_r$ 24,000, similar in size to NP-1 (18).

We searched the mid ORF of BPV and the three mid ORFs of AAV for regions homologous to the C-terminal region of NS-2 (i.e., mid ORF A in MVM), scoring conservative amino acid changes like D↔E, K↔R, T↔S, and A↔V↔I↔L as matches, and none were found. In fact, no such regions were found anywhere in the genome of BPV or AAV. When the mid ORF of BPV was compared with that of the other parvoviruses, there was no homology of the putative BPV mid ORF protein with any peptides that may be encoded anywhere in the genome of AAV, MVM, or H-1. The partial nt sequences of CPV and FPV were searched, and the results were the same. Likewise, when the three small mid ORFs of AAV were compared with those of other parvoviruses, no homologies were found either.

The analysis of comparisons of the homologous regions of BPV with the other parvoviruses is summarized in Table 2. The similarity coefficient ($S_{AB}$) between two proteins A and B is defined as (2 × number of identical residues in A and

TABLE 2. Similarity coefficients[a] between putative parvoviral proteins of AAV, BPV, and MVM

| ORF | Virus | $S_{AB}$ for proteins of: | | |
|---|---|---|---|---|
| | | AAV | BPV | MVM |
| Left | AAV | 1 | 0.15 | 0.15 |
| | BPV | 0.15 | 1 | 0.13 |
| | MVM | 0.15 | 0.13 | 1 |
| Right | AAV[b] | 1 | 0.27 | 0.18 |
| | BPV | 0.27 | 1 | 0.20 |
| | MVM | 0.18 | 0.20 | 1 |
| Mid | AAV | 1 | NS[c] | NS |
| | BPV | NS | 1 | NS |
| | MVM | NS | NS | 1 |

[a] The similarity coefficient $S_{AB}$ between two proteins A and B is defined as (2 × number of identical residues in A and B)/[(number of residues in A) + (number of residues in B)].

[b] The putative amino acid sequence used for AAV2 is translated from nt 2191 to 2436 in frame 1 and then nt 2438 to 4324 in frame 2 (see the Discussion [The right ORF]).

[c] NS, Not significant.

B)/[(number of residues in A) + (number of residues in B)].

## DISCUSSION

Although the genomic organizations of all parvoviruses studied to date are similar, the extent of DNA sequence homology among them varies. Previous heteroduplex mapping by electron microscopy (5) indicated that the defective parvovirus AAV has little DNA sequence homology with the autonomous parvoviruses (BPV, H-1, LuIII, KRV, and MVM). The autonomous parvoviruses H-1, LuIII, KRV, and MVM have greater than 70% of their sequences conserved, whereas BPV, like AAV, has little homology with these viruses. Immunological cross-reactivity leads to the same conclusion. AAV and BPV do not cross-react with other autonomous parvoviruses, whereas CPV, FPV, H-1, KRV, LuIII, and MVM have some cross-reactivity (48).

From our sequence data and the published sequence data on AAV, CPV, FPV, H-1, and MVM (4, 15, 42, 43, 50), the overall DNA homologies of the BPV left and right ORFs with those of the other parvoviruses (AAV, CPV, FPV, H-1, and MVM) are not significant (approximately 28%), whereas the sequences in CPV, FPV, H-1, and MVM are quite similar to each other, with greater than 50% conservation. Based on the evidence given above, we suggest that there are three groups of parvoviruses: the defective parvoviruses with AAV as a representative, the autonomous parvovirus BPV, and the other autonomous parvoviruses (CPV, FPV, H-1, KRV, LuIII, and MVM). If the human parvovirus B19 is also considered (47; see Comparison with B19 at the end of the Discussion), it would represent an additional group.

Several points merit discussion when the sequence of BPV is compared with those of other parvoviruses.

