


Transcription and encapsidation in parvoviruses LuIII
and bovine parvovirus


by

Nanette Diffoot Carlo


Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Biology


APPROVED:


M. Lederman, Co-chairman


R. C. Bates, Co-chairman


J. L. Johnson


T. Sitz


E. R. Stout


S.A. Tolin

April, 1992

Blacksburg, Virginia

**Transcription and encapsidation in parvoviruses LuIII
and bovine parvovirus**

by

Nanette Difffoot Carlo

M. Lederman, Co-chairman

R. C. Bates, Co-chairman

Biology

(ABSTRACT)

The termini of the autonomous parvovirus LuIII, which encapsidates plus and minus DNA strands equally, were cloned and sequenced. The left and right termini of LuIII differ in nucleotide sequence and these termini can assume T- and U-shaped intra-strand base-paired structures, respectively. The LuIII termini are virtually identical in nucleotide sequence and secondary structure to those of the rodent parvoviruses MVM and H-1. The presence of non-identical LuIII termini demonstrated that identical ends are not required for the encapsidation of both DNA strands with equal frequency, as suggested for parvoviruses B19 and AAV. An infectious genomic clone of LuIII was constructed and sequenced. The LuIII genome is 5135 bases and it shares over 80% sequence identity with the sequence of the genomes of MVM and H-1. The genome organization of LuIII is virtually identical to that of the rodent parvoviruses of known sequence. The major ORFs, the left and right ORFs, are restricted to the plus strand. Promoter-like sequences are present at map units 4 and 38. The transactivation responsive element (TAR), characterized in H-1, upstream of P38, is also present in LuIII. Regulatory sequences and splice donor-acceptor consensus sequences, characterized in MVM and H-1, are also present in LuIII. This suggests that both LuIII promoters are functional, and that the transcription map for LuIII could be very similar to that of MVM. The LuIII sequence has only a single copy of a repeat present in tandem at the right end of the MVMP genome. Downstream of this sequence, an A-T rich region of 47 nt is present in LuIII. Since this A-T rich region is absent from the genomes of MVM and H-1, we propose that it represents a putative encapsidation signal responsible for the encapsidation pattern observed for LuIII.

Northern analysis of BPV RNAs suggests that, like the human parvovirus B19, most, if not all, BPV transcripts initiate at promoter sequences localized at map unit 4. Amplification of BPV cDNA ends by the polymerase chain reaction resulted in a number of BPV-specific fragments. Four of these fragments were cloned and sequenced. Sequencing revealed two splices, one of which is very likely a major splice for several BPV transcripts. cDNA fragments were assigned to transcripts possibly coding for three BPV non-structural proteins. Amplification of BPV transcripts with primers specific to the mid-ORF suggests that the amino terminus of the capsid protein VP1 is not coded for by the mid-ORF as suggested by earlier studies, but instead results from one or both of the two small ORFs present upstream of the right ORF, in the same reading frame.

Dedication

I dedicate this work to my most challenging project of all, my daughter, Ariadhne Padilla-Diffot

Acknowledgements

During my years in graduate school I have experienced just about every emotion known to exist. All these moments, both the exciting and disappointing, have been shared with very special people which I would like to acknowledge, and say thank you to.

Special thanks must go to my advisors, Dr. Muriel Lederman, and Dr. Robert C. Bates for their advice, help, time, support, and most importantly, for their respect. Thank you for making the working environment an enjoyable one.

My committee members, Dr. John Johnson, Dr. Thomas Sitz, Dr. Ernest Stout, and Dr. Sue Tolin, who were always there to help and make invaluable suggestions. I thank you for your help, support and your committment in seeing this project through.

To my dear friends, Dr. Katherine Chen, Dr. Brock Metcalf, and Dr. Bruce Shull, for caring when something went wrong, for your good suggestions, and even for those suggestions one would never think of trying, for your invaluable friendship, that I will cherish forever.

Finally to my family, especially my husband, Edwin H. Padilla, my mom Nayda Carlo Garcia, my sister Aileen Rivera Graniela, and my brother John Marcus Difffoot Carlo, for their sacrifices, encouragement, patience and support during these years.

To all those people that go without mentioning, because the list would be extensively long, and to all those mentioned above, I have just two words left to say, thank you!

Table of Contents

Literature review	1
Characteristics and taxonomy of parvoviruses	1
Genome organization	3
Replication of parvoviruses	5
a) Hairpin transfer model	5
b) Modified rolling hairpin transfer model	6
c) Kinetic hairpin transfer model	8
Transcription	10
Encapsidation	17
LITERATURE CITED	18
OBJECTIVES	21
IDENTICAL ENDS ARE NOT REQUIRED FOR THE EQUAL ENCAPSIDATION OF PLUS AND MINUS STRAND PARVOVIRUS LU111 DNA	24
INTRODUCTION	24
MATERIALS AND METHODS	25

RESULTS	28
DISCUSSION	34
LITERATURE CITED	36

**THE SEQUENCE OF PARVOVIRUS LUIII AND LOCALIZATION OF A PUTATIVE
SIGNAL RESPONSIBLE FOR ITS ENCAPSIDATION PATTERN** 39

INTRODUCTION	39
MATERIALS AND METHODS	41
Materials	41
Plasmid propagation	41
Analyses of infectivity of genomic clones	42
Sequencing of LuIII DNA and computer analysis	42
Nucleotide sequence accession number	43
RESULTS	45
LuIII genomic clone construction	45
Infectivity of LuIII genomic clones	45
LuIII sequence determination	46
Genomic organization	57
Assignment of coding domains and transcription map	59
Major sequence differences between LuIII and the rodent parvoviruses MVMp and H-1 ..	60
DISCUSSION	63
LITERATURE CITED	66

TRANSCRIPTION MAP OF BPV AS GENERATED BY AMPLIFICATION OF cDNA

ENDS	69
INTRODUCTION	69
MATERIALS AND METHODS	72
Virus propagation and cell culture	72

RNA Isolation	72
Total RNA isolation using guanidium isothiocyanate	72
Isolation of poly-A tailed mRNA from BPV-infected cells	73
Northern Analysis of BPV RNAs	74
Reverse Transcription of BPV RNA	74
Amplification of BPV transcripts by the Polymerase Chain Reaction	75
Cloning and sequencing of PCR-generated cDNA fragments.	76
RESULTS	81
Nature of BPV transcripts	81
All BPV transcripts originate from P4	81
Analysis of cDNA fragments	86
DISCUSSION	93
BPV RNA species	94
All BPV transcripts initiate from P4	95
BPV infection results in minimum amounts of BPV transcripts	95
Analysis of PCR generated cDNA fragments	96
cDNA fragments of mRNAs for the BPV non-structural protein, NS-1	96
cDNA fragments of mRNAs for the BPV non-structural protein, NP-1	97
cDNA fragments of the mRNAs for the BPV structural proteins	99
LITERATURE CITED	102
SUMMARY	104

List of Illustrations

Figure 1. Genomic organization of parvoviruses MVM, B19, BPV and AAV	4
Figure 2. Hairpin transfer model	7
Figure 3. Modified rolling hairpin transfer model	9
Figure 4. Transcription map for MVM	11
Figure 5. Transcription map for AAV	14
Figure 6. Transcription map for B19	15
Figure 7. Preliminary transcription map for BPV	16
Figure 8. Strategy for cloning LuIII genome fragments into pUC18 and pUC19	27
Figure 9. DNA sequence at the left palindrome of the minus strand of LuIII and comparison to that of MVMp and H-1	29
Figure 10. DNA sequence at the right palindrome of the minus strand of LuIII and comparison to that of MVMp and H-1	30
Figure 11. End-label analysis of double-stranded LuIII virion DNA	33
Figure 12. Strategy for cloning the LuIII genome	44
Figure 13. Sequence comparison of LuIII (LU3), MVMp (4) and H-1 (32)	47
Figure 14. Genomic organization of LuIII	58
Figure 15. Comparison of the translated left ORF coding for NS1 (A) and NS2 (B) among LuIII (LU3), MVMp and H-1	61
Figure 16. Comparison of the translated right ORF coding for VP1 and VP2 among LuIII (LU3), MVMp and H-1	62
Figure 17. Schematic representation of the RACE protocol	77
Figure 18. Sequence of primers used for the amplification of BPV cDNAs ends by the RACE protocol	78

Figure 19. Location in the BPV genome of specific primers used in the RACE protocol . . .	79
Figure 20. Strategy used for the amplification of BPV cDNAs	80
Figure 21. Northern blot analysis of BPV RNA	82
Figure 22. Temporal appearance of BPV RNA by Northern blot analysis of total RNA isolated from BPV-infected BFL cells	83
Figure 23. Alkaline gel analysis of poly-A RNA isolated from BPV-infected BFL cells	84
Figure 24. Northern blot analysis of BPV RNA using BPV-specific probes	85
Figure 25. DNA fragments generated by PCR from amplifications of BPV cDNA ends	88
Figure 26. Southern blot analysis of BPV cDNA fragments generated by PCR	89
Figure 27. Southern blot analysis of BPV cDNA fragments generated with MID-II primers by PCR	90
Figure 28. cDNA recombinant clones generated by the RACE protocol	91
Figure 29. Sequence of splice junctions illustrated in Fig. 28	92
Figure 30. Schematic diagram of potential transcripts for BPV non-structural proteins, NS1 and NP-1 (NS2)	101

Chapter I

Literature review

Characteristics and taxonomy of parvoviruses

The family *Parvoviridae* is divided into three genera and assignment into a genus is based on replication requirements, encapsidation pattern and host of the virus. Members of the genus *Parvovirus* replicate autonomously in vertebrates and encapsidate primarily the minus strand. The *Dependoviruses* replicate in vertebrates but only in the presence of a helper virus and they encapsidate both strands with equal frequency. The third genus, the *Densoviruses*, replicate autonomously in insects and encapsidate both DNA strands with equal frequency. Classification of the mammalian parvoviruses requires reconsideration because there are exceptions for each of the genera. The autonomously replicating parvoviruses B19, the cause of erythema infectiosum and of a form of aplastic crisis in various hemolytic anemias (28), and LuIII, a parvovirus isolated from a human cell line (38), replicate autonomously in vertebrates, like the genus *Parvovirus*, but like the *Dependoviruses*, both encapsidate equal amounts of plus and minus DNA strands. For some time it was assumed that adeno-associated virus (AAV) could only replicate in the presence of adeno-

or herpesvirus, yet recent studies showed that AAV can replicate autonomously in UV-irradiated or hydroxyurea-synchronized HeLa cells (44).

The family *Parvoviridae* is the only family of DNA viruses known to replicate in the nuclei of both vertebrate and invertebrate cells. They have an absolute requirement for S-phase cells. The viral single-stranded DNA genome of approximately 5 kb has palindromic sequences at the termini which can exist in the form of stable hairpin duplexes. These hairpin duplexes can form cruciform or panhandle structures with alternative sequence orientations designated flip and flop. The viral DNA is contained in a naked icosahedral capsid, 15 to 28 nm (20) in size. This virion is composed of 60 subunits. Bovine parvovirus (BPV) codes for three capsid proteins. These are VP1 (82 kDa), VP2 (72 kDa), and VP3 (62 kDa). The fourth capsid protein made by BPV, VP4, results from a proteolytic cleavage of VP3. The rodent parvoviruses code for two capsid proteins, VP1 (83 kDa) and VP2 (64 kDa), and VP3 results from a proteolytic cleavage of VP2. The *Dependoviruses* code for three capsid proteins, VP1 (87 kDa), VP2 (73 kDa), and VP3 (63 kDa). RNA studies suggest that VP3 results from an independent transcript and not from a proteolytic cleavage of VP2. All the structural proteins encoded by the parvoviral genome contribute to the capsomer arrangement. This implies that structurally complementary regions of VP-1, VP-2, and VP-3 or VP4 would interact in the formation of the capsid.

The mature virus particles have a density of 1.41 g/ml as measured by CsCl centrifugation and sediments at 110S (28). They have a relatively high buoyant density in aqueous CsCl because of their high DNA to protein ratio (20). They are extremely stable at 56 °C for 60 min (39).

The host range of most parvoviruses is rather limited. If not restricted to one species, it can only be extended experimentally to animals closely related to its natural host. All parvoviruses appear to have a worldwide distribution. (37). Evidence suggests that every susceptible organism infected with a parvovirus will develop into a virus carrier. Under epidemic conditions, the acutely ill organism is the source of the infecting parvovirus. It is not clear whether persistently infected animals shed parvoviruses continuously or whether latent infections become repeatedly reactivated for limited periods of time (37). Animals were found to shed parvoviruses in their feces, urine, saliva, and nasal secretions. Widespread infections have been noted for BPV, canine parvovirus (CPV),

lapine arbovirus (LPV), minute virus of mice (MVM), and rat virus (RV). The diseases they cause are most severe in the young and newborn of the species they infect. The extraordinary stability of parvoviruses favors the prolonged persistence of these viruses in a contaminated environment (15, 37). Infection occurs via the fecal-oral route and vertical transmission occurs through the placenta (15, 37), while porcine parvovirus (PPV) can be sexually transmitted (37).

Genome organization

The mammalian parvoviruses of known nucleotide sequence share a similar genomic organization (Fig.1). The two major ORFs, the left and right ORFs, are restricted to the plus strand. The left ORF codes for the non-structural proteins and the right ORF for the structural proteins (29). Unlike those of other mammalian parvoviruses, the right half of the B19 genome has a large continuous ORF which overlaps the left ORF by several codons. The genomic organization of BPV and the human parvovirus, RA-1, is unique in that a third ORF (mid ORF) exists. In BPV, this ORF is suggested to code for the small non-structural protein NP-1 (Fig.1). In contrast to the genomic organization of mammalian parvoviruses is that of the densovirus *Junonia coenia* (JcDNV) and *Galleria mellonella densovirus* (GmDNV) which contain ORFs in the plus and minus DNA strands (5, 42) Functional promoters in the rodent parvoviruses, MVMp (genome size of 5149 nt) and H-1 (genome size of 5177 nt), have been located to map units 4 and 38. The AAV genome, 4674 nt in length, has functional promoters at map units 4, 19, and 40. The nucleotide sequences of BPV and B19 revealed potential promoter sequences at map units 4, 12, 38 in a genome of 5517 nt, and at 6, 23, 42, 43, 55 in a genome of 5596 nt, respectively (29).

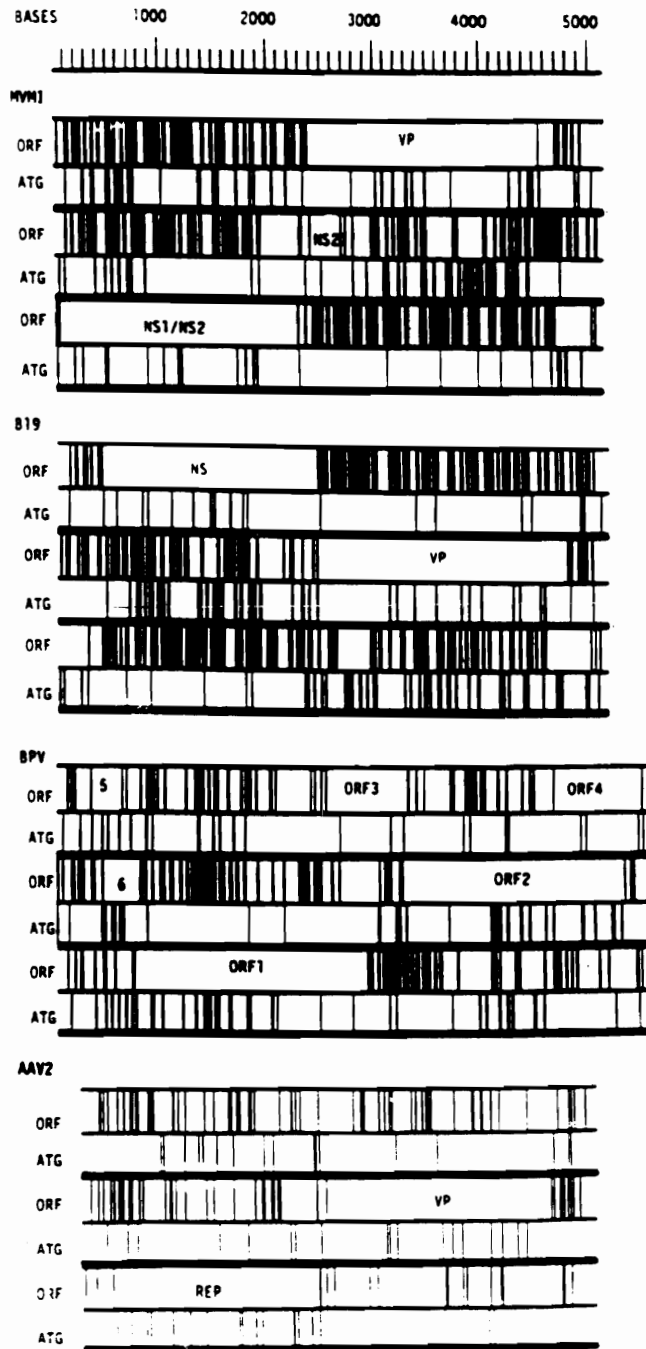


Figure 1. Genomic organization of parvoviruses MVM, B19, BPV and AAV: The diagram illustrates the three reading frames on the plus strand for each virus. The positions of the ATG initiation codons and the termination codons in each frame are indicated by vertical lines. The open reading frames (ORF) are indicated and where possible, known proteins are assigned to their respective ORFs. Abbreviations: NS, REP, non-structural proteins; VP, capsid proteins. (Reprinted from Rhode and Iversen, 1990).

Replication of parvoviruses

Parvoviruses have an absolute requirement for actively dividing cells. The virus may enter the cell at any time but virus replication cannot occur until the cell reaches the S-phase of the cell cycle. When the replication of H-1 was examined by pulse chase experiments (43), no Okazaki fragments were detected. This suggests that no lagging strand DNA synthesis is occurring and that the hairpin terminus is likely the primer for DNA synthesis. Replication of the viral DNA is thought to occur by the mechanism proposed by Cavalier-Smith (12), in which the resolution of the terminal hairpin occurs by a single-stranded nick followed by self-priming DNA synthesis. All models proposed for parvovirus replication agree on and incorporate three basic steps. These are: 1) conversion of single-stranded DNA to double stranded DNA; 2) amplification of replicative form (RF) DNA; 3) and encapsidation of the progeny DNA. Three models have been proposed for parvoviral DNA replication. The hairpin transfer model (7, 19), accounting for replication of AAV, and the modified rolling hairpin model (1) for MVM replication, cannot account for the terminal orientations and polarity of BPV and LuIII viral DNA. The third model, the kinetic hairpin transfer model (14), can account for both the polarity of the viral DNA and the ratios of the sequence orientations at the left and right termini specific for each parvovirus by a single mechanism.

a) Hairpin transfer model

The hairpin transfer model for AAV replication (Fig. 2) (19), proposes that the hairpin termini serve as a primer for DNA synthesis by host cell polymerases, converting the single-stranded viral genome to a double-stranded form (steps 1 and 2). A site-specific endonuclease then introduces a single-stranded nick internal to the terminal palindrome of the closed end molecules (step 2), generating a free 3'OH on the parental viral strand. This 3'OH then serves as a primer for elongation of the viral strand by synthesis of the reverse complement of the original palindrome (step 3). Once resolution of the hairpin is complete, the open-ended RF molecule can be amplified

by continued hairpin formation, DNA synthesis and terminal resolution (steps 3' to 6). This replication scheme can account for the generation of the equal flip and flop conformations of the AAV terminal palindromes (24). Im et al. found that the AAV non-structural protein *rep 68* binds and nicks the AAV terminal hairpin in the presence of ATP, enabling synthesis of the complementary strand DNA (21, 40).

b) Modified rolling hairpin transfer model

The sequence at the left end of the viral genome of the rodent parvoviruses exists solely in the flip conformation and unlike AAV, these viruses encapsidate the minus strand almost exclusively. The hairpin transfer model cannot account for these phenomena. To explain replication of the rodent parvoviruses, the rolling hairpin model was proposed (1) (Fig. 3). According to this model the host polymerases will prime synthesis of a complementary strand from the 3' terminal palindrome of the infecting DNA. This results in a monomer duplex RF covalently closed at the left end (step 1). Either the 5' and 3' ends of the molecule are ligated to give rise to a transient endless duplex intermediate (step 4), or the extended right end assumes a U-shaped hairpin structure and synthesis of the complementary strand occurs to produce a dimer RF molecule with an extended right end and a covalently closed turn-around terminus at the left end (step 2). A nick, 18 nts inboard of the 5' end of the parental strand, opposite the 3' end of the newly synthesized viral strand, is introduced at the left end of the parental minus strand (steps 2 and 3). A second nick is introduced on the opposite strand and the 3' terminus of the progeny minus strand is ligated to the 5' terminus of the plus strand (step 4). The remaining 3' OH at this nicking site is used as a primer for strand extension of the parental viral strand conserving the sequence orientation at the 3' terminus (step 3). These forms are monomer double-stranded RFs (step 5), all with the flip conformation at the left end and a 1/1 ratio of flip to flop at the right end, that can enter the pathway again at step 1.

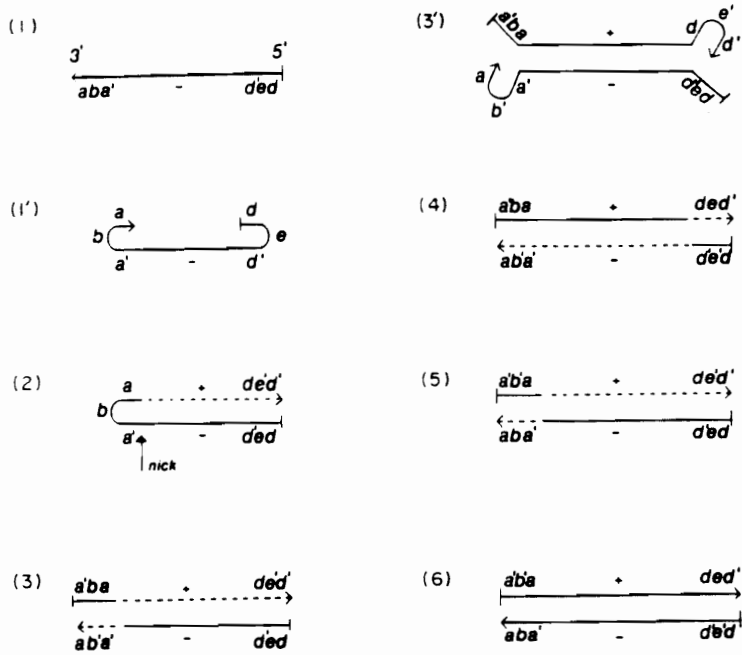


Figure 2. Hairpin transfer model: General scheme for the replication of AAV. (–) and (+) represent viral DNA of minus and plus polarity. Lower case letters a, b, c, d, e, and a', b', c', d', e' represent reverse complementary sequences, respectively. The site of nicking during hairpin transfer is shown by the arrow. (Reprinted from Chen et al., 1989).

Synthesis of a dimer molecule is a critical step in this replication model since it is required for the regeneration of viral strands with a single sequence orientation at the 3' terminus. A covalently closed circular RF molecule likely generated in step 4 or by a ligation reaction in step 1 has been identified (17) The protein responsible for producing these nicks is probably NS1 (18).

c) Kinetic hairpin transfer model

The sequences of the terminal palindromes of parvovirus LuIII are not known, but evidence suggests that they are similar to those of the rodent parvoviruses MVM and H-1. Unlike the rodent parvoviruses, LuIII encapsidates both strands with equal frequency. BPV also encapsidates plus strand (10%). Due to the encapsidation pattern observed for these viruses, the modified rolling hairpin model cannot explain the replication of parvoviruses BPV and LuIII. A third model, the unified kinetic hairpin model (14), suggests that replication of all parvoviruses proceeds predominantly through four monomer ds RFs. This model takes into account the polarity of the virion DNA and the sequence conformation of the terminal palindromes. The conversion of the ss virion DNA to a ds RF molecule would occur through a self-priming event that utilizes the host cell replication proteins. The closed end replicative intermediate generates all four possible monomer RFs by successive cycles of terminal resolution, hairpin formation and DNA synthesis. The amplification of a particular RF is based on the differential kinetics of hairpin formation and complementary strand synthesis. These rates are defined by constants associated with the location of the terminus (left versus right) and the sequence orientation (flip versus flop) at that terminus.

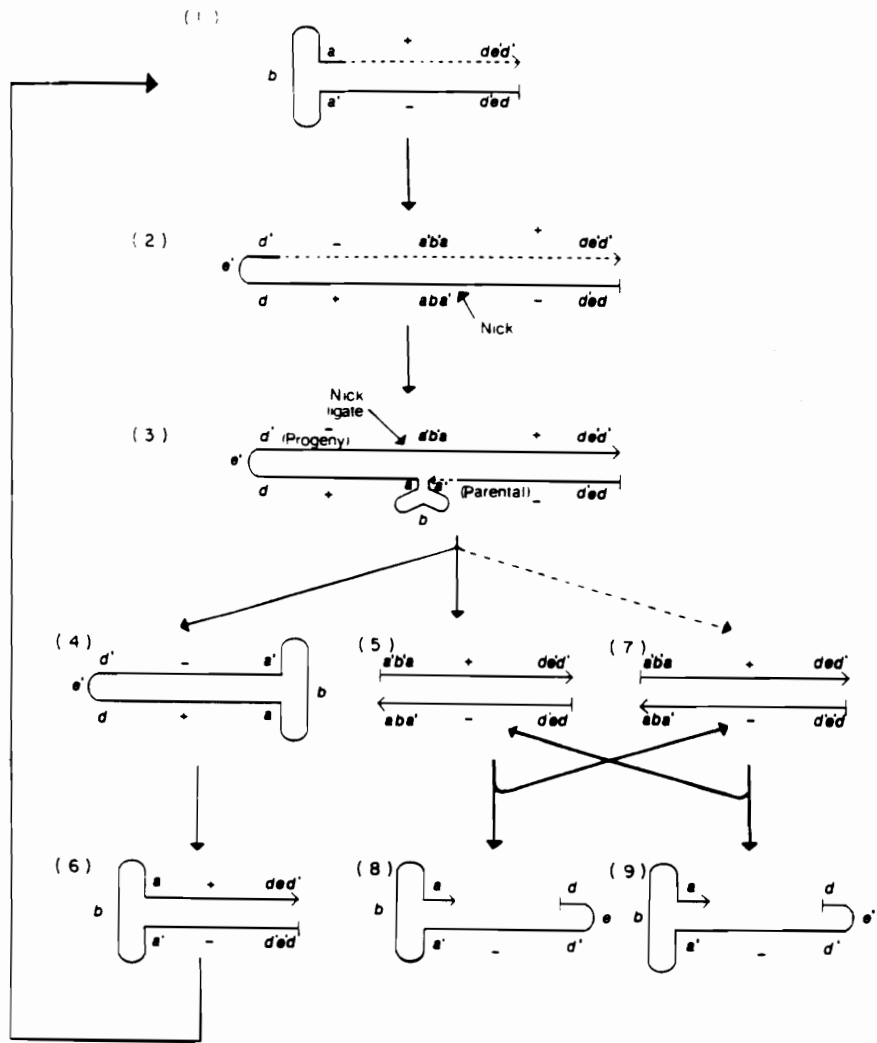


Figure 3. Modified rolling hairpin transfer model: Model proposed for the replication of rodent parvoviruses such as MVM. (Reprinted from Chen et al., 1989).

Transcription

RNAs coding for the non-structural and structural proteins of parvoviruses are each transcribed as overlapping transcription units initiating at a single promoter (B19) or multiple promoters (MVM, H-1, AAV) in the viral genome. The MVM genome codes for three major RNA species, R1, R2 and R3 (16) (Fig. 4). The P4 promoter has a TATA box and a near-consensus SP1 binding site 12 nucleotides upstream. R1 and R2 initiate transcription at P4 and code for the nonstructural proteins NS1 (84 kDa) and NS2 (25 kDa) respectively. The P38 promoter has a classical TATA-box sequence, and 22 nts upstream of P38 a consensus SP1 binding site, highly conserved in this group of viruses, is present. In H-1, a transactivation responsive element (TAR) is required in *cis* for transactivation of P38 (32). This sequence has been mapped to positions -146 to -116 for the H-1 promoter. R3 initiates transcription at P38 and codes for the structural or capsid proteins VP1 (83 kDa) and VP2 (64 kDa). Three forms of a small splice exist in each of the three classes of viral mRNA, resulting from the use of two donor (donor a at nt 2280 and donor b at nt 2398) and two acceptor (acceptor a at nt 2376 and acceptor b at nt 2398) splice sites. Since NS1 is encoded from a continuous sequence in frame three, terminating upstream of the small splice, all three spliced forms of the R1 transcript are expected to generate identical proteins. The R2 transcripts coding for NS2 would contain an additional splice, known as the large splice in MVMp, from nt 520 to nt 1996. These R2 transcripts splice a second time, the majority from donor a to acceptor a and a small number from donor b to either acceptor a or acceptor b. VP1 and VP2 are translated from R3 transcripts generated from P38. Transcripts for VP1 splice from donor b to acceptor b and initiate translation at the first ATG in the right ORF, at m.u. 44. The majority of transcripts directing the synthesis of VP2 contain a splice from donor a to acceptor a, while a small number contain a splice from donor a to acceptor b. Both VP2 transcripts would initiate translation at m.u. 54 (nt 2792).

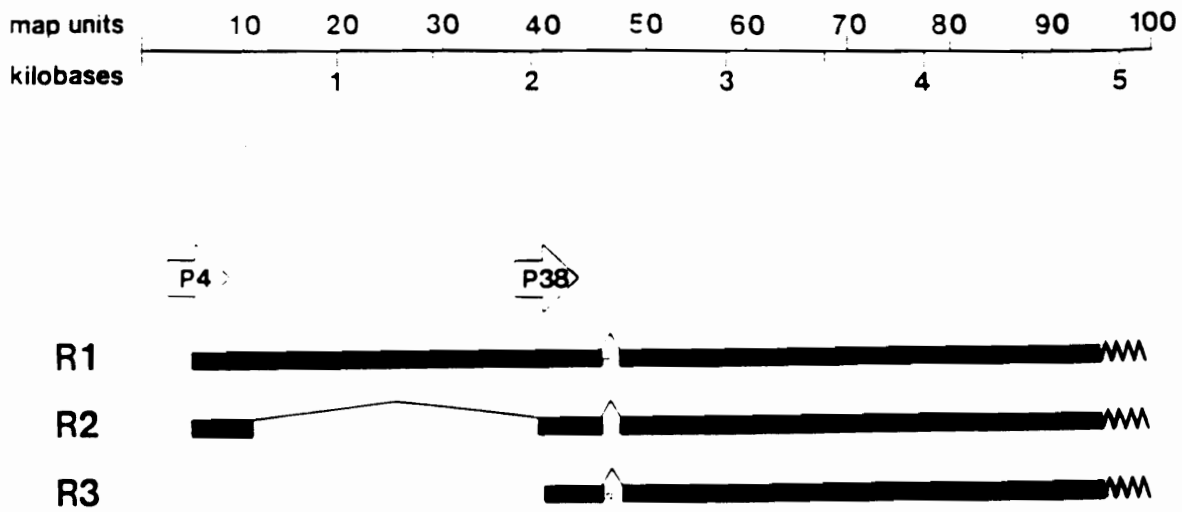


Figure 4. Transcription map for MVM: MVM transcripts are aligned below a line diagram of the plus strand DNA. The solid boxes indicate genomic sequences present in each transcript. The lines represent sequences removed during processing. (Reprinted from Cotmore, 1990).

The dependent parvovirus, AAV, has three promoters (Fig. 5). There are three overlapping RNA families transcribed from promoter sequences at map positions 5, 19 and 40. The 4.2 kb and 3.6 kb transcripts, for the nonstructural proteins rep78 and rep52, respectively, initiate at P5 and terminate at the UAA codon at nt 2184 or 2193. The 3.9 kb and the 3.3 kb transcripts for the nonstructural proteins rep68 and rep40, respectively, initiate at at P19 and terminate translation at the UGA codon at nt 2250. The transcripts generated from P19 represent spliced versions of the P5 generated transcripts. This splice spans nts 1908 to nt 2227. The AAV capsid proteins VP1 (87kDa), VP2 (73 kDa) and VP3 (63 kDa), are translated from a 2.3 kb mRNA generated from P40. VP2 begins translation at the AUG at nt 2615 and VP3 begins at an ACG at nt 2810. The start of the VP1 transcript is still unknown.

The B19 genome codes for at least nine overlapping transcripts all initiating at a strong promoter at map unit 6 (Fig. 6). All these transcripts contain a short 5' leader sequence of approximately 60 nt. The transcript coding for the non-structural protein NS1 (77 kDa), uses an internal polyadenylation signal. The RNAs coding for the capsid proteins VP1 (84 kDa) and VP2 (58 kDa) contain a large splice. Two transcripts containing sequences from the middle of the genome are known not to code for capsid proteins, but the translation products have not been identified.

Mapping of BPV transcripts suggests that, like B19, all transcripts may be initiated at a single promoter located at map unit 14 (9), although the nucleotide sequence revealed potential promoter sequences at map units 4.5, 12.8 and 38.7 of the viral genome (13).

BPV transcription *in vivo* produced four size classes of polyadenylated RNA, the 5.25 kb, 3.6 - 3.1 kb, 2.6 - 2.25 kb, and 1.45 - 0.8 kb species (Fig. 7). Mapping by S1 nuclease digestion and two dimensional neutral and alkaline agarose gel electrophoresis showed that the members of each size class contain a common main body to which smaller segments are spliced. By *in vitro* translation of size-fractionated cytoplasmic RNA from BPV-infected BFL cells, the most abundant BPV RNA, the 2.6 kb family, was mapped to the right end of BPV genome and was shown to encode all three BPV capsid proteins (9). Transcription for this RNA family starts at map unit 14 and continues to map unit 94. Leader RNA segments, 0.35 kb or smaller, from map positions 14

through 20, are joined to a 2.25 kb main body from map positions 53.5 to 94. Three 5' donor sites (map positions 18, 19 and 20) contained in the leader segments, would join a single acceptor site at map position 53.5 contained in the main body of the transcript. The transcripts for the 2.6 kb BPV RNA family apparently also produces the 1.45 - 0.8 kb RNAs. These would be contained within the large intron of the 2.6 kb RNAs, initiating transcription at map position 13.5. This suggests that these RNAs arise from multiple splicing patterns of a single primary transcript. The 5.25 kb and 3.6 - 3.1 kb RNAs are less abundant. The 5.25 kb RNA species consists of a 3.1 kb and 2.25 kb RNA segments while 3.6 - 3.1 kb family contains a 3.1 kb main body. A small amount of unspliced 3.6 kb RNA was observed. Those transcripts coding for the non-structural proteins NS1 (72 and 83 kDa) and NS2 (28 kDa) were not identified.

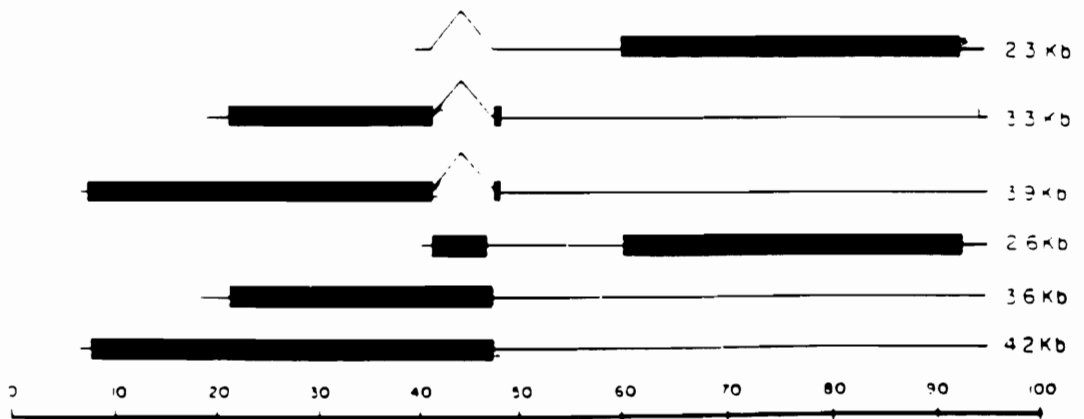


Figure 5. Transcription map for AAV: AAV transcripts are aligned above a line diagram of the plus strand DNA. The solid boxes indicate genomic sequences present in each transcript. The lines represent sequences removed during processing. (Reprinted from Carter et al., 1990).

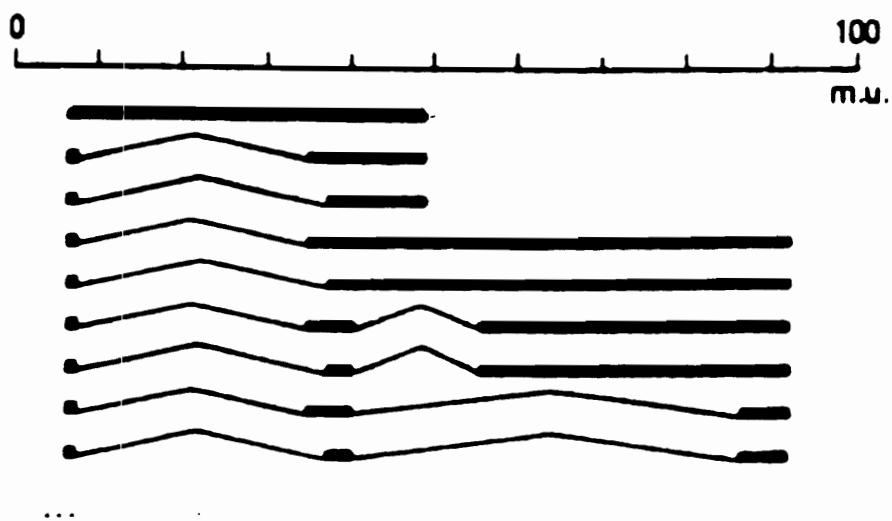


Figure 6. Transcription map for B19: B19 transcripts are aligned below a line diagram of the plus strand DNA. The solid boxes indicate genomic sequences present in each transcript. The lines represent sequences removed during processing. (Reprinted from Ozawa et al., 1987).

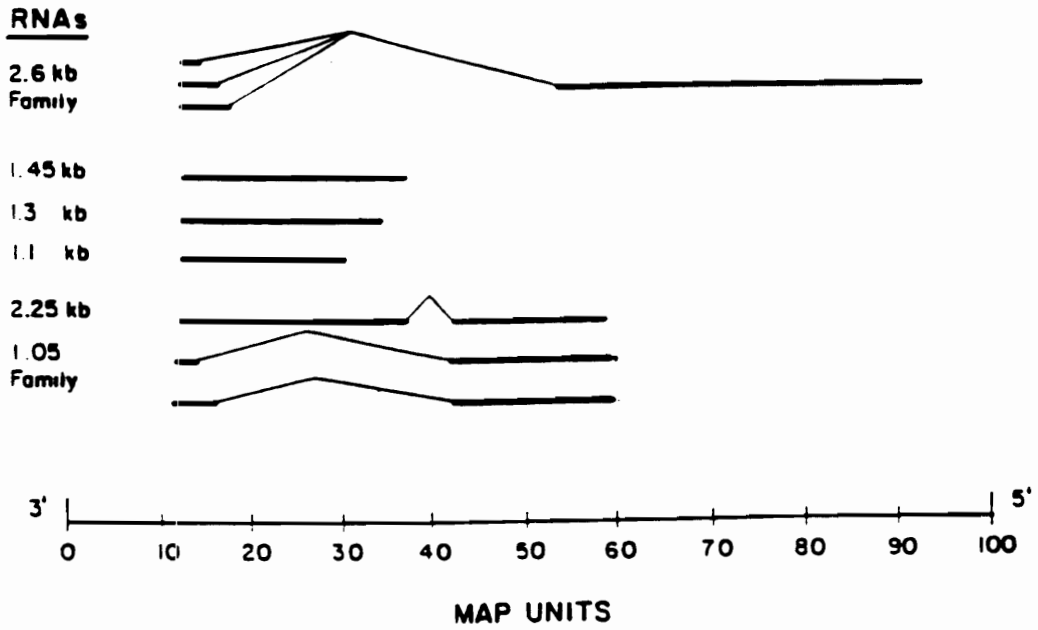


Figure 7. Preliminary transcription map for BPV: BPV transcripts are aligned above a line diagram of the plus strand DNA. The solid boxes indicate genomic sequences present in each transcript. The lines represent sequences removed during processing. (Reprinted from Burd, 1982).

Encapsidation

Studies with genomes containing terminal deletions (22, 25, 34, 36) or site-specific mutations (23) suggest that the *cis* signals necessary for DNA replication and encapsidation reside in the genomic termini, possibly in the 200 nucleotides at each end. AAV has identical terminal repeats and separately encapsidates equal amounts of plus and minus strands (6, 8). All autonomous parvoviruses of known sequence, with the exception of B19, have non-identical ends (35). Their encapsidation patterns allow separation into three distinct subgroups (for a review see 39). The rodent parvoviruses MVM (3), RV, (2, 33) and H-1 (30, 31) have closely related terminal sequences and encapsidate 99% minus strand. BPV (14) and LPV both encapsidate about 90% minus strand. Parvoviruses B19 and LuIII have encapsidation patterns similar to that of the dependent parvovirus AAV. Like AAV, both viruses encapsidate plus and minus strands with equal frequency (4, 39). The equal production of both strands of AAV is thought to result from the symmetry of the terminal palindromes (7). This may be the case for B19 (35) but not for LuIII, since it did not form panhandle structures during annealing reactions suggesting that identical ends are not required for the equal encapsidation of plus and minus DNA strands. The signals required for encapsidation of single-stranded virion DNA are unknown. Evidence suggests that late in infection complete capsids interact with the viral DNA. Experiments inhibiting AAV particle formation showed that the AAV capsid protein, VP3 or an assembled AAV capsid is required for accumulation of AAV single-stranded DNA. It is not clear whether the capsid proteins or assembled capsids were required for the actual strand displacement or merely to sequester single strand DNA into particles. Since both defective interfering particles and empty AAV particles were as stable as mature AAV particles, a full length DNA molecule is not required for particle stability. DNA appears to be packaged into preformed particles. The thought is that the newly synthesized AAV protein is assembled into empty capsids which then associate with progeny DNA strand concomitant with, or soon after, displacement of the single strand DNA, and this displacement synthesis of the parvoviral progeny strand may be driven by concomitant encapsidation (26).

LITERATURE CITED

1. 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.
2. Astell, C. R., M. Smith, M. B. Chow, and D.C. Ward. 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at the origins of DNA replication. *Cell* 17:691-703.
3. Astell, C. R., M. Thompson, M. Merchlinsky, and D. C. Ward. 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. *Nuc. Acids Res.* 11:999-1018.
4. Bates, R. C., C. E. Snyder, P. T. Bannerjee, and S. Mitra. 1984. Autonomous parvovirus LuIII encapsidates equal amounts of plus and minus DNA strands. *J. Virol.* 49:319-324.
5. Bergoin, M., M. Jourdan, M. Gervais, F. X. Jousset, S. Skory and B. Dumas. 1989. Molecular cloning, nucleotide sequence and organization of an infectious genome of the *Junonia coenia* Densovirus (JcDNV). Abstract. EMBO Workshop: Molecular biology of parvoviruses. Kibbutz Ma'ale Hachamisha, Israel.
6. Berns, K. I. and R. A. Bohensky. 1987. Adeno-associated viruses: an update. *Adv. Vir. Res.* 32:243-306.
7. 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.
8. Berns, K. I. and M. Labow. 1987. Parvovirus gene regulation. *J. Gen. Virol.* 68:601-614.
9. Burd, P. 1982. Characterization and localization of in vivo bovine parvovirus transcription products. Dissertation, Virginia Polytechnic Institute and State University.
10. Carter, B. J., C. A. Laughlin, and C. J. Marcus-Sekura. 1984. Parvovirus transcription. in (K. I. Berns, Ed.), *The Parvoviruses*. Plenum Press, New York, NY.
11. Carter, B. J., E. Mendelson, and J. P. Trempe. 1990. AAV replication, integration, and genetics, in (Tijssen, P., Ed.), *Handbook of parvoviruses* Vol. I. CRC Press, Boca Raton, FL.
12. Cavalier-Smith, T. 1974. Palindromic base sequences and replication of eukaryotic chromosome ends. *Nature* 250:467-470.
13. Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates. 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. *J. Virol.* 60:1085-1097.