**Terminal structures.** The terminal palindromes of BPV show very little homology with the other autonomous parvoviruses at the level of DNA sequences, but the secondary structures of the termini are highly conserved. This suggests that the secondary structure rather than the primary sequence is important in the function of the palindromes of the autonomous parvoviruses. This is in agreement with the data of Lefebvre et al. (33), who noted that changes in the primary sequence of AAV terminal palindromes are tolerated if the secondary structure is conserved. The 3′ end palindrome of BPV shows two G-C-rich stems, a feature conserved in all parvoviruses sequenced to date. The high G-C content of these stems may be important in stabilizing the 3′ terminal palindrome. The sequence TAAAAAT at or near the viral 3′ terminus is conserved among all autonomous parvoviruses sequenced to date. The conservation of this sequence element suggests a functional role in the replication cycle of the autonomous parvoviruses. It is interesting to note that the 3′ terminus of the autonomous parvoviruses is A-T rich with twelve consecutive A or T residues in BPV and 9 of the first 10 bases being A or T in H-1 and MVM.

**The left ORF.** In the left ORF, the region within the one marked as GKRN (55 amino acids) is highly conserved among all parvoviruses and this conserved sequence may be used as a diagnostic probe for parvovirus identification.

**The right ORF.** There are three regions in the right ORF that are conserved among all parvoviruses (marked as NPYL, TPW, and PIW in Fig. 4), and three are conserved among the autonomous parvoviruses (marked as PGY, GGG, and YNN).

The sequence NPYL (m.u. 62.6 to 65.1; 47 amino acids) is present in the largest capsid protein VP1 in AAV, CPV, FPV, H-1, and MVM. Based on the assignment of coding domains presented in the Results section, we propose that VP2 of BPV also contains this region. The significance of this highly conserved sequence in the large capsid protein is unknown at present.

A glycine-rich region (GGG, m.u. 68.7 to 69.8; 20 amino acids) is highly conserved in all of the sequences reported for the autonomous parvoviruses. It is located about 75 nt downstream from the initiation codon for the smallest coded capsid protein VP3. A proteolytic cleavage site (which generates the smallest capsid protein sometimes detectable in the mature virus, i.e., VP4 in BPV and VP2 in MVM and H-1) is often found in the vicinity of the GGG region (15, 42). It has been suggested (42) that a run of glycines distorts the alpha helical structure of the protein; the resulting extended region could be the substrate for the proteolytic cleavage which generates the smallest capsid protein. A less-conserved form of this sequence is found in AAV (31); it is only 9 nt downstream from the initiation methionine for VP3, and VP3 is not cleaved in the mature virion.

The PGY region (m.u. 60.4 to 62.2; 33 amino acids) is conserved among autonomous parvoviruses but not in the defective parvovirus AAV2, if the published sequence data of AAV2 (50) is used for the translation of the right ORF (nt 2261 to 4324, frame 2). However, a PGY region is found in a different frame (frame 1) as shown below:

```
AAV  nt 2332  GLVLPGYKYLGPFNGLDKGEPVNEADAAALEHVQS
BPV  nt 3319  GLTLPGYNYLGPFNSLFAGAPVNKADAAARKHDFG
MVM  nt 2398  GWVPPGYKYLGPGNSLDQGEPTNPSDAAAKEHDEA
```

This suggests that frame 1 rather than frame 2 in this region (nt 2332 to 2427) is used in the translation of VP1 in AAV. The same suggestion has been made by Shade et al. (47). The data of Janik et al. (26) also support this hypothesis. They showed that a 186-bp deletion of *Xho*I fragment (plasmid pLH1XH3) resulted in the synthesis of a new protein 7 kd smaller (which is equivalent to the coding capacity of the excised 186-bp tract) than the normal VP1 protein. The results of Janik indicated that the entire *Xho*I fragment (nt 2234 to 2419) is translated into VP1, but the only ORF in that region is frame 1. Presumably, VP1 has to switch from reading in frame 1 to frame 2 in a small region (nt 2420 to 2464) to acquire the NPYL conserved sequence and the rest

FIG. 7. Homology of the translated right ORFs among parvoviruses AAV2, BPV, and MVMp. Homologous regions are enclosed by boxes.

of the right ORF. However, there is no splice donor or acceptor in that small region that satisfies the consensus sequences (36). The other possibility is that ribosomal frameshifting (as demonstrated for the expression of the Rous sarcoma virus *pol* gene) (25) occurs in this small region. With either ribosomal frameshifting or a low-frequency RNA splicing event (despite the absence of conventional splice donor and acceptor sites), the right ORF would be in frame 1, and reading of VP1 would be uninterrupted from nt 2191 onward.

**The mid ORF.** No significant homology was detected among the mid ORFs of AAV, BPV, and MVM (Table 2). Even the closely related viruses MVM, CPV, and FPV showed little homology when the regions corresponding to the coding domain for the C terminus of NS-2 in MVM (nt 1990 to 2236) were analyzed. A total of 20 conserved amino acids are scattered in a total of 82 residues. Whatever proteins these ORFs in the middle of the genomes may code for, they have little amino acid sequence homology, although they may have similar functions.

**Homology with RA-1.** Although BPV has little homology with AAV, CPV, FPV, H-1, KRV, LuIII, or MVM, it is highly homologous to a recently described human parvovirus RA-1 (49) isolated from the synovial fluid of a patient suffering from rheumatoid arthritis (D. Van Leeuwen, Abbott Laboratories, North Chicago, Ill., personal communication). The similarity coefficients ($S_{AB}$) between BPV and RA-1 in the left, right, and mid ORFs are considerably higher than the values shown in Table 2 (manuscript in preparation). The relationship between these two parvoviruses deserves further investigation.

**Comparison with B19.** While the present manuscript was in preparation, the nt sequence of the human parvovirus B19 became available (47). B19 is probably an autonomous

TABLE 3. Similarity coefficients between putative parvoviral proteins of B19, AAV, BPV, and MVM

| ORF | Virus | $S_{AB}$ for proteins of: | | | |
|---|---|---|---|---|---|
| | | B19 | AAV | BPV | MVM |
| Left | B19 | 1 | 0.20 | 0.17 | 0.13 |
| | AAV | 0.20 | 1 | 0.15 | 0.15 |
| | BPV | 0.17 | 0.15 | 1 | 0.13 |
| | MVM | 0.13 | 0.15 | 0.13 | 1 |
| Right | B19 | 1 | 0.19 | 0.16 | 0.13 |
| | AAV | 0.19 | 1 | 0.27 | 0.18 |
| | BPV | 0.16 | 0.27 | 1 | 0.20 |
| | MVM | 0.13 | 0.18 | 0.20 | 1 |

parvovirus, since it appears to replicate in erythropoietin-stimulated human bone marrow cultures without the aid of a helper virus. However, it is more similar to the defective parvovirus AAV in having inverted terminal repeats at both ends and in lacking the two conserved regions (GGG and YNN) in its right ORF which are characteristic of other autonomous parvoviruses. It does contain the conserved GKRN region in its left ORF and the conserved PGY, NPYL, TPW, and PIW regions in its right ORF. From the sequence analyses of B19, AAV2, and MVM, Shade et al. (47) concluded that B19 is only distantly related to the other parvoviruses and that B19 is as different from MVM and AAV2 as those two viruses are different from each other.

When we compared the homology of the putative proteins translated from the left and right ORFs of B19 with the corresponding proteins from the other parvoviruses (Table 3), we found that B19 is closer to AAV than it is to BPV or to MVM and that the evolutionary relationship among them is B19↔AAV↔BPV↔MVM. The same relationship was deduced when the algorithm described by Goad and Kanehisa (19) was used to calculate the evolutionary distance between two protein sequences (data not shown).

In conclusion, from the genomic sequences reported so far, there seem to be four classes of parvoviruses: (i) B19; (ii) AAV; (iii) BPV; and (iv) MVM, H-1, FPV, and CPV. The highly conserved amino acid sequences observed among all parvoviruses may prove to be useful in the development of molecular probes for the detection and identification of viruses in the family *Parvoviridae*. For example, a synthetic peptide including a consensus amino acid sequence derived from the GKRN region found in the left ORF of all parvoviruses might be used to prepare antibodies for diagnostic purposes. Antibodies against the conserved GKRN regions of B19 and MVM did give positive reactions in an immunofluorescence assay with BPV-infected cells (M. Lederman and S. F. Cotmore, personal communication). The smaller conserved sequences in the right ORF, if shown to be associated with antigenic domains, could provide the basis for the development of peptide vaccines protective against all parvoviral infections. Synthetic oligonucleotides containing the consensus sequences could be used as probes for detection of parvoviral DNA in clinical specimens. Also, it is now possible to search the BPV sequence for regions identified as functionally important for replication and transcription in other viruses and in eucaryotic cells.