14. Chen, K. C., J. J. Tyson, M. Lederman, E. R. Stout, and R. C. Bates. 1989. A kinetic hairpin transfer model for parvoviral DNA replication. *J. Mol. Biol.* 208:283-296.
15. Churn, C. C., Bates, R. C., and Broadman, G. G. 1983. Mechanism of chlorine inactivation of DNA-containing parvovirus H-1. *Appl. Environ. Microbiol.* 46, 1394.
16. Cotmore, S. F.. 1990. Gene expression in the autonomous parvoviruses, in (Tijssen, P., Ed.), *Handbook of parvoviruses*. Vol. I. CRC Press, Boca Raton, Fl.
17. Cotmore, S. F., M. Gunther, and P. Tattersall. 1989. Evidence for a ligation step in the replication of the autonomous parvovirus minute virus of mice. *J. Virol.* 63:1002-1006.
18. Cotmore, S. F. and P. Tattersall. 1988. The NS-1 polypeptide of minute virus of mice is covalently attached to the 5' termini of duplex replicative-form DNA and progeny single strands. *J. Virol.* 62:851-860.
19. Hauswirth, W. W. and Berns, K. I., 1977. Origin and termination of adeno-associated virus DNA replication, *Virology*, 78, 488.
20. Hoggan, M. D. 1971. Small DNA viruses, in (Maramorosh, K. and E. Kurstak, Eds.), *Comparative Virology* Academic Press, New York.
21. Im, D. S. and Muzyczka, N. 1989. Factors that bind to adeno-associated virus terminal repeats. *J. Virol.* 63:3095-3104.
22. Labow, M. A., L. H. Graf, Jr., and K. I. Berns. 1987. Adeno-associated virus gene expression inhibits cellular transformation by heterologous genes. *Mol. Cell. Biol.* 7:1320-1325.
23. 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.
24. Lusby, E., K. H. Fife, and K. I. Berns. 1980. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* 34:402-409.
25. Merchlinsky, M. J., P. J. Tattersall, J. J. Leary, S. F. Cotmore, E. M. Gardiner, and D. C. Ward. 1983. Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J. Virol.* 47:227-232.
26. Myers, M. W. and B. J. Carter. 1980. Assembly of adeno-associated virus, *Virology* 102:71.
27. Ozawa, K., J. Ayub, H. Yu-Shu, G. Kurtzman, T. Shinada, and N. Young. 1987. Novel transcription map for the B19 (human) pathogenic parvovirus. *J. Virol.* 61:2395-2406.
28. Pattison, J. R., Ed. 1988. *Parvoviruses and Human Disease*, CRC Press, Boca Raton, Fl.
29. Rhode, S. L., and Iversen, P., 1990. Parvovirus genomes: DNA sequences, in (Tijssen, P., Ed.), *Handbook of parvoviruses*, Vol. I. CRC Press, Boca Raton, Fl.
30. Rhode, S. L., and B. Klaassen. 1982. DNA sequence of the 5' terminus containing the replication origin of parvovirus replicative form DNA. *J. Virol.* 41:990-999.
31. Rhode, S. L., and P. R. Paradiso. 1983. Nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. *J. Virol.* 45:173-184.

32. Rhode, S. L., and S. M. Richard, 1987, Characterization of the *trans*-activation-responsive element of the parvovirus H-1 P38 promoter. *J. Virol.* 61:2807-2815.
33. Salzman, L. A., and P. Fabisch. 1979. Nucleotide sequence of the self-priming 3' terminus of the single-stranded DNA extracted from the parvovirus kilham rat virus. *J. Virol.* 30:946-950.
34. Samulski, R. J., A. Srivastava, K.I. Berns, and N. Muzyczka. 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33:135-143.
35. Shade, R. O., B. 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.
36. Shull, B. C., K. C. Chen, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Genomic clones of bovine parvovirus: construction and effect of deletions and terminal sequence inversions on infectivity. *J. Virol.* 62:417-426.
37. Siegl, G. 1990. Variability, adaptability, and epidemiology of autonomous parvoviruses, in *Handbook of parvoviruses*, Vol. II, Tijssen, P., Ed. CRC Press, Boca Raton, Fl.
38. Siegl, G. 1976. Parvoviruses isolated from human cell lines, p. 72-85. In *The Parvoviruses* Springer-Verlag, New York.
39. Siegl, G., R. C. Bates, K. I. Berns, B. J. Carter, D. C. Kelley, E. Kurstak, and P. Tattersall. 1985. Characteristics and taxonomy of parvoviridae. *Intervirology* 23:61-73.
40. Snyder, R. O., R. J. Samulski, and N. Muzyczka. 1990. *In vitro* resolution of covalently joined AAV chromosome ends. *Cell* 60:105-113.
41. Summers, J., S. E. Jones, and M. J. Anderson. 1983. Characterization of the genome of the agent of erythrocyte aplasia permits its classification as a human parvovirus. *J. Virol.* 64:2527-2532.
42. Tijssen, P. 1989. Nucleotide sequence and organization of genome of *Galleria mellonella* Densonucleosis Virus (GmDNV). Abstract. EMBO Workshop: Molecular biology of parvoviruses. Kibbutz Ma'ale Hachamisha, Israel.
43. Tseng, B. Y., R. H. Grafstrom, D. Revie, W. Oertel, and M. Goulian. 1978. Studies on early intermediates in the synthesis of DNA in animal cells. *Cold Spring Harbor Symp. Quant. Biol.* 43:263-270.
44. Yakobson, B., T. A. Hymko, M. J. Peak, and E. Winocour. 1989. Replication of adeno-associated virus in cells irradiated with UV light at 254nm. *J. Virol.* 63:1023-1030.

OBJECTIVES

Parvoviruses are small DNA viruses with intricate mechanisms of replication. The overall mechanism of replication is understood but specific details of the replication processes for individual viruses are still not known. To address some of the remaining questions in the field of parvovirus research, two different, but somewhat related, aspects of the biology of parvoviruses will be investigated.

The first area of research will focus on the encapsidation pattern of a parvovirus initially isolated from a human cell line, LuIII. Heteroduplex analysis suggests that LuIII virus is related to the rodent parvoviruses MVM and H-1. Unlike MVM and H-1, which encapsidate primarily the minus strand, LuIII encapsidates both viral DNA strands with equal frequency. Since LuIII, MVM and H-1 replicate in the same cell type, host proteins are not expected to be the sole determining factor for encapsidation. For AAV and the human parvovirus B19, encapsidation of equal amounts of plus and minus DNA is thought to result from the presence of identical left and right terminal palindromic sequences. Even though LuIII encapsidates equal amounts of plus and minus DNA the nucleotide sequences of the left and right termini of LuIII are expected to differ significantly from each other since they do not form panhandle structures during annealing reactions. If true, LuIII, unlike AAV and B19, would lack the symmetry which results from the presence of identical terminal palindromes. If both terminal palindromes of LuIII contained copies of a short identical

nucleotide sequence, this sequence could result in the equal encapsidation of plus and minus strand. Clones containing intact LuIII termini will be constructed and sequenced. If sequence analysis of the LuIII palindromes demonstrates considerable sequence identity to the termini of MVM and H-1, the complete nucleotide sequence of the LuIII genome will have to be determined and analyzed to identify differences among these viruses. Those sequences that differ between the viruses could be responsible for the encapsidation patterns observed for these viruses.

The second area of research involves the study of BPV transcription. Previous studies from this laboratory using S1 nuclease mapping and two dimensional gel electrophoresis showed that BPV codes for a heterogenous population of mRNA species. Hybridization of these RNAs with BPV DNA restriction fragments provided preliminary information on potential splices and on the site of initiation for these transcripts. Further analysis of the 2.6 kb species, likely coding for the capsid proteins, showed that this species contained a 2.25 kb main body with heterogenous 5' termini. These studies also mapped the initiation site for BPV transcripts to map unit 14 of the viral genome. The coding assignment for these transcripts has not been determined. Three non-structural proteins with molecular weights of 83, 75 and 28 kDa and three capsid proteins with molecular weights of 80, 72 and 62 kDa have been identified. The small nonstructural protein, NP-1, is a phosphoprotein that has been shown to have amino acid sequence homology with BPV capsid protein. Unlike the parvoviruses of known sequence, BPV contains a MID ORF which could code for NP-1. In order to further characterize the BPV transcripts and determine their coding relationship with BPV proteins, cDNAs will be prepared from BPV RNAs, amplified by the polymerase chain reaction and analyzed.

My dissertation will address the following objectives and answer the following questions:

1. Determine the nucleotide sequence of clones containing intact left and right termini of LuIII. The nucleotide sequence will tell us whether this virus contains identical ends or short identical nucleotide sequences within both termini, that could be responsible for the equal encapsidation of plus and minus strand DNA by LuIII.

2. If the termini of LuIII are not similar in nucleotide sequence, the complete nucleotide sequence of the LuIII genome will be determined. Comparison of this sequence with those of MVM and H-1 will allow us to identify a putative encapsidation signal(s), describe the genomic organization and compare the well-characterized regulatory sequences and splice-donor/acceptor consensus sequences of MVM and H-1 with possible similar sequences in LuIII.

3. Determine the nucleotide sequence of BPV cDNAs clones constructed either by conventional methods or generated by amplification of cDNAs by the polymerase chain reaction. We will use the sizes and genome location of the amplified fragments, in conjunction with the known sizes of the BPV-coded proteins and transcripts, to refine the BPV transcription map obtained from earlier studies.

Chapter II

IDENTICAL ENDS ARE NOT REQUIRED FOR THE EQUAL ENCAPSIDATION OF PLUS AND MINUS STRAND PARVOVIRUS LUIII DNA

INTRODUCTION

Parvoviruses are small, icosahedral, single-stranded DNA viruses with characteristic palindromes at the left and right termini. Studies of defective interfering particles (14, 15), genomes with terminal deletions (17, 19, 26, 30) and site specific mutants (18) suggest that the "cis" signals necessary for DNA replication and encapsidation reside in the genomic termini, possibly in the 200 nucleotides at each end. The helper dependent adeno-associated virus (AAV) has identical terminal

repeats and separately encapsidates equal amounts of plus and minus strands (6, 8, 16). All autonomous parvoviruses of known sequence have non-identical ends (with the possible exception of B19)(28), and their encapsidation patterns allow separation into three distinct subgroups (For a review see 13, 32). The rodent parvoviruses minute virus of mice (MVM)(4), rat virus (RV), (3, 25) and H-1 (22, 23) have closely related terminal sequences and encapsidate 99% minus strand. Bovine parvovirus (BPV)(10) and lapine parvovirus (LPV; Metcalf et al., in preparation) both encapsidate about 90% minus strand. Parvoviruses B19 and LuIII have encapsidation patterns similar to that of the defective parvovirus AAV. Like AAV both viruses encapsidate both plus and minus strands with equal frequency (5, 33). The equal production of both strands of AAV is thought to result from the symmetry of the terminal palindromes (7). This may be the case for B19 (28) but not for LuIII. Bates et al., (5) showed that unlike AAV, LuIII, given appropriate reannealing conditions, did not form panhandles, suggesting that it has non-identical ends. We have sequenced the left and right termini and found them to be nonidentical; therefore, identical termini are not required for equal encapsidation of plus and minus strands.

MATERIALS AND METHODS

LuIII was propagated in newborn human kidney cells (NBE) transformed by SV40 (29). The virion DNA was purified as described previously for BPV (30) resulting in fully double-stranded DNA due to reannealing of plus and minus strands. This DNA was end repaired with the Klenow fragment of *E. coli* DNA polymerase I (Amersham Corp., Arlington Heights, Illinois) and digested with either EcoRI or HindIII. LuIII has one recognition site for each of these restriction enzymes (21), at map units 21 and 51, respectively. The vectors were prepared by digesting pUC18 DNA with SmaI and EcoRI and plasmid pUC19 DNA with SmaI and HindIII followed by treatment with calf intestinal phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, Indiana). The

two HindIII fragments of LuIII were cloned into pUC19 (Fig. 8A) and the two EcoRI fragments were cloned into pUC18 (Fig. 8B). Ligations were carried out at 15 C for approximately 18 hours. Restriction enzymes were purchased from Bethesda Research Laboratories Inc. (Gaithersburg, Maryland).

The resulting recombinant plasmids were sequenced directly by the dideoxy method (27). Two primers which anneal to plasmid sequences just outside the multiple cloning site were used, one to obtain plus strand sequence of the left end and the other to obtain minus strand sequence of the right end. Based on this sequence, two additional primers were synthesized with sequences complementary to the viral DNA just inside the terminal palindromes to obtain the sequence of the opposite strands. At the left palindrome, the primer annealed to bases 157-171 of the plus strand and provided minus strand sequence. At the right palindrome, the primer annealed to bases 279-291 of the minus strand (see Fig. 9 and 10 for numbering convention) and provided plus strand sequence. Sequencing reactions were carried out using Klenow fragment or avian myeloblastosis reverse transcriptase (Boehringer Mannheim Biochemicals, Indianapolis, Indiana) and ^{35}S -dATP (500 Ci/mmol) or ^{35}S -dCTP (1290 Ci/mmol, New England Nuclear Research Corp. Boston, Massachusetts) as the label. The nucleotide analogue deazaguanosine triphosphate (deaza GTP, American Bionetics, Hayward, California) was used in some reactions with Klenow fragment to resolve regions high in G-C content (20).

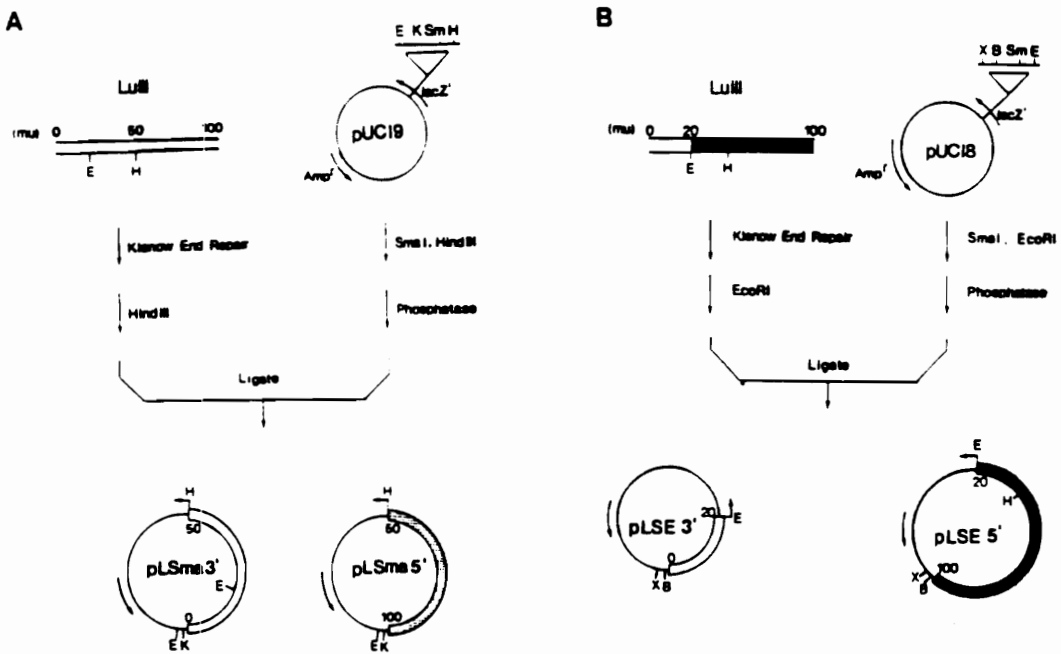


Figure 8. Strategy for cloning LuIII genome fragments into pUC18 and pUC19: Restriction sites of the vector point outward, while restriction sites of the LuIII insert point inward. **A.** Clones of the pLSma series were constructed by ligation of the HindIII fragments of LuIII into the HindIII-SmaI sites of pUC19. **B.** Clones of the pLSE series were constructed by ligation of the two EcoRI fragments of LuIII into the EcoRI-SmaI sites of pUC18. Abbreviations for the restriction enzyme sites: B, BamHI; E, EcoRI; H, HindIII; K, KpnI; Sm, SmaI and X, XbaI.

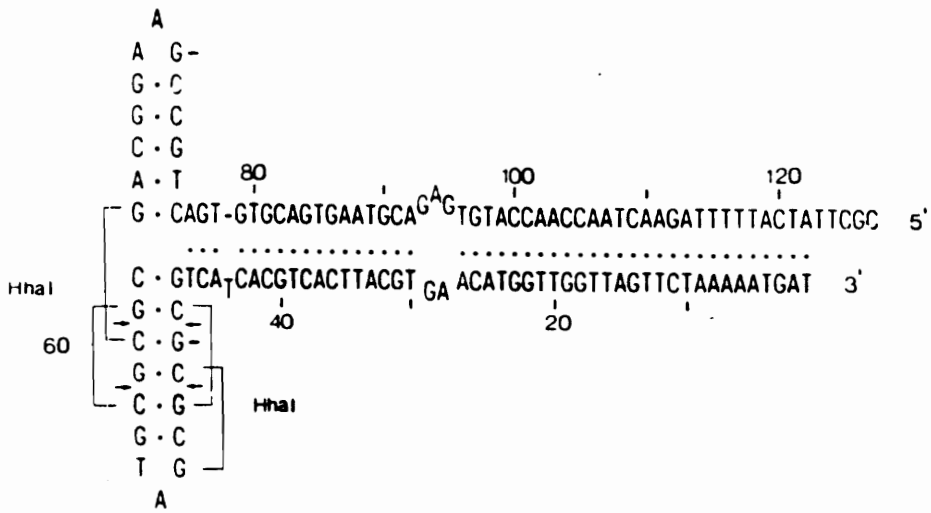
RESULTS

The left palindrome of LuIII consists of 122 bases which can assume a T-shaped intra-strand base paired structure. The sequence of this terminus was obtained by analyzing 15 independent left end clones. Nine of these clones began 5'ATC (Fig. 9A). The other eight each had a different terminal nucleotide and therefore appeared to represent terminal deletions ranging in size from 11 to 37 nucleotides.

The sequence of the left terminus of LuIII is virtually identical to the left terminus of MVMp (4), and of H-1 (23), (Fig. 9B). All 15 clones had the sequence conformation designated flip by Astell et al., (4) in studies of MVMp. One of the arms of the LuIII hairpin differs from the published sequence of MVMp in having an A-T base pair at nt 64 and 74 whereas MVMp has an unpaired T at nucleotide 69 (4). Astell et al. (2) recently suggested that this T residue may in fact occur at nucleotide 70, the equivalent to the T residue at nucleotide 74 in LuIII. However, the sequence of MVMi shows the A-T pair at the same location as in LuIII (24). As these differences are unlikely to affect the stability of the hairpin, the secondary structure of the left palindrome appears fully conserved between the viruses.

The right palindrome of LuIII can assume a 211 nt U-shaped intra-strand base paired hairpin. A portion of this terminal sequence was obtained from the analyses of four independent clones, while the sequence of nt 131-229 was obtained from a clone with a deletion at the right terminus. The structure of the right hairpin in the two possible orientations is shown in Fig. 10A. When the sequence of the right palindrome of LuIII was compared to that of parvoviruses MVMp and H-1 (Fig. 10B), only minor differences were observed. Nucleotides in LuIII which are not in common with MVMp and H-1 have their complements altered as well to maintain base pairing within the right palindrome. The differences seen are unlikely to affect the overall secondary structure of the hairpin.

A



B

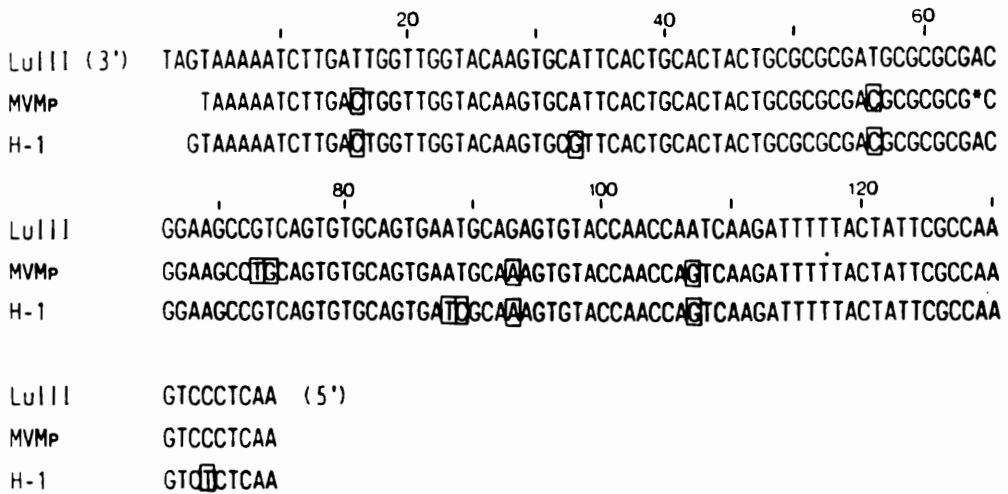


Figure 9. DNA sequence at the left palindrome of the minus strand of LuIII and comparison to that of MVMp and H-1: A. DNA sequence at the left terminus of the minus strand of LuIII. B. Comparison of the DNA sequence at the left termini of rodent parvoviruses MVMp (4) and H-1 (22) and LuIII. The cleavage sites for HhaI are indicated by arrows. Boxes indicate nucleotides nonhomologous to LuIII sequence. Spaces required for maximal alignment of the sequences are indicated by *.

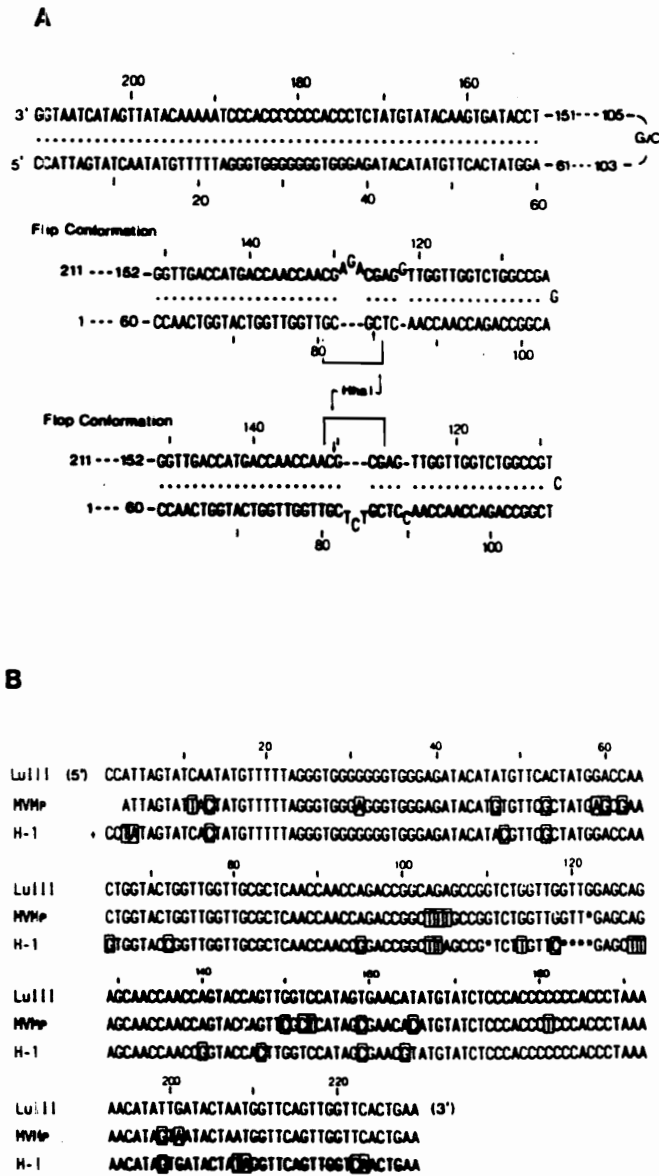


Figure 10. DNA sequence at the right palindrome of the minus strand of LuIII and comparison to that of MVMP and H-1: A. DNA sequence at the right terminus of the minus strand of LuIII. B. Comparison of the DNA sequence at the right termini of rodent parvoviruses MVMP (4) and H-1 (23) and LuIII. The cleavage sites for HhaI are indicated by arrows. Boxes indicate nucleotides nonhomologous to LuIII sequence. Spaces required for maximal alignment of the sequences are indicated by *.

The terminal sequence inversions of LuIII were analyzed by digestion with the restriction enzyme HhaI (3). The HhaI recognition sequence (GCG/C) occurs within the arms of the left hairpin (Fig. 9A) and in an unpaired region of the right hairpin (Fig. 10A) of LuIII. The HhaI site of the left end may be digested in either single-stranded or double-stranded (reannealed plus and minus strands) virion DNA, while the right end site can be digested only in double-stranded virion DNA. This experiment was performed by Dr. Bruce Shull. To confirm the results described above on the sequence analysis on the LuIII termini, Dr. Shull's data are included in the text.

Reannealed double-stranded LuIII virion DNA was end labeled at the 3' hydroxyl with ³⁵S-ddATP (1372 Ci/mmol, New England Nuclear Research Corp., Boston, Massachusetts) followed by digestion with HhaI. This results in the labeling of the minus strand at the left end and the plus strand at the right end. The labeled fragments were run on a sequencing gel using M13mp18 sequencing reactions as size markers (Fig. 11). Four possible fragments resulting from HhaI digestion were expected from the left end of LuIII as a consequence of the two overlapping restriction sites (Fig. 9A). Digestion of DNA in the flip conformation at the left end would result in fragments of 50, 52, 59, and 61 nucleotides in length, while digestion of flop conformation DNA would result in fragments of 61, 63, 70 and 72 nucleotides. Strong bands of 47, 49 and 56 nucleotides were observed which corresponded well with the expected fragment sizes for DNA whose terminus of the minus strand was in the flip orientation. The secondary structure assumed by the fragments during migration in the sequencing gel could be responsible in part for the observed deviation (9). The extra bands of lesser intensity likely reflect the heterogeneity in the position of the terminal nucleotide of LuIII, as seen in the cloning studies described above and for other parvovirus genomes (3, 16, 30). No bands of sufficient length to correspond to the flop orientation of the left terminus of the minus strand were observed. Because of the inability to label the 5' end of the plus strand (unpublished experiment), no conclusions can be made about the conformation of this strand at the left terminus.

Based on the sequence of the cloned right terminus of LuIII (Fig. 10A), two bands of 82 and 129 nucleotides, corresponding to the flip and flop conformations, are expected after end labeling and HhaI digestion. Strong bands are observed at 99 nt, 17 nucleotides longer than expected, and

at 144 nt, 15 nucleotides longer than expected, for the flip and flop conformations respectively (Fig. 11). The difference between the observed and expected sizes should be the same for the flip and flop fragments, and the small deviation seen could be due to the secondary structure assumed by the fragments during migration or to the heterogeneity at the ends of virion DNA molecules as mentioned above. Irrespective, the data suggest that the mature right terminus is longer than the cloned terminus. Since only the 3' end of the plus strand was labeled no information is available on the length of the minus strand. The sequence predicted for the ultimate nucleotides suggests that they could form a small hairpin, making molecules with this hairpin inaccessible to cloning. The presence of a covalently linked terminal protein at the mature terminus, similar to the one described for MVM (12), might also interfere with the efficient cloning of full length DNA molecules. Attempts to label the 5' end of LuIII virion DNA were unsuccessful (unpublished results), suggesting that the 5' end of both LuIII plus and minus DNA strands may also be blocked by a terminal protein. These extra nucleotides, if actually present at the right terminus, are not crucial for viral replication. A LuIII genomic clone with a right terminus identical to that shown in Fig. 10A was highly infectious (N. Diffoot, unpublished results).

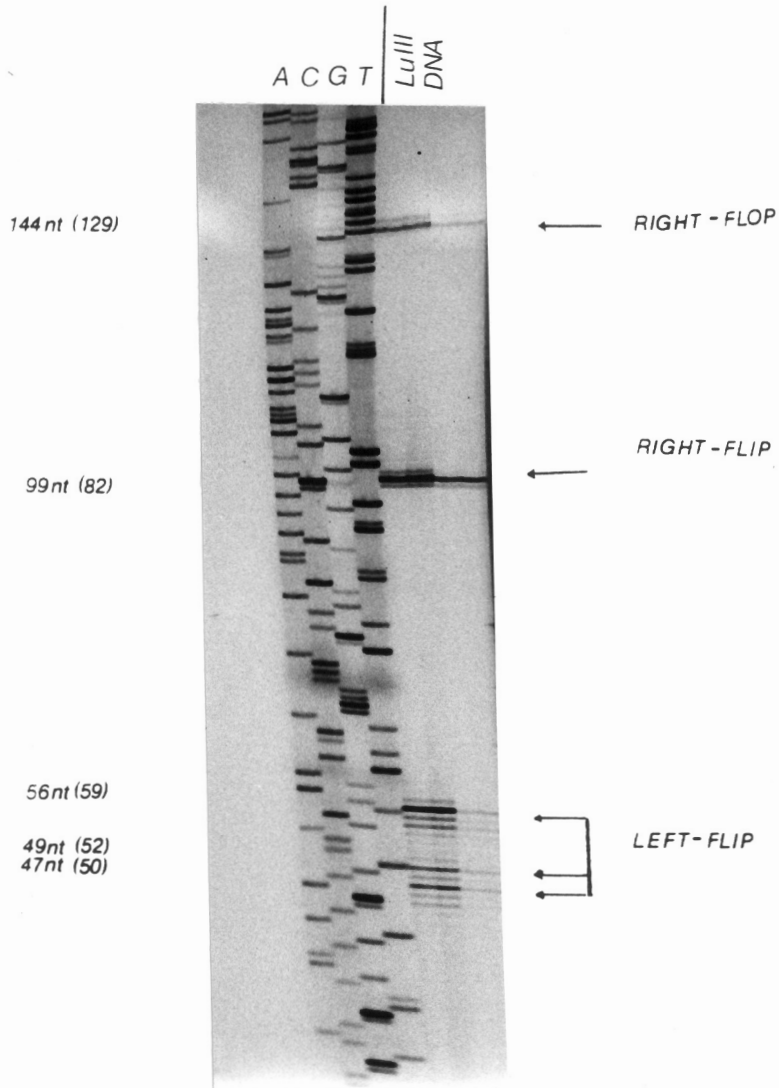


Figure 11. End-label analysis of double-stranded LuIII virion DNA: Reannealed plus and minus-strand virion DNA was end-labeled and digested with HhaI for analysis of sequence inversions at the termini. The labeled fragments were run on a sequencing gel using M13mp18 sequencing reactions as size markers. The positions of the observed and expected (shown in parenthesis) sizes of fragments in the flip and flop conformations are indicated.

DISCUSSION

The sequence homology between the termini of LuIII, MVM and H-1 parvoviruses, the conservation of secondary structure in spite of minor sequence differences and the unique flip conformation at the left terminus of the minus strand of LuIII, suggest that this virus is closely related to the rodent parvoviruses, even though it was originally isolated from a human cell line (31). The finding of equal amounts of plus and minus strand LuIII progeny DNA with a unique conformation at the left terminus of the minus strand means that neither the replication model for AAV (7) nor the replication model for MVM (1) can explain the replication of LuIII. The replication model for AAV predicts hairpin transfer at both (left and right) ends with equal efficiency. This would result in equal amounts of plus and minus strand DNA concomitant with equal proportions of flip-flop at both ends. Conversely, the replication model for MVM explains the formation of progeny DNA with only one conformation at the left terminus, but simultaneously limits the nature of the progeny DNA to strands of only minus polarity, in contrast to LuIII.

The available data (3, 9) suggest that the virion DNA is not positively selected at the encapsidation step. A kinetic analysis (11) proposes that the characteristic distribution of virion DNA is defined by the differential rates of hairpin transfer at various 3' hydroxyl terminal palindromes. These rates determine the distribution of amplified replicative intermediates during the infectious cycle and the efficiency with which replicative forms generate single-stranded DNAs of characteristic polarity and terminal orientation which are then packaged with equal efficiencies. The differential rates of hairpin transfer are determined by the flip-flop conformation of the terminal palindrome, as well as by cellular and viral factors.

It is unlikely that the signals necessary for strand selection during replication are defined solely by the termini, otherwise this would predict encapsidation of only minus strand LuIII DNA, based on its virtual sequence identity with MVM DNA, unless the minor nucleotide differences found between the terminal palindromes of LuIII and MVMp are sufficient to account for the different

encapsidation patterns observed for the two viruses. Further studies on the mechanism of replication of LuIII will be of great importance in understanding the biology of parvoviruses since LuIII shares characteristics with two highly studied extremes of the parvovirus spectrum, AAV and the rodent parvoviruses MVM and H-1.

LITERATURE CITED

1. 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.
2. Astell, C. R., E. M. Gardiner, and P. Tattersall. 1986. DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain. *J. Virol.* 57: 656-669.
3. Astell, C. R., M. Smith, M. B. Chow, and D.C. Ward. 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at the origins of DNA replication. *Cell* 17:691-703.
4. Astell, C. R., M. Thompson, M. Merchlinsky, and D. C. Ward. 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. *Nuc. Acids Res.* 11:999-1018.
5. Bates, R. C., C. E. Snyder, P. T. Bannerjee, and S. Mitra. 1984. Autonomous parvovirus LuIII encapsidates equal amounts of plus and minus DNA strands. *J. Virol.* 49:319-324.
6. Berns, K. I., and R. A. Bohensky. 1987. Adeno-associated viruses: an update. *Adv. Vir. Res.* 32:243-306.
7. 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.
8. Berns, K. I., and M. Labow. 1987. Parvovirus gene regulation. *J. Gen. Virol.* 68:601-614.
9. Chen, K. C., B. C. Shull, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Analysis of the termini of the DNA of bovine parvovirus: demonstration of sequence inversion at the left terminus and its implication on the replication model. *J. Virol.* 62:3807-3813.
10. Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates. 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. *J. Virol.* 60:1085-1097.
11. Chen, K. C., J. J. Tyson, M. Lederman, E. R. Stout, and R. C. Bates. 1989. A kinetic hairpin transfer model for parvoviral DNA replication. *J. Mol. Biol.* 208:283-296.
12. Cotmore, S. F., and P. Tattersall. 1988. The NS-1 polypeptide of minute virus of mice is covalently attached to the 5' termini of duplex replicative-form DNA and progeny single strands. *J. Virol.* 62:851-860.
13. Cotmore, S. F., and P. Tattersall. 1987. The autonomously replicating parvovirus of vertebrates. *Adv. Vir. Res.* 33:91-174.

14. de la Maza, L. M., and B. J. Carter. 1980. Molecular structure of adeno-associated virus variant DNA. 255:3194-3203.
15. Faust, E. A., and D. C. Ward. 1979. Incomplete genomes of the parvovirus minute virus of mice: selective conservation of genome termini, including the origin for DNA replication. *J. Virol.* 32:276-292.
16. Fife, K. H., K. I. Berns, and K. Murray. 1977. Structure and nucleotide sequence of the terminal regions of adeno-associated virus DNA. *Virology* 78:475-487.
17. Labow, M. A., L. H. Graf, Jr., and K. I. Berns. 1987. Adeno-associated virus gene expression inhibits cellular transformation by heterologous genes. *Mol. Cell. Biol.* 7:1320-1325.
18. 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.
19. Merchlinsky, M. J., P. J. Tattersall, J. J. Leary, S. F. Cotmore, E. M. Gardiner, and D. C. Ward. 1983. Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J. Virol.* 47:227-232.
20. Mizusawa, S., S. Nishimura, and F. Seela. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nuc Acids Res.* 14:1319-1324.
21. Rhode, S. L. 1978. Replication process of parvovirus H-1 X. Isolation of a mutant defective in replicative-form DNA replication. *J. Virol.* 25:215-223.
22. Rhode, S. L., and B. Klaassen. 1982. DNA sequence of the 5' terminus containing the replication origin of parvovirus replicative form DNA. *J. Virol.* 41:990-999.
23. Rhode, S. L., and P. R. Paradiso. 1983. Nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. *J. Virol.* 45:173- 184.
24. 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. *Nuc. Acids Res.* 13:3617-3633.
25. Salzman, L. A., and P. Fabisch. 1979. Nucleotide sequence of the self-priming 3' terminus of the single-stranded DNA extracted from the parvovirus kilham rat virus. *J. Virol.* 30:946-950.
26. Samulski, R. J., A. Srivastava, K. I. Berns, and N. Muzyczka. 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33:135-143.
27. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
28. Shade, R. O., B. 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.
29. Shein, J. M., and J. F. Enders. 1962. Multiplication and cytopathogenicity of simian vacuolating virus 40 cultures of human tissues. *Proc. Soc. Exp. Biol. Med.* 109:495-500.

30. Shull, B. C., K. C. Chen, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Genomic clones of bovine parvovirus: construction and effect of deletions and terminal sequence inversions on infectivity. *J.Virol.* 62:417-426.
31. Siegl, G. 1976. Parvoviruses isolated from human cell lines, p. 72-85. in *The parvoviruses*. Springer-Verlag, New York.
32. Siegl, G., R. C. Bates, K. I. Berns, B. J. Carter, D. C. Kelley, E. Kurstak, and P. Tattersall. 1985. Characteristics and taxonomy of parvoviridae. *Intervirology* 23:61-73.
33. Summers, J., S. E. Jones, and M. J. Anderson. 1983. Characterization of the genome of the agent of erythrocyte aplasia permits its classification as a human parvovirus. *J. Virol.* 64:2527-2532.

Chapter III

**THE SEQUENCE OF PARVOVIRUS LUIII
AND LOCALIZATION OF A PUTATIVE
SIGNAL RESPONSIBLE FOR ITS
ENCAPSIDATION PATTERN**

INTRODUCTION

Parvoviruses are small animal viruses with linear single-stranded DNA genomes of either plus or minus polarity. The parvoviral genomes have short palindromic sequences at both ends of the DNA molecule. These terminal sequences can be heterogeneous. Each end may display two alternative conformations ("flip" and "flop"), in a ratio characteristic for an individual virus. Studies of both defective interfering particles and genomes with mutations in the terminal regions

(11,19,26,41) show that these palindromes are essential for replication and encapsidation. The ratio of plus to minus DNA strands encapsidated varies among the Parvoviridae. The dependent parvovirus adeno-associated virus (AAV) encapsidates separately the plus and minus DNA strands with equal frequency. AAV has identical T-shaped imperfect palindromic sequences at its left and right termini. The presence of identical termini was suggested to be necessary for the encapsidation pattern observed for AAV (9,10). The autonomous parvovirus B19 also encapsidates both strands with equal frequency (43) and it, too, has identical termini (39). Bovine parvovirus encapsidates 10% plus strands and 90% minus strands, and it has nonidentical termini (12). The rodent parvoviruses, minute virus of mice (MVMp), H-1, and rat virus are at the other end of the encapsidation spectrum, encapsidating almost exclusively the minus strand. Their left and right palindromes have different nucleotide sequences and secondary structures (2,3,4,16,32,36). The parvovirus LuIII, which has been shown to be closely related to MVMp and H-1 (6), shows a completely different encapsidation pattern. It encapsidates both the plus and minus strand with equal frequency (7), yet the termini of LuIII share over 90% sequence identity with those of MVMp and H-1 (18). These findings indicate that identical ends are not the critical requirement for equal encapsidation of plus and minus strands and that there are nucleotide sequences other than the termini that influence the encapsidation pattern.

The factors determining the unique encapsidation patterns observed among different parvoviruses are unknown. The pattern could be due to selective synthesis, selective encapsidation of a particular DNA strand or a combination of both mechanisms. Studies of the distribution of flip and flop conformations at the termini of replicating form DNAs and virion DNAs of MVMp (1) and BPV (12) argue against selective encapsidation since the ratios of flip to flop conformations at the termini were the same for RF and virion DNA. A kinetic hairpin transfer model for parvoviral DNA replication proposes that these flip/flop ratios result from differential rates of hairpin transfer dependent on the sequences and conformations of the termini (14).

Since MVMp and LuIII infect the same cell type, have very similar terminal sequences but show different encapsidation patterns, we set out to determine the nucleotide sequence of LuIII to search for possible sequences that could influence this process. Comparison of the nucleotide se-

quence of LuIII with those of MVMp and H-1 shows that these viruses are virtually identical with respect to genome organization, location of regulatory signals, mRNA splicing patterns and amino acid sequences of viral proteins. The presence of a putative encapsidation signal in LuIII is discussed.

MATERIALS AND METHODS

Materials

Restriction endonucleases and other enzymes were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, MD), and Boehringer Mannheim Biochemicals, Inc. (Indianapolis, IN). The Sequenase kit (United States Biochemical, Inc. Cleveland, OH) was used for sequence determination. ³⁵S- and ³²P- deoxyribonucleoside triphosphates were obtained from New England Nuclear Corp. (Boston, MA). Oligonucleotide primers used for DNA sequencing reactions were synthesized using an Applied Biosystems Model 381A oligonucleotide synthesizer (Foster City, CA).

Plasmid propagation

Transformation of JM107 and DH5 α *E. coli* cells with pUC18 and pUC19 vectors containing LuIII inserts was done according to the Hanahan procedure (Fig. 12) (20). Plasmid DNA was isolated as described by Rodriguez and Tait (34) and purified by banding on CsCl-ethidium

bromide gradients (23). Recombinant clones containing full-length LuIII inserts were screened by restriction endonuclease digestion and verified by sequencing (37) and Southern blot analysis (23).

Analyses of infectivity of genomic clones

Newborn human fetal kidney (NBE 324K) cells (40) were grown in monolayer culture and maintained in Eagle's minimal essential medium (MEM) supplemented with 10% fetal bovine serum (27). Cells (1×10^6 per 60 mm or 1.5×10^6 per 100 mm diameter dish) were transfected with LuIII recombinant plasmids by the DEAE-Dextran method (22) 24 to 48 hours after being plated. Cell layers were observed for up to 10 days for cytopathic effect (CPE). Infectivity of genomic clones was also analyzed by plaque assay, indirect immunofluorescence (41), and hemagglutination (8) as described previously. For indirect immunofluorescence, the first antibody was rabbit antiserum prepared against LuIII virions (gift from Dr. G. Siegl) and the second antibody was fluorescein-conjugated goat anti-rabbit IgG (Cooper Biomedical Inc., West Chester, PA).

Sequencing of LuIII DNA and computer analysis

Recombinant clones were sequenced by the dideoxy method (37). LuIII-specific primers were synthesized to sequence overlapping clones and to obtain extended sequence from large fragments. The Pustell and Kafatos (29) sequencing programs (International Biotechnologies Inc., New Haven, CT) and the UWGCG (17) programs were used for sequence analyses.

Nucleotide sequence accession number

The GenBank accession number of the LuIII sequence given in Fig.13 is M81888.