## ACKNOWLEDGMENTS

We are grateful to R. Foote for preparing the synthetic oligonucleotides and to S. Boyle for making the ICR SEQuence program available to us.

This work was supported by American Cancer Society grant MV-220.

## LITERATURE CITED

1. **Armentrout, R., R. Bates, K. Berns, B. Carter, M. Chow, D. Dressler, K. Fife, W. Hauswirth, G. Hayward, G. Lavelle, S. Rhode, S. Straus, P. Tattersall, and D. Ward.** 1978. A standardized nomenclature for restriction endonuclease fragments, p. 523–526. *In* D. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Astell, C. R., M. B. Chow, and D. C. Ward.** 1985. Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J. Virol. **54**:171–177.
3. **Astell, C. R., E. M. Gardiner, and P. Tattersall.** 1986. DNA sequence of the lymphotrophic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain. J. Virol. **57**:656–669.
4. **Astell, C. R., M. Thomson, M. Merchlinsky, and D. C. Ward.** 1983. The complete nucleotide sequence of minute virus of mice, an autonomous parvovirus. Nucleic Acids Res. **11**:999–1018.
5. **Banerjee, P. T., W. H. Olson, D. P. Allison, R. C. Bates, C. E. Snyder, and S. Mitra.** 1983. Electron microscopic comparison of the sequences of single-stranded genomes of mammalian parvoviruses by heteroduplex mapping. J. Mol. Biol. **166**:257–272.
6. **Bates, R. C., C. E. Snyder, P. T. Banerjee, and S. Mitra.** 1984. Autonomous parvovirus LuIII encapsidates equal amounts of plus and minus DNA strands. J. Virol. **49**:319–324.
7. **Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson.** 1985. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc. Natl. Acad. Sci. USA **82**:7919–7923.
8. **Ben-Asher, E., and Y. Aloni.** 1984. Transcription of minute virus of mice, an autonomous parvovirus, may be regulated by attenuation. J. Virol. **52**:266–276.
9. **Bensimhon, M., J. Gabarro-Arpa, R. Ehrlich, and C. Reiss.** 1983. Physical characteristics in eucaryotic promoters. Nucleic Acids Res. **11**:4521–4540.
10. **Berget, S. M.** 1984. Are U4 small nuclear riboproteins involved in polyadenylation? Nature (London) **309**:179–182.
11. **Berns, K. I., and W. W. Hauswirth.** 1984. Adeno-associated virus DNA structure and replication, p. 1–31. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
12. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
13. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site. Cell **41**:349–359.
14. **Burd, P. R., S. Mitra, R. C. Bates, L. D. Thompson, and E. R. Stout.** 1983. Distribution of restriction enzyme sites in the bovine parvovirus genome and comparison to other autonomous parvoviruses. J. Gen. Virol. **64**:2521–2526.
15. **Carlson, J., K. Rushlow, I. Maxwell, F. Maxwell, S. Winston, and W. Hahn.** 1985. Cloning and sequence of DNA encoding structural proteins of the autonomous parvovirus feline panleukopenia virus. J. Virol. **55**:574–582.
16. **Carter, B. J., C. A. Laughlin, and C. J. Marcus-Sekura.** 1984. Parvovirus transcription, p. 153–207. *In* K. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
17. **Cotmore, S. F., L. J. Sturzenbecker, and P. Tattersall.** 1983. The autonomous parvovirus MVM encodes two nonstructural proteins in addition to its capsid polypeptides. Virology **129**:333–343.
18. **Cotmore, S. F., and P. Tattersall.** 1986. Organization of nonstructural genes of the autonomous parvovirus minute virus of mice. J. Virol. **58**:724–732.
19. **Goad, W. B., and M. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.