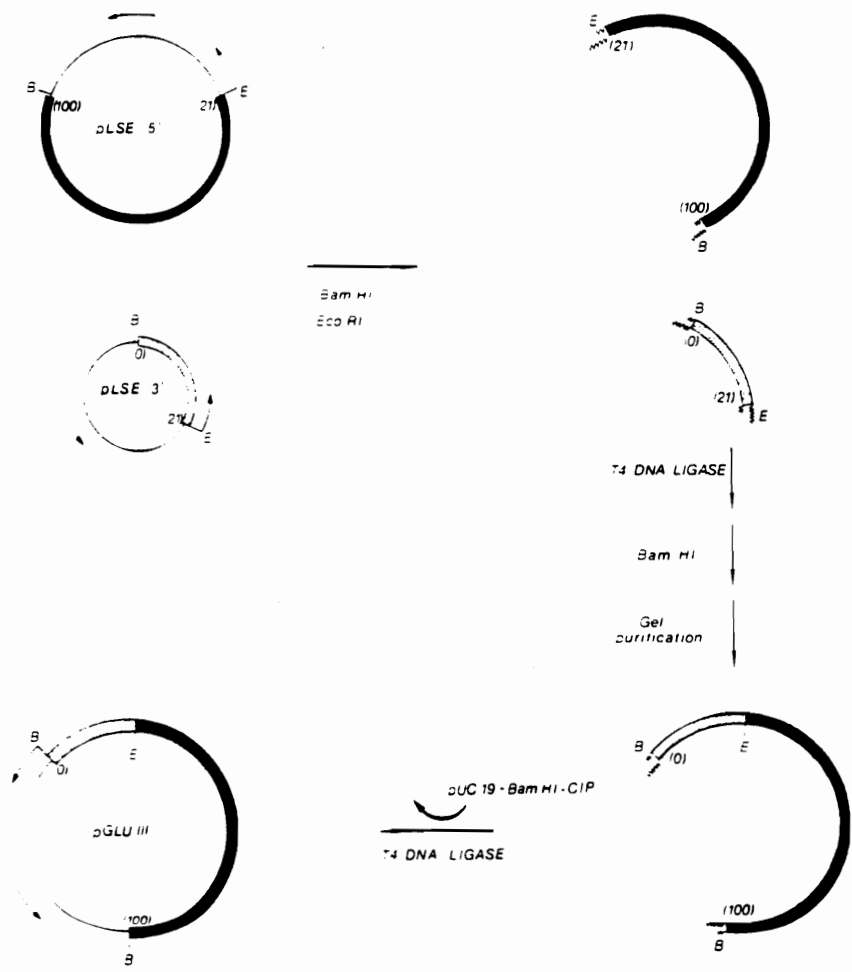


Figure 12. Strategy for cloning the LuIII genome: pLSE clones were described earlier (18). Solid and dotted thick segments represent the LuIII genome. Thin segments represent the vector. Numbers in parentheses refer to map units of the LuIII genome. Arrows indicate the 5' to 3' direction of the lac Z gene in the vector. Restriction sites of the vector point outward and restriction sites of the LuIII insert point inward. Abbreviations for the enzymes: B, BamHI; E, EcoRI; CIP, calf intestinal phosphatase.

RESULTS

LuIII genomic clone construction

A full-length LuIII genomic clone was constructed from previously described subgenomic clones of the pLSE series (18) (Fig. 12). The two LuIII *EcoRI*-*BamHI* fragments were excised from their respective vectors and ligated through the unique *EcoRI* site, resulting in a full-length (0-100 m.u.) LuIII molecule. To minimize the number of undesired species, the ligation mix was digested with *BamHI*. The full-length LuIII DNA ligation product was purified and ligated into the *BamHI* site of both pUC18 and pUC19. Screening of over 900 white colonies yielded more than 100 recombinants containing apparently full-length LuIII inserts, as verified by gel electrophoresis and Southern blot analysis. These were designated as pGLU (pUC-genomic-LuIII) clones. Further characterization of these clones was done by sequence analysis. The majority of the genomic clones contained an identical deletion from nt 4925 to nt 5033, within the right terminus. These clones were still capable of causing CPE in transfected cells within nine to ten days. Nine full-length clones with intact termini were obtained. The intact genomic clones were found to be stable during propagation in *E. coli* JM107 and DH5 α cells and one clone, pGLU 883, was selected for further studies.

Infectivity of LuIII genomic clones

The full-length genomic clone, pGLU 883 was found to be infectious. CPE was observed within five to seven days after transfection. Plaque assays were performed to assess quantitatively the transfection efficiency of the genomic clone. Maximum efficiency of 1×10^2 plaques per μg

was observed when 0.4 μ g of plasmid DNA were used to transfect 1×10^6 cells in 5 ml of medium. Replication of the LuIII DNA was demonstrated by the presence of *DpnI*-resistant replicative form DNA isolated from cells seven days post-transfection upon probing with radioactively labelled LuIII virion DNA. The presence of viral proteins and their nuclear location were detected by indirect immunofluorescence and the pattern was indistinguishable from that seen in cells infected with LuIII virions. Viral progeny resulting from transfection with pGLU 883 were also positive by hemagglutination (data not shown).

LuIII sequence determination

The nucleotide sequence of LuIII was determined from clones containing a nested set of deletions created by exonuclease III (5). The size of the LuIII genome was 5135 bases. Its complete sequence and a comparison with the sequences of the rodent parvoviruses MVMp (4) and H-1 (32) are shown in Fig. 13. The left and right palindromes of LuIII can assume a T- and a U-shaped intra-strand base-paired structure, respectively. The left and right terminal sequences shown in the figure are designated arbitrarily as the flip sequences. As evident from Fig. 13, LuIII is closely related to MVMp and H-1, showing over 80% sequence identity with each virus.

Figure 13. Sequence comparison of LuIII (LU3), MVMp (4) and H-1 (32): Dashes indicate nucleotides identical to the LuIII sequence. Spaces added for maximal alignment of the sequences are indicated by asterisks. Nucleotide and amino acid sequences conserved among parvoviruses, the A-T rich region of LuIII and the direct repeats of MVMp and H-1 are indicated by boxes. Splice-donors, acceptors, and initiation sites for mRNAs are indicated by arrows.

	LEFT TERMINAL REPEAT		
LU3	ATCATTTTTTAGAACTAACCAACCATGTTACGTAAGTGACGTGATGACGC	50	
MVM	***-----G-----	50	
H-1	**-----G-----C-----	50	
LU3	GCGCTACGCGCGCTGCCTTCGGCAGTCACACGTCACCTTACGTCTCACATG	100	
MVM	-----G-----*-----AC-----T-----	96	
H-1	-----G-----AG---T-----	98	
LU3	GTTGGTTAGTTCTAAAAATGATAAGCGGTTTCAGGGAGTTT*AAACCAAGG	149	
MVM	-----C-----*-----	145	
H-1	-----C-----A-----G-----	148	
			P4
LU3	CGCGAAAAGGAAGTGGGCGTGGTTTTAAGTATATAAGCGACACGTTAAGT	199	
MVM	-----A-----A--TAC-G----	195	
H-1	--G---C-----C-AACT--AGTCAC-CTG--	198	
	↓ START OF mRNA R1 (NS1) AND R2 (NS2)		
LU3	CAGTTACTTACTCTTTTCGCTTATTCTGTAAGTCGAGACACACAGAGTAAC	249	
MVM	-----TCT---TT-C-----G-----**--G	243	
H-1	-G-----C---*G-TT-C---TCTG-GT-T-T--G-----*-AGCG	246	
	START OF NS1 AND NS2		
LU3	CAACTAACCAACTAGCCATGGCTGGAACCGCTACTCTGATGAAGTTTTG	299	
MVM	AG-G-----A-----T--T-----	293	
H-1	AG-----A-----T---C---G-----	296	
LU3	GGAACAACCTAACTGGTTGAAGGATAAGAGCAACCAGGAAGTATTCTCATT	349	
MVM	---G---C-----A---A--A--T-----G-----	343	
H-1	---GT---A-----C-----C--A--T-G-----G--G-----	346	
LU3	TGTTTTTAAAAATGAGGATGTTTCAGCTCAATGGAAAAAATATCGGATGGA	399	
MVM	-----AA-----A--G-----G-----	393	
H-1	-----AA-C--C--A--A-----GG-C-----T-----	396	

LU3	ACAGTTACAGAAAGGAGCTGCAAGAGGAGGAGCTGAAATCTTTACAACGA	449
MVM	-T-----A--A-----G-----C-----C-----	453
H-1	-T-----A-----T--C-----G--C-----	446
LU3	GGAGCTGAAACTACCTGGGACCAGAGCGAGGACATGGAATGGGAATCTTC	499
MVM	-----G-----T-----A-----A-CA-	493
H-1	--G--G--G--C--T-----A-----GAGCG-	496
↓ SPLICE-DONOR FOR R2		
LU3	AGTGGATGAACTGACCAAAAAGCAAGTATTCATTTTTGACTCTTTAGTTA	549
MVM	-----A-----T-----G-----	543
H-1	-----CA-----T-----T-----G-----	546
LU3	AAAAGTGTCTCTTTGAAGTACTGAGCACAAAGAACATAGCTCCTAGTGAT	599
MVM	----A--T-A-----G--T-A-----T--TT----G-----	593
H-1	-G-----T-G-----G--C-----A--	596
LU3	GTTACTTGGTTTGTACAGCATGAATGGGGAAAAGACCAAGGCTGGCACTG	649
MVM	----A-----G--A-----	643
H-1	-----C--G-----G--C-----	646
LU3	TCATGTGCTCATTGGAGGCAAGAAGCTTTAGCCAGGCTCAAGGAAAATGGT	699
MVM	C-----A--A-----A--G-----T--A-----A-----	693
H-1	-----G-----G-----T--AC-----	696
LU3	GGAGGAGACAATTAATGTTTACTGGAGTAGATGGTTGGTAACAGCCTGT	748
MVM	----A--G--C-----C-----	743
H-1	----A--G--GC-----G-----G--T-----	746
LU3	AGCGTGCAGCTATCACCAGCTGAAAGAATTAACCTAAGAGAAAATAGCAGA	799
MVM	-AT-----A--A-----	793
H-1	-AT--T--A--A-----G-----	796
LU3	AGACCAAGAATGGGTTACTCTGCTTACTTATAAGCATAAGCAAACCAAAA	849
MVM	----A-T--G-----A-----	843
H-1	G---AGT-----C--T-----C-----C-----G-	846
LU3	AAGACTATACTAAGTGTGTTTGTCTTTGGAAATATGGTTGCTTACTACTTT	899
MVM	-----C-----CTT-----C--A-----T---	893
H-1	-G-----C-----CTT-----C--A-----T-----	896
LU3	TTAACCAAAAAGAAAATATGTACCAGTCCACCAAGGGACGGAGGCTATTT	949
MVM	-----T-----A-C--T-----A-----	943
H-1	----G-----	946

LU3	TCTCAGTAGTGACTCTGGCTGGAAAACTAACTTTTTGAAAAGAAGGCGAAC	999
MVM	---T--C-----A-----G-	993
H-1	---T--C-----G-----G-	996
LU3	GCCATCTAGTGAGCAAACATATACTGATGACATGCGGCCAGAAACGGTT	1046
MVM	-----C-----	1043
H-1	-----G-----G---AAA-----C	1046
LU3	GAGACCACAGTAACCACAGCCGAGGAACTAAGCGCGGCAGAATTCAAAC	1099
MVM	--A-----T-----	1093
H-1	-----G-----T--A-----G-----	1096
LU3	TAAGAAGGAAGTCTCTATTAAGACTACACTTAAAGAGCTGGTACATAAGA	1149
MVM	---A--A-----T-----A-----G-----A-	1143
H-1	--GAG---G---G---A--C---C---T-----A-	1146
LU3	GAGTAACCTCACCAGAAGACTGGATGATGATGCAGCCAGACAGTTACATT	1199
MVM	-----G-----	1193
H-1	-----	1196
LU3	GAAATGATGGCTCAACCAGGGGAGAAAACCTACTTAAGAATACGCTAGA	1249
MVM	-----T-----G--G--A-----	1243
H-1	-----T-----T-G---A---A-----	1246
LU3	GATCTGTACGCTGACTCTAGCCAGAACCAAACAGCCTTTGACTTGATTT	1299
MVM	---T---A--A-----A-----A---	1293
H-1	-----A-----A-----C	1296
LU3	TAGAAAAAGCTGAAACCAGCAAACCTAACCACCTTTTTACTGGCTGATACA	1349
MVM	-----C---C---C---	1343
H-1	-G-----G-----CCA---AGC--C	1346
LU3	AGAACCTGTAGAATCTTTGCTTTTCATGGCTGGAACCTACATCAAAGTCTG	1399
MVM	-----C---T-----TG-T---T--	1393
H-1	-----GAG-----T--T---CT--	1396
GKRN		
LU3	TCATGCTATTTGTTGTGTCTTGAACAGACAGGGAGGCAAAAAGAAATACTG	1449
MVM	C-----C---T--A-----A-----	1443
H-1	C---C--C-----GC---T---A-----G--C-----	1446
LU3	TTCTGTTTCATGGACCAGCCAGTACAGGCAAATCAATCATTGCACAGGCC	1499
MVM	--T-A-----C-----T--T-----A---	1493
H-1	-G--C---C-----C-----T--T-----A---	1496

LU3 ATAGCACAGGCAGTTGGTAATGTTGGTTGTTATAACGCAGCCAATGTGAA 1549
MVM -----A-----C-----C-----T-----A-- 1543
H-1 -----A-----C-----T--T----- 1546

WVEE

LU3 CTTTCCATTTAATGACTGTACCAACAAGAACTTAATCTGGGTGGAAGAAG 1599
MVM -----G--T-----A----- 1593
H-1 -----A--G--T----- 1596

LU3 CTGGTAACTTTGGACAGCAAGTAAACCAGTTTTAAAGCCATTTGTTCTGGT 1649
MVM -----C----- 1643
H-1 -----C-----A--C-----T-----C 1646

LU3 CAGACCATTTCGCATTGACCAAAAAGGAAAAGGCAGCAAACAGATTGAACC 1699
MVM --A--T-----T----- 1693
H-1 --A--A-----T----- 1696

LU3 AACACCAGTGATCATGACCACAAATGAAAACATCACAGTGGTCAAAATAG 1749
MVM -----C-----G--T-----G----- 1743
H-1 -----T--T-----C--C--G--T--C-----T-G----- 1746

LU3 GGTGTGAAGAGAGACCAGAACACACTCAACCAATCAGAGACAGAATGTTA 1799
MVM -C--C-----A-----C-T 1793
H-1 -C-----G--A-----C-C 1796

LU3 AACATTCATCTGACACATACATTGCCTGGTGACTTTGGTTTGGTTGATAA 1849
MVM -----A-----C-----T-- 1843
H-1 -----C-----G--C-A-----G----- 1846

TAR ELEMENT

LU3 AAACGAATGGCCTATGATATGTGCTTGGTTGGTAAAGAACC GTTACCAAT 1899
MVM ---T-----C--T-----T----- 1893
H-1 GC-----C--C-----G--T----- 1896

LU3 CGACCATGGCAAGTTACTGTGCTAAATGGGGCAAAGTTCCTGATTGGACA 1949
MVM -T-----C-----T-- 1943
H-1 -T-----TT-----T-- 1946

SPLICE-ACCEPTOR R2

LU3 GAAACTGGGCGGAGCCAAAAGTAACGACTGAAATAAATTCCGGTAGGTTC 1999
MVM -----G--GC-A--CCT-----TAC----- 1993
H-1 --GG-----G--GC--GAC--CCT-----C----- 1996

↓ START OF mRNA FOR R3 (VP1 AND VP2)

LU3 AACCAACTCACCAT*****CTCCGAAAAGTACGCCTCTCAGCCAGAACT 2042
MVM GG-ACG-----TCACGA-A----- 2043
H-1 --TGCG-----CTCTGA-----G-----A----- 2046

LU3 ACGCACTAACTC*****CGTCGGATCTCGAGGACCTGGCTCTGGAGCCT 2087
MVM -T-----CACTTG-A-----T-A----- 2093
H-1 ----T--T---CACTTG-A----C--T-C-----A-----A----- 2096

LU3 TGGAGCACACCAAGTACTCCTGTTGTGGGCACTGTCAAAACCCCGAACAC 2147
MVM -----A-----C-----CAG-----A----- 2143
H-1 -----A-----C-----CAGC--G--AA----- 2146

LU3 TGGGGAAACTGGTTCAACAGCCTGTCAAGAAGCTCAACGGAGCCCAACTT 2187
MVM -----G-----C-A-----C----T-G----T----- 2193
H-1 -----GG-----C-----C---GT-----C- 2196

LU3 GGTCCGAGATCGAGGAGGATTTGAGAGCGTGCTTCAGTTCGGAACACTGG 2237
MVM ----A-----G--G-----CG-T- 2243
H-1 -----C-----T-----CAA----G-T- 2246

SPLICE-DONOR a
END OF NS1 ↓ START OF VP1

LU3 AAGAGCGACTCCGAACAGCTACCAAACCTTGGATTAAAGGTACGATGGCGCC 2287
MVM ----AA----T-AGCG---CG-TG-----C----- 2293
H-1 G-----T-A-CG--GAG-TG-C-----C-----A-----A-- 2296

↓ **SPLICE-DONOR b**

LU3 TCCGGCTAAAAGAGCTAAAAGAGGTAAGGGGTTAAGGGATGGTTGGTAGG 2337
MVM ---A-----T-----*****-- 2329
H-1 ---A-----C-----*****-- 2342

SPLICE-ACCEPTOR a ↓

LU3 TTGGTGGGGTATTAATATGTGACTACCTGTTTTACAGGCCTGAAATCACT 2387
MVM -----G-T-A-T----- 2389
H-1 -----C---G-A----- 2392

↓ **SPLICE-ACCEPTOR b**
PGY

LU3 TGGTTCTAGGTTGGGTGCCTCCAGGCTACAAGTACCTGGGACCAGGGAAC 2437
MVM -----T-----T----- 2439
H-1 -----T-----

LU3 AGCCTTAACCAAGGAGAAACCAACCAATCCATCTGACGCTGCTGCTAAAGA 2487
MVM -----G-----C---C----- 2489
H-1 -----G-----C--T-----C---C----- 2492

NPYL

LU3 GCAGGACGAGGCCTACGACCAATACATCAAATCTGGAAAGAAATCCTTACC 2537
MVM -----T--T-----A----- 2539
H-1 A-----A-----A----- 2542

LU3 TGTACTTCTCTCCTGCTGATCAACGCTTCATTGACCAAACCAAGACGCT 2587
MVM -----G-----T-----G-----C 2589
H-1 -----C-----

THE SEQUENCE OF PARVOVIRUS LU111 AND LOCALIZATION OF A PUTATIVE SIGNAL RESPONSIBLE FOR ITS ENCAPSIDATION PATTERN

LU3 AAAGACTGGGGCGGAAAGGTTGGTCACTACTTCTTTAGAACCAAGCGTGC 2637
MVM -----A--C-----T-----C-- 2639
H-1 --G-----C-----T-----A-- 2642

LU3 TTTTGCACCTAAGCTTTCTACTGACTCTGAGCCTGGGACTTCTGGTGTGA 2687
MVM -----G-----A---A-----A- 2689
H-1 -----A---C----- 2692

LU3 GCACAGCTGGTAAACGTACTAAACCACCTGCTCACATCTTTATTAACCAA 2747
MVM ---G-----C---G-----T---T----- 2739
H-1 ---G-C-----A-----T---G-A--T--- 2742

LU3 GCCAGGGCTAAAAAAAAACGTACTTCTCTTGCTGCGCAGCAGAGGACTCA 2787
MVM -----A-----T-----TC-----A-----A--C-G--- 2789
H-1 -----A-----CG-----A-----T 2792

START OF VP2

LU3 GACAATGAGTGATGGCACCGACCAATCTGACAGCGGAAACGCTGTCCAGT 2837
MVM A--C-----AG---C-----C- 2839
H-1 -----***--A-AA--CAACC-G--A---GAATCG 2839

GGG

LU3 CAGCTGCTAGAGTTGAGCGAGCAGCTGACGGTCCTGGAGGCTCTGGGGGC 2887
MVM -----A-----A-----C-----T 2889
H-1 -TAA-----T-----AGG---A-----T 2889

LU3 GGGGGCTCTGGTGGGGTGGGGTTGGCGTTTCTACTGGCAGTTATGATAA 2937
MVM -----C-----T-----C-TC--GTC----- 2939
H-1 -----C-----A---T-----G-C----- 2939

LU3 TCAAACACATTATAAGTTTCTAGGGGATGGGTGGGTAGAGATTACTGCTT 2987
MVM -----G-----GA--CT-G--T--C--C-----A-----AC 2989
H-1 -----GAC-----T-G--A-----A-----A--A-----AC 2989

LU3 ACAGCACACGCATGGTACACTTGAACATGCCTAAATCAGAAAACACTACTGT 3037
MVM TAGCA--TA-AC-A-----T--A-----T--C 3039
H-1 -TGCTT-TA-AC-TT-G-----GGA-----CCT-----C 3039

LU3 AGGGTGC GGTACACAACACAAATGACACAGGTACAGCAAGTCACATGGC 3087
MVM --AA-CA-A--T-----T-----CA-----TCAGTCAA-G-CA----- 3089
H-1 C-C--CAC---T-----T-ATC-AACA-----ACAC-G--C-A-GG-AAA 3089

TPW

LU3 TATGGAC*****GATGCTCATGAACAGATTTGGACACCATGGAGTC 3128
MVM A-AA--*****T-----G--A-----CT 3130
H-1 GGGAA--ATGGCCTAT--CA-A--C---A-----CT 3139

LU3 TGGTTGATGCTAATGCTTGGGGAGITTTGGTTTTCAACCAAGTGACTGGCAG 3178
 MVM ----G-----C-C--G-----A 3180
 H-1 ----A-----C----- 3189

LU3 TACATTTCTAATAATATGATTCACATCAATTTACATTCACTTGACCAAGA 3222
 MVM -----GC--C-CC----GC--GC-T--C--GGTA-----T----- 3224
 H-1 -T----CAA--C-GC---GAATCGC-G---C-TG-C---T-GAG----- 3239

LU3 ATTGTTTAATGTGGTCATCAAAACAGTGACTGAACAGAA***CACAGGAG 3275
 MVM -A-A--C-----A--GC-G-----T--T--A--G--AG-***-TT----- 3277
 H-1 -C-A-----A--AG-----C-----AC-AGGAG-T--CC 3289

YNN

LU3 CTGAGGCCATTAAGGTCATACAACAATGACCTCACTGCTGCCATGATGGTT 3325
 MVM G-C-A--T--A--AA-A-----T--A--TG----- 3327
 H-1 AA--T-----A--T-T-----T-G--G--CTGT----- 3339

LU3 GCTCTTGATTCTAACAACATACTGCCTTACACACCAGCCATAGACAATCA 3375
 MVM --AG-A--C--A-----TT----A-----T--AGC-A--TCAAT 3377
 H-1 ----G--AG-----T--AGCTC-A-CATC 3389

LU3 AGAGACACTTGGTTTTCTATCCATGGAAACCAACCATAACCAAGTCCTTACA 3425
 MVM G--A-----C--C-----G--TCA--A---- 3427
 H-1 ---A-----C-----GC---GC----- 3439

LU3 GATACTATTTTAGCTGTGACAGAACTTATCAGTTA***CTTACAAAGAC 3472
 MVM -G-----T--GT-----G-TC-T-----G-***-C--G--A-T 3474
 H-1 -----C--TT-ATGCCT---C-AC-CAGT--T-CCT--AG--CTCT 3489

LU3 GAAGCAGGAACCAT***CACTGACACAATGGGTTTGGCCAGTGGCCTGAA 3519
 MVM C---A---C--AG-TGAACA-A-TGTG-----AACAC-A-AA--AA---- 3524
 H-1 -CT-A-----TCAAAT---A-----C--T--AGA-C-ACAG-CA--A-- 3539

LU3 CTCCCAATTTTTTACCATTGAGAACACTCAGCGTATTAACCTACTCAGAA 3569
 MVM T--T-----A--A-AA--C-CAT-G----- 3574
 H-1 ---T-----T-----CTT--C-----CT--C--GC-C- 3589

LU3 CTGGGGATGAGTATGCTACTGGAACCTTACTACTTTGACACAGAACCAATC 3619
 MVM -A-----C--A-T--C--A--T-----A-TT--G-T 3624
 H-1 -A--T-----T-A-A-----C--C--AT---A---T--C--C-T 3639

LU3 AGACTAACTCACAGGTGGCAAACCAACAGACACCTGGGTCAGCCTCCACA 3669
 MVM -A---C--A-----C-T--A--T--A-----T 3674
 H-1 -A---T-----A-----T---CATGC-TC-A-GG 3689

LU3 AATTACTGAACTACCAAGCTCTGACACTGCTAACGCTACTTTAACAGCTA 3719
 MVM GC-GT-AACCT-T--TGAAG-----A-GCA-G---AC-T--T---C 3724
 H-1 ---A-----C-----CA--A--T--A--A-CA--AT-AC----T--A- 3739

LU3 GAGGTTACAGATCAGGTCTGACTCAAATTCAGGCAGAAATGATGTGACT 3769
 MVM A---GAG----CAT--AACAA--A-----***G--GGTT--CTGG----G- 3771
 H-1 AT--AG-----TT--ATCA--A----CA--GAATGTG--CT----C--A 3789

LU3 GAAGCTACTAGGGTCAGACCTGCACAGGTTGGATTTTGTTCAGCCTCATGA 3819
 MVM -----A-TC--AAC-----T--A--A-----A--A--CA- 3821
 H-1 --G---TTGC-CAC---G-----T---A----C--CATG--A----- 3839

LU3 CAATTTTGAACCAGCAGAGCGGGCCCTTTCAAGGTTCCGGTAGTGCCAG 3869
 MVM TG-C-----G-----T--A--A--TGCT-CC--AAA---T---- 3871
 H-1 ---C-----G-A-A-----GT--C--A--T-----A--G--A--GC 3889

LU3 CAGACATCACACAAGGCCTAGACCATGATGCCAATGGTAGCCTGAGATAT 3919
 MVM ----T--T--T-----AG-----A-A--A-----C--TG-T-----C 3921
 H-1 T-----A--GCT--GAG-----A--C--AGC-A-AC---T- 3939

LU3 ACCTATGACAAACAACATGGTCAAAGCTGGGCAAGTCA**GAACAACA*A 3966
 MVM -GT---G-----G-----G---AT-----TTCA--TG--C--G--CC 3971
 H-1 -A-----G-----CG--GAT-----C-AA--AG--G--G--CC 3989

LU3 AGACAGGTACACTTGGGATGCTGTAACTATGATTCTGGCAGATGGACTA 4016
 MVM ---GC-C-----A-----AAACA-G--T--G---A--T---GAC--C- 4021
 H-1 ---A-----A-----AA--G-TAG--CAG---G--GGAC--AG 4039

LU3 ACAACTGTTTTATTCAATCAGTACCATTTACATCAGAACC AAAATGCTAAC 4066
 MVM -AG-TG-----C-----C-AGTTGTTCC----CCACTA--T 4071
 H-1 CT-GA--C---G-A---AGT-C-----A-AT-TATTCC-----CCAA--- 4089

LU3 CAAATACTTACTAACCCTGACAACCTAGCGGGTAAGACTGACATACATTT 4116
 MVM GGC--T-----A--TGCAA--CCTA-T-G-AC---A-A-----T---- 4121
 H-1 --G--CT-GCAGCGAGAA--GC-A---T--C-GA---A---G-----A 4139

LU3 TACCAACGCATTTAACAGTTATGGACCACTAACTGCTTTTCCACATCCTG 4166
 MVM -T-A--T-TT-----C-----T-----A--T---C--AA 4171
 H-1 ---T--T-TT-----C-----T-----T-G---A-----T-----A- 4189

Q1W

LU3 CGCCGATTTACCCACAAGGGCAGATTTGGGACAAAGAACTTGATCTTGAA 4216
 MVM GT--TG-A-----T-----A--A--A-----A----- 4221
 H-1 AT--C-----T---A-T--A--A-----T-G--C--G--- 4239

LU3 CACAAGCCAAGACTGCACACACAGGCTCCTTTTGTCTGTAAAAACAATGC 4266
 MVM -----A--T-----T-----T-ACT-----A-----T----- 4271
 H-1 -----A--T-----A---GT-ACT--A--A-----T-----CC- 4289

LU3 TCCAGGTCAGCTTCTGGTTAGGCTAGCACCTAACTTGACTGACCAGTATG 4316
 MVM A--T--A--AA-GT-----AT---G---A---C-A-----A---- 4321
 H-1 A-----A--AT-T---CACT-G-GG-----TC-----A-T-- 4339

LU3 ATCCTAATAGTTCTAACCTATCTAGAATTGTCACCTATGGCACCTTCTTC 4366
 MVM -----A--CG-AG-C-CA--T-----T--A--C--T--A--T--- 4371
 H-1 -C--A--C--CA-A-CTG-T---C-C-----T--A---A---T--T-A- 4389

LU3 TGGAAGGGCAAACCTAACTCTAAAAGCAAAGATGAGACCTAATGCTACTTG 4416
 MVM -----A--A-----CA-G-G-----AC-T---G---CA-C----- 4421
 H-1 -----T-TTT-G-AAT-C-----C--AC-A-----A---CTG--C-- 4439

LU3 GAACCCAGTCTTCCAAATAAGTGCTACCAACCAAGGAACCAATGACTACA 4466
 MVM -----G-A---G-----GAAG--A-T--***---CTCA---- 4468
 H-1 ---T--T--A-A---GC--***-C--AG--TCT-TTG-----TCT---- 4486

LU3 TGAGCATTGAAAGATGGTTACCAACTGCTACTGGCAACATAACAAATGTG 4516
 MVM ----TG-AACT-A-----A-----GCAGTC---- 4508
 H-1 ---ATG--A-G-A---C-C---T---A-----GCAGTC--AT 4536

END OF VP1 AND VP2

LU3 CCTCTGCTTTCTAGACCTGTTGCTAGAAACACTTACTAACTAATGCT 4566
 MVM --G--TA-AA-A-----T-----***** 4551
 H-1 --AT--A--G-----GC--CAC-TG--A--CTAACTAATGCT-----C-----***** 4580

A-T RICH REGION

LU3 CTATGCTTCATATATATTATATATATTATACTAACTAACCATGTTTA 4616
 MVM *****-----C--T 4580
 H-1 *****-A-CTA-GT 4589

POLY A

LU3 CTCTTACATTACTTCATATAATATTAAGACTTAATAAAAATACAACATAGA 4666
 MVM T---T-TG-----T-----G----- 4620
 H-1 T---CTGT--G-----C-----C-T-A-----CT-G-C-----A- 4639

LU3 AATATAATATTACATATAGA*TATAAAGAATAGAATAATATGGTACTTAC 4715
 MVM -----G-----*T---GA-----G 4669
 H-1 -----CAC---ATA-----T---T--A-----AC-----GG-** 4687

LU3 TTACTGTTAGAAATAATAGAACTTTTGAATAACAAGATAGTTAGTTGGT 4765
 MVM -A-----A-----C----- 4719
 H-1 -A-----A-A-----TTAGTTAGTTGGT-----**--T----- 4735

DIRECT REPEAT I

LU3	*****	4765
MVM	<u>TAATGTTAGATAGAATAAGAAGATCATGTATAATGAATAAAAAGGGTGGAA</u>	4769
H-1	<u>TAATGTTAGATAGAAT**ATAAAAAGATTTTGTATT</u> TAAAAATAATATA	4783

DIRECT REPEAT II

LU3	*****TTATGTTATATAGAATATAAGAAGATGATGTACAA	4800
MVM	GGGTGGTTGGTAGGT-A-----G-----**-----C-----T--	4817
H-1	<u>GTTAGTTGGTTAATG---GA-AGA---T**-A---TTT--TAT-TGGG</u>	4831

POLY A

LU3	<u>AGAATAAN</u> AGGGTGGGAGGGTGGTTGGTTGGTACTCCCTTAGACTGAATG	4850
MVM	T-----A-----A-----A-----T-----TG---	4867
H-1	-A-TA-----TT-----	4871

MAJOR POLY A

LU3	TTAGGGACCAAAAAA <u>TAATAAAA</u> *TTCTTGAAAACCCAACAAGGACTAC	4899
MVM	---A-----C--T--T----T---C-A-----	4917
H-1	-----TAAT-***--ATG-----	4928

RIGHT TERMINAL REPEAT

LU3	TGTCATATTCAGTGAACCAACTGAA <u>CCATTAGTATCAATATGATTTTAGG</u>	4949
MVM	----*-----T-C---T-----	4967
H-1	----*- <u>TG-----TA-----C---T-----</u>	4967

LU3	GTGGGGGGGTGGGAGATACATATGTTCACTATGGACCAACTGGTACTGGT	4999
MVM	-----A-----G-----G-----AG-G-----	5016
H-1	-----C-----G-----G-----C---	5027

LU3	TGGTTGCTCTGCTCCAACCAACCAGACCGGCTCTGCCGGTCTGGTTGGTT	5049
MVM	-----*-----AAA-----	5065
H-1	-----AA---G-*****--AGA-----AA-----C-----	5072

LU3	GAGCGCAACCAACCAGTACCAGTTGGTCCATAGTGAACATATGTATCTCC	5099
MVM	-----C-CT-----C-----C-----	5115
H-1	-----G-----C-----C-----G-----	5122

LU3	CACCCCCCACCCTAAAAACATATTGATACTAATGG*****	5135
MVM	-----T-----G-A-----*****	5149
H-1	-----G-----TA--TTCAGTTGGTCAAC	5172

LU3	*****	5135
MVM	*****	5149
H-1	TGAAT	5177

Genomic organization

The organization of the LuIII genome is similar to that of all vertebrate parvoviruses sequenced to date. The stop codons in each of the three reference frames of the LuIII plus and minus strand DNA are shown in Fig. 14. There are two large open reading frames (ORFs) (designated as left and right ORFs respectively), and two small ORFs in the plus strand. No ORFs of significant size were found in the minus strand DNA. Search for promoter-like consensus sequences (TATAA) in LuIII revealed two possible promoters at m.u. 4 (P4, nt 181) and m.u. 38 (P38, nt 1982) (Figs. 13 and 14). The sequence and location of P4 are virtually identical to those of MVMp and H-1. For P38, the sequence in LuIII is AATAA rather than the TATAA found in MVMp and H-1. The transactivation responsive element (TAR) sequence identified in H-1 (33), is present in a similar location in LuIII, upstream of P38. The similarities of P4 and P38 among the three viruses suggest that both LuIII promoters are likely functional. Search for polyadenylation signals in the LuIII sequence revealed three possible signals at map units 90, 94 and 95, whereas MVMp and H-1 have four signals (4,32).

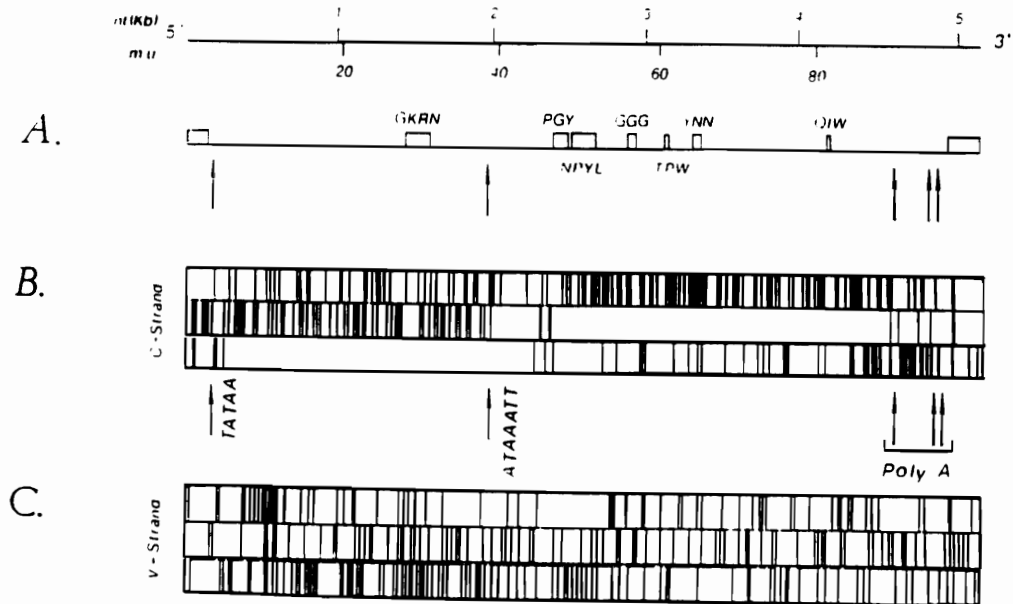


Figure 14. Genomic organization of LuIII: A. The location of the conserved amino acid sequences are depicted by boxes. Position of promoter-like sequences and poly adenylation signals are indicated by arrows. Vertical lines depict stop codons in all three reading frames of B. plus strand (C-strand) and C. minus strand (V-strand).

Assignment of coding domains and transcription map

Based on the genomic organization and conserved amino acid sequences of vertebrate parvoviruses (2,4,13,32,39,42), we assign the left ORF of LuIII to code for the nonstructural proteins NS1 and NS2. The GKRN and WVEE amino acid sequences, conserved in NS1 among parvoviruses of known sequence, are also conserved in LuIII. Most of the amino acid changes in NS1 and NS2 between LuIII and MVMp are conservative (Fig.15). The majority are localized to the carboxyl ends of the proteins. The same degree of variability was observed when MVMp and H-1 were compared. Similarly, we assign the right ORF to code for the major parts of the two capsid proteins, VP1 and VP2 (Fig. 16). Most of the amino acid changes of the capsid proteins between LuIII and MVMp are conservative (Fig. 16). The majority are localized to the carboxyl half of the proteins. Consequently, the amino acid sequence of VP2 differs the most between LuIII and the rodent parvoviruses.

Given the sequence identity of LuIII with MVMp, we propose that the transcription map of LuIII is similar to that of MVMp for the major RNA species R1, R2 and R3 (15,21,25). Three different forms of a small splice would exist in each of the three classes of viral mRNA, resulting from the use of two donor and two acceptor splice sites. The most frequently used pattern would splice from nt 2274, donor a, to nt 2375, acceptor a, and the two less frequently used patterns splice from nt 2310, donor b, to nt 2397, acceptor b or from nt 2274, donor a to nt 2397, acceptor b. In LuIII, the mRNAs R1 and R2, coding for NS1 and NS2, respectively, would be generated from P4. Since NS1 is encoded from a continuous sequence in frame 3 terminating at nt 2270, upstream of the small splices, all three spliced forms of the R1 transcripts are expected to generate identical proteins. R2 transcripts coding for NS2 would contain an additional splice, known in MVMp as the large splice, from nt 520 to nt 1996. These R2 transcripts would splice a second time. The majority of them splice from donor a to acceptor a and a small amount from donor b to either acceptor a or acceptor b. VP1 and VP2 for LuIII would be translated from R3 transcripts generated from P38. Transcripts for VP1 would splice from donor b to acceptor b and initiate translation at

the first ATG, at m.u. 44 (nt 2280) (Fig. 16). The majority of transcripts directing the synthesis of VP2 contain a splice from donor a to acceptor a, a small amount contain a splice from donor a to acceptor b. Both VP2 transcripts would initiate translation at m.u. 54 (nt 2792).

Major sequence differences between LuIII and the rodent parvoviruses

MVMp and H-1

Two regions of the LuIII sequence differ noticeably from those of the rodent parvoviruses. Both stretches of sequence are downstream from the capsid protein coding domain. At m.u. 92 (Fig. 13), MVMp and H-1 contain a direct repeat of about 60 bases, whereas LuIII has only one copy of this sequence, as does the lymphotropic strain of MVM (MVMi) (2). Since each sequence contains an AATAAA signal for polyadenylation, LuIII and MVMi have one signal less than MVMp and H-1.

At m.u. 89, six bases downstream of the stop codon of the capsid gene, LuIII has a unique A-T rich region region 47 bases long (TATGCTCTATGCTTCATATATATTATATATATTATTATACTAACTAA) (Fig. 13). This region contains a direct repeat of the sequence TATATATTA.

A.

LU3	MAGNAYSDEVLTGTTNWLKDKSNQEVFSFVFKNEDVQLNGKNIQWNSYRKEIQEEELKSLQRGAEITVDQSEDMEWESSVDELTKKQVIFDLSLVKKCLFE	100
MVM	-----A-----E-----N-----D-----K-----D-----P-----TT---M-----	100
H-1	-----V-----S-----N-----D-----DD-----A---DM-----	100
LU3	VLSTKNIAPSDVITVFWQHEWGKDOGWCHVLIIGGKNSQAQGGKWRRLQNVYWSRWLVLTACSVQLSPAERIKLREIAEDQEWVLLTYKHKQTKKDYTKC	200
MVM	--N---F-G--N-----D-----N---T-----N-----	200
H-1	-----N-----P-----D---P-----N---T-----S-----H-----	200
LU3	VCFGNMVAYYFLTKKICTSPPRDGGYFLSSDSGWKINFLKEGERHLVSKLYTDDMRPETVETTIVTAQETKRGRQIQTKKEVSIKTLKELVHKRVTSPE	300
MVM	-L---I-----S-----E-K-----A-----RE-----	300
H-1	-L---I-----S-----E-K-----A-----RE-----	300
LU3	DWMHMQPDSYIEMMAOPGGENLLKNTLEICTLTLARTKTAFDLILEKAETSKLTNFLADTRTCRIFAFHGWNYIKVCHAICCVLNROGKRNIVLFGHP	400
MVM	-----S-P-----V-----	400
H-1	-----A-SM-S-----E-----A-----	400
LU3	ASTGKSIIAQAIQAQAVGNVGCYNAANVNFPPNDCTNKNLWVEEAGNFQQVNOFKAICSGQTIKIDQKGGKSKQIEPTPVIMTTNENITVVKIGCEERP	500
MVM	-----R-----	500
H-1	-----R-----	500
LU3	EHTQPIRDRMLNIHLTHLPGDFGLVDKNEWPHICAWLVKNGYQSTMASYCAKRWKVPDWTENWAEPKVITTEINSVGSNSP*S*PKSTPLSQNYALTP*	597
MVM	-----S-----P-P--LL--AR--FTT-----R-----L	600
H-1	-----R-----H---L-----C-----S-D---LD-P--L--MR--SLT-R-----L	600
LU3	*SDLEDLALPEWSTPSTPVVGTVKTPNTGETGSTACQEAQRSPTWSEIEEDLRACFSSEHWKSD*SEQLPNLD	668
MVM	A-----N---A--AE-Q---A-K---DG-L-----GA-PL-K-F--P-*N--	672
H-1	A---A-----N---A--AASQ---A-----G-----A-----Q-QLE--FN-E-*T--	672

B.

LU3	MAGNAYSDEVLTGTTNWLKDKSNQEVFSFVFKNEDVQLNGKNIQWNSYRKEIQEEELKSLQRGAEITVDQSEDMEWESSVDELTKKFNOLTI**SEKYASQI	98	LARGE SPLICE												
MVM	-----A-----E-----N-----D-----K-----D-----P-----TT---H---GT---HDT-----I	100													
H-1	-----V-----S-----N-----D-----DD-----A---DM-----A---SD-----I	100													
LU3	PELRTNS**VGSRPGSGALEHTKYSCCGRCPNPEHVGWVFNLSRSSTEPNLVDRGGFESVLQFGTLEERLRTATKLGRLI	178	DONOR a												
MVM	---C---TCI---FR-----R-----S--Q--P-W-----RC--V---QR-AE---I	182													
H-1	-K--SY--TCI--PC--S-R-----SK-K---G--H--P-C-----L--SR-VG---QRGAD---I	182													
<table border="0" style="width: 100%;"> <tr> <td style="width: 33%;">DONOR a-ACCEPTOR a</td> <td style="width: 33%;">DONOR b-ACCEPTOR b</td> <td style="width: 33%;">DONOR a-ACCEPTOR b</td> </tr> <tr> <td>LU3 1. PEITVF 184</td> <td>2. YDGASG 184</td> <td>3. LGASRLQVPGTREQP 194</td> </tr> <tr> <td>MVM 1. ----- 188</td> <td>2. -----S 188</td> <td>3. ---W----- 197</td> </tr> <tr> <td>H-1 1. ----- 188</td> <td>2. -N-T-S 188</td> <td>3. ---W----- 197</td> </tr> </table>				DONOR a-ACCEPTOR a	DONOR b-ACCEPTOR b	DONOR a-ACCEPTOR b	LU3 1. PEITVF 184	2. YDGASG 184	3. LGASRLQVPGTREQP 194	MVM 1. ----- 188	2. -----S 188	3. ---W----- 197	H-1 1. ----- 188	2. -N-T-S 188	3. ---W----- 197
DONOR a-ACCEPTOR a	DONOR b-ACCEPTOR b	DONOR a-ACCEPTOR b													
LU3 1. PEITVF 184	2. YDGASG 184	3. LGASRLQVPGTREQP 194													
MVM 1. ----- 188	2. -----S 188	3. ---W----- 197													
H-1 1. ----- 188	2. -N-T-S 188	3. ---W----- 197													

Figure 15. Comparison of the translated left ORF coding for NS1 (A) and NS2 (B) among LuIII (LU3), MVMp and H-1: Dashes indicate amino acids identical to those of LuIII. Down arrows indicate characterized splice donor-acceptor junctions based on the transcription map of MVMp (15). The three possible carboxyl termini of NS2 protein resulting from alternative splicing are shown.