20. **Hall, M. N., L. Hereford, and I. Herskowitz.** 1984. Targeting of *E. coli* β-galactosidase to the nucleus in yeast. Cell **36**:1057–1065.
21. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
22. **Hauswirth, W. W.** 1984. Autonomous parvovirus DNA structure and replication, p. 129–152. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
23. **Hermonat, P. L., M. A. Labow, R. Wright, K. I. Berns, and N. Muzyczka.** 1984. Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. J. Virol. **51**:329–339.
24. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
25. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230**:1237–1242.
26. **Janik, J. E., M. M. Huston, and J. A. Rose.** 1984. Adeno-associated virus proteins: origin of the capsid components. J. Virol. **52**:591–597.
27. **Johnson, F. B.** 1984. Parvovirus proteins, p. 259–295. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
28. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
29. **Labieniec-Pintel, L., and D. Pintel.** 1986. The minute virus of mice $P_{39}$ transcription unit can encode both capsid proteins. J. Virol. **57**:1163–1167.
30. **Lederman, M., R. C. Bates, and E. R. Stout.** 1983. In vitro and in vivo studies of bovine parvovirus proteins. J. Virol. **48**:10–17.
31. **Lederman, M., K. C. Chen, E. R. Stout, and R. C. Bates.** 1986. Possible sequences for nuclear accumulation of parvoviral proteins. Cell Biol. Int. Rep. **10**:383–387.
32. **Lederman, M., J. T. Patton, E. R. Stout, and R. C. Bates.** 1984. Virally coded noncapsid protein associated with bovine parvovirus infection. J. Virol. **49**:315–318.
33. **Lefebvre, R. B., S. Riva, and K. I. Berns.** 1984. Conformation takes precedence over sequence in adeno-associated virus DNA replication. Mol. Cell. Biol. **4**:1416–1419.
34. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
35. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
36. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
37. **Paradiso, P. R.** 1984. Identification of multiple forms of the noncapsid parvovirus protein NCVP1 in H-1 parvovirus-infected cells. J. Virol. **52**:82–87.
38. **Paradiso, P. R., K. R. Williams, and R. L. Costantino.** 1984. Mapping of the amino terminus of the H-1 parvovirus major capsid protein. J. Virol. **52**:77–81.
39. **Parris, D. S., and R. C. Bates.** 1976. Effect of bovine parvovirus replication on DNA, RNA and protein synthesis in S phase cells. Virology **73**:72–78.
40. **Pintel, D., D. Dadachanji, C. R. Astell, and D. C. Ward.** 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic Acids Res. **11**:1019–1038.
41. **Pustell, J., and F. Kafatos.** 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res. **12**:643–655.
42. **Rhode, S. L., III.** 1985. Nucleotide sequence of the coat protein gene of canine parvovirus. J. Virol. **54**:630–633.
43. **Rhode, S. L., III, and P. K. Paradiso.** 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. J. Virol. **45**:173–184.
44. **Roychoudhury, R., and R. Wu.** 1980. Terminal transferase-catalyzed addition of nucleotides to the 3′ termini of DNA. Methods Enzymol. **65**:43–62.
45. **Sahli, R., G. K. McMaster, and B. Hirt.** 1985. DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus, minute virus of mice. Nucleic Acids Res. **13**:3617–3633.
46. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
47. **Shade, R. O., M. C. Blundell, S. F. Cotmore, P. Tattersall, and C. R. Astell.** 1986. Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis. J. Virol. **58**:921–936.
48. **Siegl, G.** 1984. Biology and pathology of autonomous parvoviruses, p. 296–232. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
49. **Simpson, R. W., L. McGinty, L. Simon, C. A. Smith, C. W. Godzeski, and R. J. Body.** 1984. Association of parvoviruses with rheumatoid arthritis of humans. Science **223**:1425–1428.
50. **Srivastava, A., E. W. Lusby, and K. I. Berns.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. **45**:555–564.
51. **Tattersall, P., and D. C. Ward.** 1978. The parvovirus—an introduction, p. 3–12. *In* D. C. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
52. **Tratschin, J.-D., I. L. Miller, and B. J. Carter.** 1984. Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function. J. Virol. **51**:611–619.
53. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutations prevent messenger RNA 3′ end formation. Science **226**:1045–1051.
54. **Woychik, R. P., R. H. Lyons, L. Post, and F. M. Rottman.** 1984. Requirement for the 3′ flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. USA **81**:3944–3948.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene **33**:103–119.