LU3	MAPPARAKKRWVFPGRKYLGPQNSLNOGEPINPSDAAAKLEDEATDOTIKSGINPYIITFSPADQRFIDQTKDAKLVGGKVGHYFFTKIKRAFAPKLSIDS	100
MVM	-----D-----A-----A-----	100
H-1	-----D-----	100
START OF VP2		
LU3	EPGTCVSTAGKRTKPPAHIFINQARAKKRTSLAAQORTQFSSDGTQDSGNAVQSAARVERAADGPGSGGGGGGGGGVGVSTGSDNOTHYKFLGD	200
MVM	---R---R---Y-----L-S---SS---S-P---H-----I-----R-----	200
H-1	---RP---V-----A-----L---ETNOPDTGI*AN---S-G-S---I---T-----	199
LU3	GWVCIATAYSTRMVHLMPKSENYCRVRYVNTNDGTASHMAMD***DAHEQIWTPLSLVDANAVGVWFOPSDWQYISNMHIHINLHSLDQELFNVIKTV	297
MVM	---LA---L-----I---T---SVKGN--R***-----L-----C-T-SQL--V---I---L---	297
H-1	---KAS-LL--G--P-----T---NOT--HGTKVKGMMAT-T-O-----F-Q-S-ESL--P--S-----V---	299
LU3	TE*ONTGAEAIKVTNNDLTAAHVALDSNNILPYPFAIDNQETLGFYVWKPTIPSPYRYTFSCDRNLSVITYKDEAGT**ITDTHGLASGLNSOFFTIENT	394
MVM	---DL-GQ---I---C---V-----ANSH-----A---CV-D-C---ENQE--VE*HNV--TPK-M-----	395
H-1	---Q-GA-OD---C-----AOTS-----A-A-----FHP-Q---SSNS-EGTQ---I-EPOA-----	399
LU3	ORINLLRTGDEYATGTYFDTEPIRLTHVOTNRKLGQPPQITELPSSDTANATLTARGYRSGLTQIQGRNDVTEATRVRPAQVGFQPHDNFETSRAGP	494
MVM	-Q-T---F-----NSVK-----O-----LLSTF-EA--DAG---Q-S-H-T--M*-V-W-S--I-T-----ND--A---	494
H-1	LP-T-----FT---I-N-D-LK-----ACLOG--D--T---T-S---N-D-F-S--T-NV-Y---L-T---I--M-----AN-G---	499
LU3	FKVPPVPADITQGLDEDANGSLRYTDTKONGQSWASQNMKDRITVDVNYDSGRVWNCFIQSVPTSEPNANQIILTRNDLAGKTDIHFTNAFNSYGP	593
MVM	-AA-K-----V-KE---V-S-G---EN---HGPAPF-----ETSFG---D-KDG---A-LVVP-PL-G---ANPIGT-N---S-V---	594
H-1	-----L-A-E-----A1-FN-G---ED--R-GAAPE-----IDSAA--D-AR--V--A-ISIP--Q---QRE-A1--H-NM-Y--V-----	599
LU3	LTAFPHPIIYPOQIWDKELDLEHKPRLHTQAPFVCRNNAFGOLLVRLAPNLTDOYDPNSSNLSRIVYTYGTFVYKGLTKLAKRPNATWNPVFOISAT	693
MVM	---S-S-V---I---IT-----H---G-----GAT-----MR--L-A-T---Y-V--E	694
H-1	---S---D---N---VT-----P---F-H-G---F---TTV-----S--Y---I-KF---L--L---Y-*AT-	698
LU3	NOGTNDYHSIERVLPATGNITNPVLLSRPVARNTY	729
MVM	DN--S---VTK-----MQS---IT-----	729
H-1	DSVA-S--NVKK---S---MMSD--IC---PRM--	734

Figure 16. Comparison of the translated right ORF coding for VP1 and VP2 among LuIII (LU3), MVMp and H-1: Dashes indicate amino acids identical to those of LuIII. Conserved amino acids among parvoviruses are enclosed by boxes.

DISCUSSION

We report the construction of an infectious, full-length clone of parvovirus LuIII and the complete nucleotide sequence of the genome. LuIII shares over 80% sequence identity with MVMp and H-1, suggesting that LuIII, although initially isolated from a human cell line (38), is a very close relative of the rodent parvoviruses. The transcription map of LuIII is probably similar to that of MVMp and H-1, since the sequence of the characterized regulatory regions and the splice donor-acceptor sites in these viruses are virtually identical. When translated, the LuIII ORFs coding for the nonstructural and structural proteins share a high percentage of sequence identity with those of MVMp and H-1. In LuIII, the carboxyl terminus of the capsid protein VP2 contained more substitutions than any other region of the virally coded proteins, as is also the case for MVMp and H-1. Although most of the amino acid substitutions were conservative, these changes could account for the lack of immunological cross-reactivity between LuIII and the rodent parvoviruses (38).

The encapsidation of equal amounts of plus and minus DNA strands by LuIII is unique among the parvoviruses because, unlike AAV and B19, it has non-identical left and right ends. Comparison of the terminal nucleotide sequences and conformations of LuIII to those of MVMp and H-1 did not reveal any major differences that could account for the different encapsidation patterns observed for these viruses; therefore some internal nucleotide sequences must also play a role in encapsidation.

Two regions of the LuIII sequence differ significantly from those of MVMp and H-1. One is located at map unit 92. There is a sequence present in tandem at the end of the right ORFs of MVMp and H-1, while only one copy of a similar sequence is present in LuIII. Since MVMi also has one copy and encapsidates primarily the minus strand, the presence of a single copy in LuIII is not expected to influence its encapsidation pattern. Salvino et al. (35) found that deletion of one copy of the direct repeat in MVMp resulted in 10 and 100-fold reduction in DNA replication in

A-9 and Cos-7 cells, respectively. Although the role these repeats play in replication of the virus is unknown, it has been suggested to function as an internal origin of replication. Rhode and Klaassen (31) isolated an H-1 mutant containing three copies of the repeat sequence. Different copy numbers of a similar repeat have been found in various isolates of canine (CPV), feline panleukopenia (FPV), raccoon (RPV), and mink enteritis (MEV) parvoviruses (28), but the effect of the direct repeats on encapsidation of these viruses was not studied.

The other region different from the rodent parvoviruses is at m.u. 89, six nucleotides downstream of the end of the right ORF. It is an A-T rich region of 47 nucleotides unique to the LuIII genome. Its location near the right terminus and its nucleotide sequence suggest that this A-T rich region could play an important role during LuIII replication. Its absence from MVMp and H-1 and presence in LuIII could be responsible for the characteristic virion DNA distributions of these viruses. In support of this hypothesis, transfection with a recombinant plasmid containing the left half of H-1 (m.u. 0 to the *HindIII* site) and the right half of LuIII (*HindIII* site to m.u. 100) resulted in the encapsidation of both plus and minus DNA strands (30). When a second recombinant molecule containing nts 1-270 (*NcoI* site) and the last 848 nts (from *NheI* site) of the LuIII genome bracketing a neomycin resistance gene, was transfected in the presence of either H-1, MVMp or LuIII virus, both plus and minus strands were encapsidated (30). This shows that a determinant of encapsidation is located within the the last 848 nucleotides and influences the final virion DNA distribution irrespective of the proteins provided in *trans*. The A-T rich stretch is the only sequence in this region of LuIII that differs significantly from the MVMp and H-1 sequences.

McLaughlin et al. showed that the *cis* signals required for packaging of the AAV genome reside in the terminal sequences, since molecules containing only the termini were encapsidated (24). Studies of MVMp and BPV (1,12) showed that the ratio of flip to flop conformations at the termini of replicative form DNA was similar to that of the virion DNA. If selective encapsidation is the only mechanism responsible for the observed patterns, then MVMp and H-1, like LuIII, would synthesize plus strand DNA, but not encapsidate it. Under this hypothesis, the left terminus of the minus strand is expected to occur in both flip and flop conformations and the flop conformation at the left end should be observed in the RF population. This sequence conformation was

not observed at the left end of the minus strand of MVMP. Chen et al. proposed that the ratios of the terminal conformations and strand polarity of viral DNA for an individual virus are the result of differential rates of hairpin transfer at the termini of viral replicative-form DNA (14). This differential rate of hairpin transfer is dependent on the rate constant associated with the conformation and location of the terminus. These studies show that the observed distribution of virion DNA for all characterized parvoviruses can be accounted for on the basis of characteristic rate constants without invoking selective encapsidation.

Although the differences in nucleotide sequence between LuIII and the rodent parvoviruses are minimal, they are not insignificant, for they determine the intrinsic properties of each virus which in turn results in the uniqueness of parvoviruses LuIII, H-1 and MVMP.

LITERATURE CITED

1. 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.
2. Astell, C. R., E. M. Gardiner, and P. Tattersall. 1986. DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain. *J. Virol.* 57:656-669.
3. Astell, C. R., M. Smith, M. B. Chow, and D. C. Ward. 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at the origins of DNA replication. *Cell* 17:691-703.
4. Astell, C. R., M. Thompson, M. Merchlinsky, and D. C. Ward. 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. *Nucleic Acids Res.* 11:999-1018.
5. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith and K. Struhl (Eds.). 1987. *Current protocols in molecular biology*. Vol I. p. 3.11.3. Green Publishing Associates and Wiley Interscience. New York.
6. 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.
7. Bates, R. C., C. E. Snyder, P. T. Bannerjee, and S. Mitra. 1984. Autonomous parvovirus LuIII encapsidates equal amounts of plus and minus DNA strands. *J. Virol.* 49:319-324.
8. Bates, R. C., J. Storz, and D. E. Reed. 1972. Isolation and comparison of bovine parvoviruses. *J. Infect. Dis.* 126:531-536.
9. Berns, K. I., and R. A. Bohenzky. 1987. Adeno-associated viruses: an update. *Adv. Virus Res.* 32:243-306.
10. 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.
11. Bohenzky, R. A., R. B. LeFebvre, and K. I. Berns. 1988. Sequence and symmetry requirements within the internal palindromic sequences of the Adeno-associated virus terminal repeat. *Virology* 166:316-327.
12. Chen, K. C., B. C. Shull, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Analysis of the termini of the DNA of bovine parvovirus: demonstration of sequence inversion at the left terminus and its implication on the replication model. *J. Virol.* 62:3807-3813.

13. Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates. 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. *J. Virol.* 60:1085-1097.
14. Chen, K. C., J. J. Tyson, M. Lederman, E. R. Stout, and R. C. Bates. A kinetic hairpin transfer model for parvoviral DNA replication. 1989. *J. Mol. Biol.* 208:283-296.
15. Cotmore, S. F.. 1990. Gene expression in the autonomous parvoviruses, p. 141-154. in (Tijssen, P., Ed.) *Handbook of parvoviruses*. Vol I. CRC Press, Florida.
16. Cotmore, S. F., and P. Tattersall. 1987. The autonomously replicating parvovirus of vertebrates. *Adv. Virus Res.* 33:91-174.
17. Devereux, J., P. Haeberli and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12:387-395.
18. Diffoot, N., B. C. Shull, K. C. Chen, E. R. Stout, M. Lederman, and R. C. Bates. 1989. Identical ends are not required for the equal encapsidation of plus- and minus-strand parvovirus LuIII DNA. *J. Virol.* 63:3180-3184.
19. Faust, E. A., and D. C. Ward. 1979. Incomplete genomes of the parvovirus minute virus of mice: selective conservation of genome termini, including the origin for DNA replication. *J. Virol.* 32:276-292.
20. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166: 557-580.
21. Jongeneel, C. V., R. Sahli, G. K. McMaster, and B. Hirt. 1986. A precise map of splice junctions in the mRNAs of minute virus of mice, an autonomous parvovirus. *J. Virol.* 59:564-573.
22. Lopata, M. A., D. W. Cleveland, and B. Sollner-Webb. 1984. High level transient expression of a chloramphenicol acetyl transferase gene by DEAE-dextran mediated DNA transfection coupled with a dimethyl sulfoxide or glycerol shock treatment. *Nucleic Acids Res.* 12:5707-5717.
23. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka. 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. *J. Virol.* 62:1963-1973.
25. Morgan, W. R. and D. C. Ward. 1986. Three splicing patterns are used to excise the small intron common to all minute virus of mice RNAs. *J. Virol.* 60:1170-1174.
26. Muzyczka, N., R. J. Samulski, P. Hermonat, A. Srivastava and K. I. Berns. 1984. The genetics of adeno-associated virus. *Adv. Exp. Med. Biol.* 179:151-161.
27. 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.
28. Parrish, C. R., C. F. Aquadro, and L. E. Carmichael. 1988. Canine host range and a specific epitope map along with variant sequences in the capsid protein gene of canine parvovirus and related feline, mink, and raccoon parvoviruses. *Virology* 166:293-307.

29. 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.
30. Rhode, S. L. (University of Nebraska Medical Center). 1991. Personal communication.
31. Rhode, S. L., and B. Klaassen. 1982. DNA sequence of the 5' terminus containing the replication origin of parvovirus replicative form DNA. *J. Virol.* 41:990-999.
32. Rhode, S. L., and P. R. Paradiso. 1983. Nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. *J. Virol.* 45:173- 184.
33. Rhode, S. L., and S. M. Richard, 1987, Characterization of the *trans*-activation-responsive element of the parvovirus H-1 P38 promoter. *J. Virol.* 61:2807-2815.
34. Rodriguez, R. L., and R. C. Tait. 1983. Recombinant DNA techniques: an introduction. The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA.
35. Salvino, R., M. Skiadopoulus, E. A. Faust, P. Tam, R. O. Shade, and C. R. Astell. 1991. Two spatially distinct genetic elements constitute a bipartite DNA replication origin in the minute virus of mice genome. *J. Virol.* 65:1352-1363.
36. Salzman, L. A., and P. Fabisch. 1979. Nucleotide sequence of the self-priming 3' terminus of the single-stranded DNA extracted from the parvovirus kilham rat virus. *J. Virol.* 30:946-950.
37. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
38. Seigl, G., 1976. The Parvoviruses, Virology Monographs. Vol 15. Springer-Verlag, New York. p. 72-94.
39. Shade, R. O., B. 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.
40. Shein, J. M., and J. F. Enders. 1962. Multiplication and cytopathogenicity of simian vacuolating virus 40 cultures of human tissues. *Proc. Soc. Exp. Biol. Med.* 109:495-500.
41. Shull, B. C., K. C. Chen, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Genomic clones of bovine parvovirus: construction and effect of deletions and terminal sequence inversions on infectivity. *J. Virol.* 62:417-426.
42. 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.
43. Summers, J., S. E. Jones, and M. J. Anderson. 1983. Characterization of the genome of the agent of erythrocyte aplasia permits its classification as a human parvovirus. *J. Virol.* 64:2527-2532.

Chapter IV

**TRANSCRIPTION MAP OF BPV AS
GENERATED BY AMPLIFICATION OF cDNA
ENDS**

INTRODUCTION

The DNA genomes of the mammalian parvoviruses are organized into two major coding regions, the left open reading frame (ORF), which codes for the non-structural proteins and the right ORF for the structural or capsid proteins (18). The genomes of bovine parvovirus (BPV) and a human parvovirus, RA-1, are unique in that they have a third ORF in a different reading frame in the middle of the genome (the mid ORF). The BPV mid ORF, with a coding capacity of 255

amino acids, is thought to encode the major part of the BPV non-structural protein NP-1 (5). In all mammalian parvoviruses for which the sequence is known, the protein coding regions appear to be restricted to the plus strand. This is in contrast to the recent finding that the densoviruses *Junonia coenia* (*JcDNV*) and *Galleria mellonella* densovirus (*GmDNV*) contain ORFs in both strands (1, 23).

RNAs coding for the non-structural and structural proteins of parvoviruses are each transcribed as overlapping transcription units initiating at a single promoter (B19) or multiple promoters (MVM, H-1, AAV) in the viral genome. The non-structural and structural proteins of the rodent parvoviruses (MVM, H-1) are transcribed from promoters at map units 4 and 38, respectively. The dependent parvovirus, AAV, has three promoters. Transcripts for the AAV non-structural proteins initiate at promoter sequences located at map units 5 and 19, and transcripts for the structural proteins initiate from the promoter at map unit 40. The nucleotide sequence of the human parvovirus B19, revealed five potential promoters at map units 6, 23, 42, 43, and 55 (21), yet all B19 transcripts were shown to initiate from promoter sequences at map unit 6. Promoter activity at map unit 44 in the B19 genome, potentially used for transcription of capsid proteins, was identified when these sequences were cloned upstream of a chloramphenicol acetyltransferase (CAT) gene and CAT expression was detected after transfection into HeLa cells (8). However, other studies showed that this level of expression of another reporter gene expressed from p44 did not exceed that of a promoter-less control (14). Previous mapping of BPV transcripts suggests that, like B19, all transcripts may be initiated at a single promoter located at map unit 14 (3), although the nucleotide sequence revealed potential promoter sequences at map units 4.5, 12.8 and 38.7 of the viral genome (5).

BPV transcription *in vivo* produced four size classes of polyadenylated RNA, the 5.25 kb, 3.6 - 3.1 kb, 2.6 - 2.25 kb, and 1.45 - 0.8 kb species. Mapping by S1 nuclease digestion and two dimensional neutral and alkaline agarose gel electrophoresis showed that the members of each size class contain a common main body to which smaller segments are spliced. By *in vitro* translation of size-fractionated cytoplasmic RNA from BPV-infected BFL cells, the most abundant BPV RNA, the 2.6 kb family, was shown to encode all three BPV capsid proteins (3). Transcription for

this RNA family was shown to start at map unit 14 and continue to map unit 94. Leader RNA segments, 0.35 kb or smaller, from map positions 14 through 20, were joined to a 2.25 kb main body from map positions 53.5 to 94. Three 5' donor sites (map positions 18, 19 and 20) contained in the leader segments, would join a single acceptor site at map position 53.5 contained in the main body of the transcript. The transcripts for the 2.6 kb BPV RNA family apparently also produces the 1.45 - 0.8 kb RNAs. These would be contained within the large intron of the 2.6 kb RNAs, initiating transcription at map position 13.5. This suggests that these RNAs arise from multiple splicing patterns of a single primary transcript. The 5.25 kb and 3.6 - 3.1 kb RNAs are less abundant. The 5.25 kb RNA species consists of a 3.1 kb and 2.25 kb RNA segments while 3.6 - 3.1 kb family contains a 3.1 kb main body. A small amount of unspliced 3.6 kb RNA was observed.

These data on the BPV transcripts was obtained before the nucleotide sequence of BPV was known. The sizes and map positions of the transcripts were estimates based on size markers and a limited restriction map. Since then the nucleotide sequence of BPV has been determined (5). Potential promoters are located at map units 4, 12 and 38. With this new information we set out to construct cDNAs and characterize BPV transcripts. Due to the low yield of BPV RNAs from BPV-infected BFL cell lysates, construction of cDNAs by conventional methods (15) proved to be a difficult task. However, recent findings on the initiation site of BPV transcripts permitted the amplification of BPV low abundance mRNAs by the polymerase chain reaction (PCR) (10). Using primers specific to each ORF, BPV cDNA ends were amplified. We used the sizes and genome location of the amplified fragments, in conjunction with the known sizes of the BPV-coded proteins and transcripts, to refine the BPV transcription map obtained from earlier studies.

MATERIALS AND METHODS

Virus propagation and cell culture

Bovine fetal lung cells (BFL) were grown in monolayer culture and maintained in Eagle's Minimal Essential Medium (MEM) supplemented with 10% fetal calf serum as described earlier (17). RNA was obtained from BFL cells infected at 10 pfu/cell and harvested 24-48 hours post-infection, when more than 50 % CPE was observed. Cells were collected by centrifugation and washed three times with cold Tamm's phosphate-buffered saline (15). The resulting cell pellet was either frozen in liquid nitrogen and stored at -80° C or processed immediately.

RNA Isolation

Total RNA isolation using guanidium isothiocyanate

Two methods employing guanidium isothiocyanate were used to isolate RNA from BPV-infected BFL cell lysates. The first method followed the protocol described by Burd (3) with minor modifications. BPV-infected cells from 1 roller bottle (about 1×10^7 cells) were resuspended in 1 ml of a solution containing guanidium isothiocyanate (5 M guanidium isothiocyanate, 50 mM Tris-Cl, pH 8.0, 25 mM EDTA, pH 8.0, 0.1 M β -mercaptoethanol, 0.5% N-lauroyl sarcosine) and overlaid on a 4 ml cushion of 5.7 M CsCl-0.1 M EDTA, pH 7.5. Gradients were centrifuged at 35,000 rpm in the SW41 rotor for 21 hrs at 15° C. Following centrifugation, the RNA pellets were immediately resuspended in H₂O and repeatedly extracted with chloroform-isoamyl alcohol (24:1). The RNA was then ethanol precipitated, resuspended in 50 μ l H₂O (when starting with 8×10^7 cells) and stored at -80°C.

Total RNA was also prepared essentially as described by Chomczynski and Sacchi (6) with some modifications. For this procedure, cells were grown and infected in 100 mm culture dishes. Each dish contained about 10^7 cells. BPV-infected cells contained in a single dish were resuspended in 0.5 ml of a guanidium isothiocyanate solution (4 M guanidium isothiocyanate, 25 mM sodium citrate, 0.5% Sarkosyl, 0.1 M β -mercaptoethanol). The resulting lysate was brought to a concentration of 2 M sodium acetate and extracted with an equal volume of 5 phenol/1 chloroform/0.02 isoamyl alcohol. RNA present in the aqueous phase was precipitated with isopropanol and resuspended in one-tenth the original volume of guanidium isothiocyanate solution and reprecipitated with isopropanol. The RNA was resuspended in 50 μ l H₂O (when starting with 8×10^7 cells) and stored at -80°C.

Isolation of poly-A tailed mRNA from BPV-infected cells

Poly-A mRNA was isolated as described by Bradley et al. (2) with slight modifications. 10^7 BPV-infected BFL cells were resuspended in 1 ml lysis buffer (0.2 M NaCl, 0.2 M Tris-Cl, pH 7.5, 1.5 mM MgCl₂, 2% SDS, 200 μ g/ml proteinase K). After shaking for a minimum of two hours at 45°C, 0.2 gm of oligo dT-cellulose (Collaborative Research Laboratories) equilibrated in 0.5 M NaCl-0.1 M Tris were added to the lysate. The lysate was shaken at room temperature for 30 min. The oligo dT-cellulose was collected by a low speed centrifugation (not exceeding 500 x g), resuspended in 0.5 ml of binding buffer (0.5 M NaCl-0.1 M Tris pH 7.5) and transferred to a 1.5 ml microcentrifuge tube. The cellulose was washed repeatedly with 0.5 ml of binding buffer until the A₂₆₀ of the eluate was less than 0.5. The poly-A mRNA was eluted from the oligo-dT cellulose with 0.5 ml H₂O and precipitated repeatedly with ethanol. The RNA was resuspended in 50 μ l H₂O (when starting with 8×10^7 cells) and stored at -80°C.

Northern Analysis of BPV RNAs

RNA samples were electrophoresed as described by Charmichael and McMaster (4). RNA blotting onto nitrocellulose (BioRad Laboratories) was performed as described by Thomas (22). Electrophoresis was performed using 1% agarose prepared in 6% formaldehyde, 9 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, pH 7.0. The running buffer was 2.2 M formaldehyde, 9 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, pH 7.0. Electrophoresis was performed for 4 hrs at 4 V/cm with buffer recirculation. Size determinations were based on the mobility of an RNA ladder (Bethesda Research Laboratories) with a size range of 0.3 to 9.5 Kb.

Following electrophoresis, lanes in the gel containing the RNA ladder were sliced from the gel and stained with ethidium bromide (Sigma Chemicals) for visualization under ultraviolet light. The remaining gel was soaked in 50 mM NaOH - 10 mM NaCl for 45 min, neutralized in 0.1 M Tris-Cl, pH 7.5 for 45 mins and then soaked in 20 x SSC for one hour prior to transfer onto nitrocellulose by wicking with 10 x SSC.

Prehybridization and hybridization were performed in 50% formamide, 5 x SSC, 50 mM Tris-Cl, pH 7.5, 250 $\mu\text{g}/\text{ml}$ salmon sperm DNA and 5 x Denhart's solution (15) for 2 hrs at 37° C. The probe consisted of 2×10^5 cpm/ml of ^{32}P -BPV DNA labeled by the random primer labeling method (Random primer labeling kit, Boehringer Mannheim Biochemicals) using ^{32}P - α -dATP (New England Nuclear, 3,000 Ci/mM). Hybridizations were done at 37° C for 18 hrs. Following hybridization, the membrane was washed twice in 2 x SSC-0.2% SDS and twice in 0.2 x SSC. All washes were done at 37° C for 30 mins.

Reverse Transcription of BPV RNA

A cDNA synthesis system kit (Bethesda Research Laboratories) using Moloney murine leukemia virus reverse transcriptase (MMLV) was used to reverse transcribe the first strand of BPV

total or BPV poly A-mRNA according to manufacturer's instructions. The protocol suggests using 10 μg of template RNA. Since the yield of RNA from 8×10^7 BPV-infected cells was less than this amount, 90 % of the recovered RNA was used for cDNA synthesis. The final products were re-suspended in 50 μl of H_2O .

Amplification of BPV transcripts by the Polymerase Chain Reaction

The RACE (Rapid Amplification of cDNA Ends) protocol described by Frohman et al. (10) was used, with modifications, to amplify BPV-specific transcripts using the polymerase chain reaction (PCR) (Fig. 17). Template for amplification was 1 μl of the cDNA product obtained as described above. A BPV-specific synthetic primer, P4 (5' CTTGTCGACAGAGAACACACGTCTCTGCGCT 3'), complementary to nucleotides 296 through 322 of the BPV minus strand, in conjunction with internal minus strand BPV-specific primers (Figs. 18, 19, 20) for all three open reading frames, was used to amplify the 5' end of BPV transcripts. A second synthetic primer, T17, (5' ACTAGTCGACATCGATATCTTTTTTTTTTTTTTTTTTTT 3'), complementary to the poly A tail of mRNA, in conjunction with internal plus strand BPV-specific primers to the three open reading frames, was used for initial amplification of the 3' end of BPV transcripts (Figs. 18, 19, 20). Due to the instability of the T₁₇ primer, after the first amplification of the 3' end of the mRNAs, a primer consisting of only the restriction sites included in the T₁₇ primer (5' ACTAGTCGACATCGATATC 3') was used for subsequent amplifications. All primers used had restriction endonuclease recognition sites at their 5' ends, unique to or absent from the BPV genome (Fig. 18). Reaction conditions were those suggested by the manufacturer (Perkin Elmer Cetus) for basic PCR, with the addition of 0.1 unit/ μl of *Escherichia coli* single-stranded DNA binding protein (United States Biochemicals). Five pmoles of T₁₇ and 300 pmoles of all the other primers were used in the reactions. The parameters of the reaction cycle were one cycle of 94° C for 5 min, 55° C for 20 min, 72° C for 40 min followed by 40 cycles of 94° C for 1 min, 65° C for

2 min, 72° C for 3 min, and a final extension of 72° C for 15 min. 10% of the DNA obtained from these amplifications were electrophoresed on a 1% Nusieve GTG, 1% Seakem GTG (FMC Corp.) agarose gel buffered with TAE (40 mM Tris, 20 mM sodium acetate, 2 mM EDTA). The DNA bands were cut from the gel and melted by boiling for 5-10 min. One milliliter of boiling water was added to the molten gel, diluting the template and the gel so that repolymerization was impossible. Approximately 1 pg of template in conjunction with the appropriate primers were used for a second round of amplification.

Cloning and sequencing of PCR-generated cDNA fragments.

Amplified BPV cDNA fragments were extracted with phenol/chloroform/isoamyl alcohol (1/1/0.04) and passed through a P30 column (Biorad Laboratories) to remove residual primer. After isopropanol precipitation, the DNA was digested with the appropriate restriction endonucleases (Bethesda Research Laboratories, Boehringer Mannheim Biochemicals) and ligated into similarly digested pUC19 vectors. Ligation mixtures were transformed into *Escherichia coli* JM107 or DH5 α cells by the Hanahan procedure (11). DNA was isolated by alkaline lysis (19). The Sequenase kit (United States Biochemical) was used to sequence the recombinant plasmids as instructed by the manufacturer.

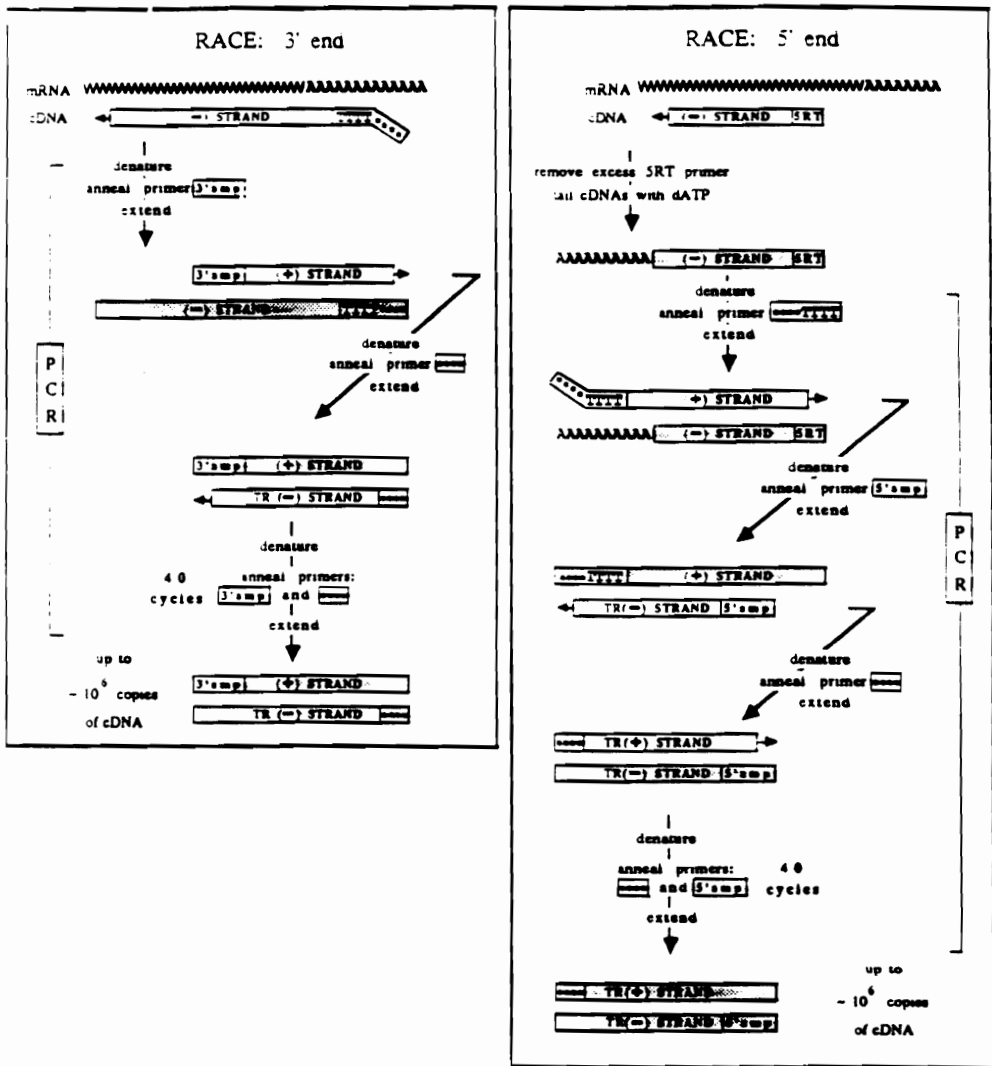


Figure 17. Schematic representation of the RACE protocol: Diagram illustrating the exponential amplification of the 3' and 5' ends of cDNAs. Abbreviations: SRT, primer used for initial amplification of the 5' end; TR, Truncated; 5'amp, 3'amp, primers used for the amplification of cDNA ends; PCR, polymerase chain reaction. (Reprinted from Frohman et al., 1988.)

BPV	P4	T17* & Adapter**
ORFs	5' CTTCGACAGAGAACACAGCCTCTCTGGCT 3' 1	* 5' ACTAGTCGACATCGATAAICTTTTTTTTTTTTTTT 3' 2 3 4 5 ** 5' ACTAGTCGACATCGATATC 3'
Left	5' TCTCAGCTACCGATGGACCTGAT 3' 6 5' ACGCGAGCTTTTGGTAGCTGGAGTGA 7	5' ATGGTACCTCAGAGCCGA 3' 6 5' CCAAAAGAGCTCGCTGGAGGAAGGT 3' 7
Mid	5' TCTATACCAGTCTCTAAACATTT 3' 5' CACCTGGCTTTATAGAGATGGCCG 3'	5' ATCATTTCAGTCAAAAAGGATTGAC 3' 5' CAATTGAACCTGGCAGCTGGTGGCT 3'
Right		

Figure 18. Sequence of primers used for the amplification of BPV cDNAs ends by the RACE protocol: Sequence of 5' amplification primers are BPV plus strand sequence and the sequence of all 3' amplification primers are BPV minus strand sequence. The numbers represent restriction endonuclease recognition sites: 1, Sall; 2, SpeI; 3, ClaI; 4, Sall; 5, EcoRV; 6, KpnI; 7, SstI.

PRIMER	POLARITY	SIZE	LOCATION IN BPV GENOME
P4	+	31mer	296 - 322
5'LT	-	23mer	2392 - 2414
3'LT	+	21mer	2401 - 2421
5'MID-I	-	30mer	2611 - 2640
3'MID-I	+	26mer	2625 - 2651
5'MID-II	-	25mer	3111 - 3134
3'MID-II	+	25mer	3038 - 3062
5'RT	-	25mer	3912 - 3936
3'RT	+	25mer	3743 - 3767

Figure 19. Location in the BPV genome of specific primers used in the RACE protocol: Nucleotide numbers are for the plus strand of BPV irrespective of the polarity of the primer. +, plus strand BPV sequence; -, minus strand BPV sequence. *

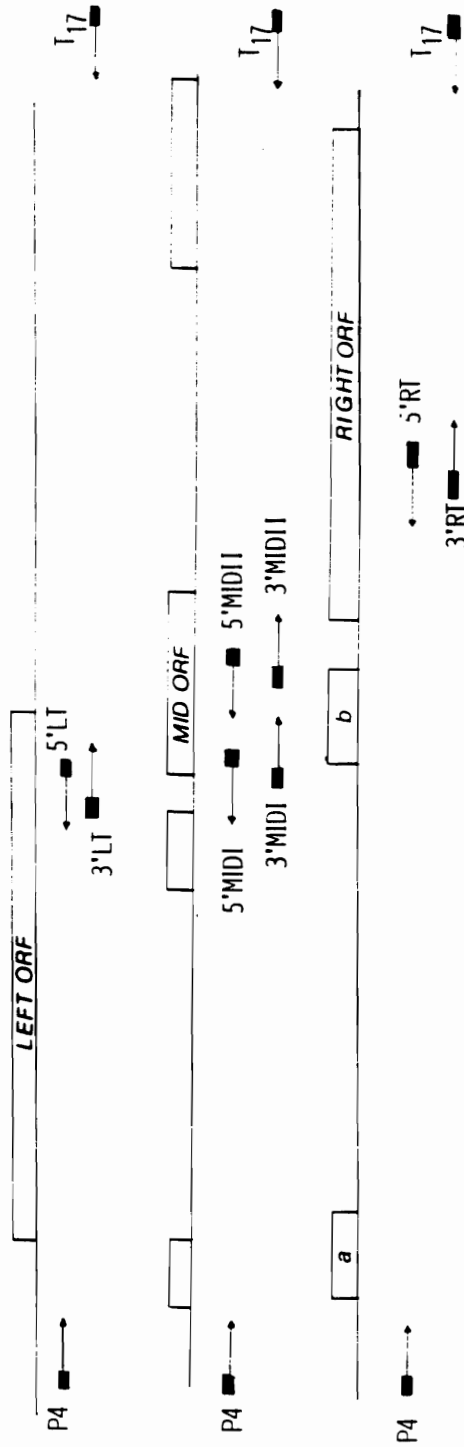


Figure 20. Strategy used for the amplification of BPV cDNAs: RNA isolated from BPV infected cells was converted to cDNA by MIMLV reverse transcriptase. This cDNA template was then amplified by PCR. The 5' ends of BPV cDNAs were amplified with P4 and one of the three possible 5'amp primers specific for each ORF. The 3' ends were amplified with T₁₇ and one of the three possible 3'amp primers specific for each ORF. Arrows indicate the direction of synthesis during amplification.

RESULTS

Nature of BPV transcripts

Northern blot analysis of BPV poly A-mRNA transferred to nitrocellulose and probed with ³²P-BPV DNA showed RNA species of 7, 5.8, 5.1, 4.2, 3.8, 3.0, 2.4, 1.8 and 1.6 kb (Fig. 21). RNA species in the range of 2.3 - 4.2 appear as doublets. BPV RNA yields were greatest at approximately 30 hours post-infection (Fig. 22), with the 2.4 Kb species being the most abundant. The apparent decrease in viral RNA at 30 hours post-infection is presumably due to cell lysis. Alkaline gel analysis and DNase I treatment demonstrated that the 5 kb and larger RNA species likely represent contaminating DNA in the BPV RNA (Fig. 23).

All BPV transcripts originate from P4

Northern blots containing identical RNA samples were probed with different ³²P-BPV probes labeled by random priming. One probe consisted of full length BPV DNA and the other of nucleotides 1 through 352 (NheI fragment) including promoter-like sequences localized at map unit 4 (P4) of BPV. Both probes hybridized with the same BPV RNA species (Fig. 24) suggesting that most if not all BPV transcripts initiate from promoter sequences localized at map unit 4.

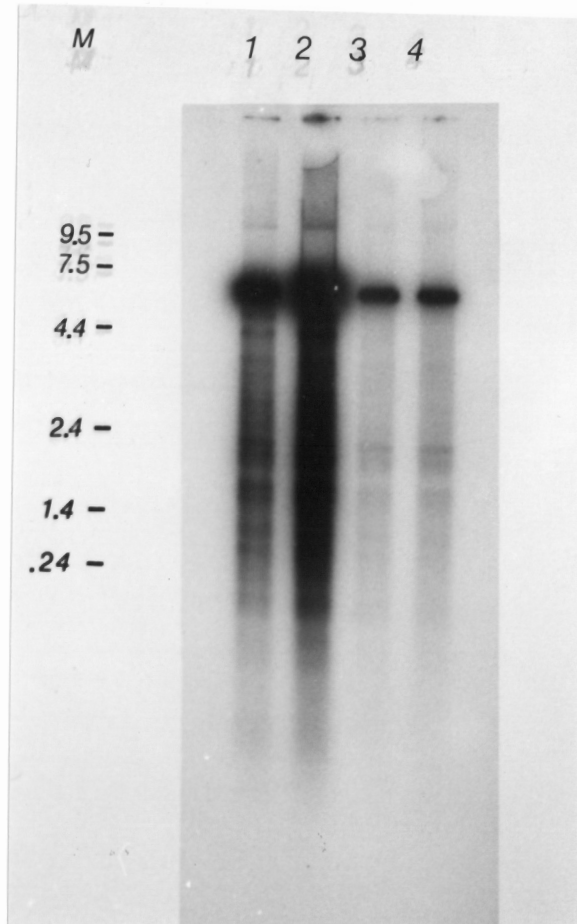


Figure 21. Northern blot analysis of BPV RNA: Poly-A mRNA from BPV-infected BFL cells was hybridized to ^{32}P -BPV DNA. Lanes 1 and 2 contain RNA from a preparation different from than in lanes 3 and 4. Lanes 2 and 4 contain twice the amount of RNA in lanes 1 and 3, respectively. Sizes of the RNAs in the RNA size ladder (M) are given in kilobases.



Figure 22. Temporal appearance of BPV RNA by Northern blot analysis of total RNA isolated from BPV-infected BFL cells: Total RNA was isolated from BPV-infected BFL cells at 6, 11, 20, 30 and 44 hours postinfection and probed with ^{32}P -labeled full-length BPV DNA.

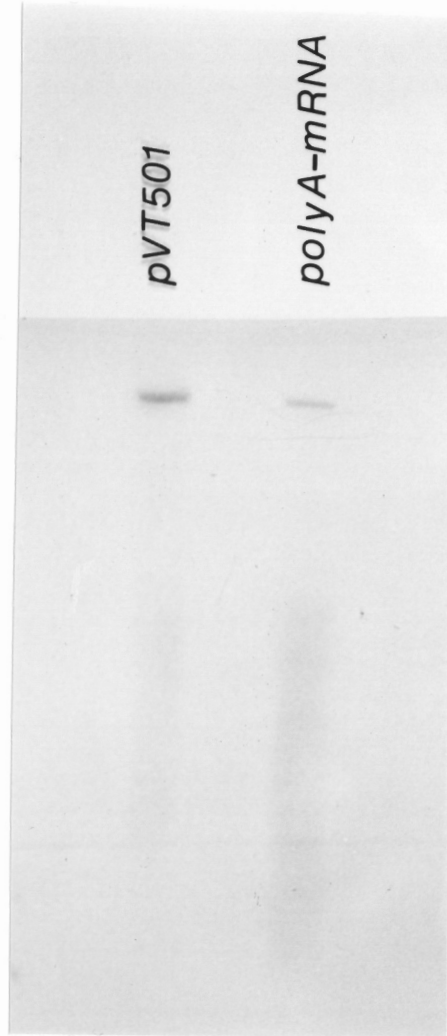


Figure 23. Alkaline gel analysis of poly-A RNA isolated from BPV-infected BFL cells: RNA isolated from BPV-infected BFL cells was electrophoresed on a 1% agarose gel prepared in 50 mM NaCl, 1mM EDTA and electrophoresed with 30 mM NaOH, 1mM EDTA. The RNA was transferred to a Zeta-Probe membrane and hybridized with ³²P-labeled full-length BPV DNA. The Sall fragment (the complete BPV genome) of pVT501 was run as a size marker.

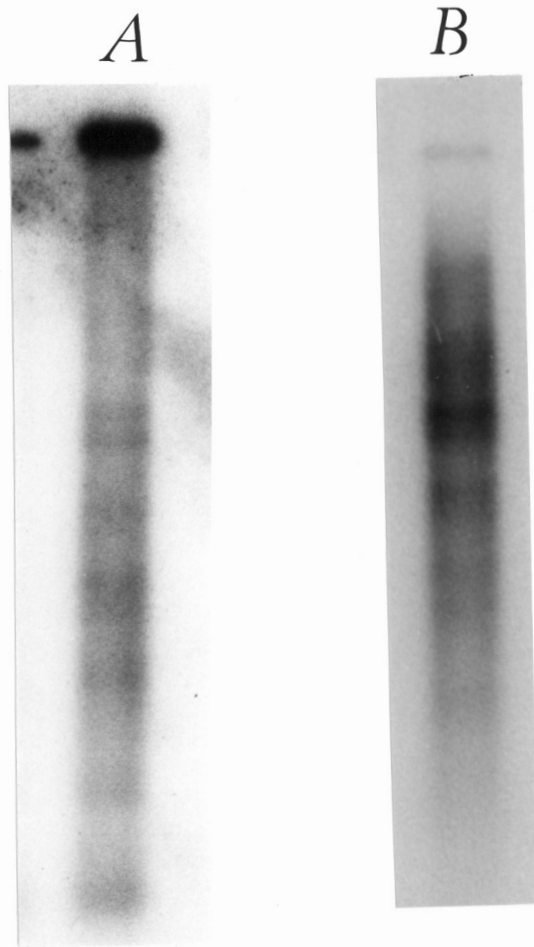


Figure 24. Northern blot analysis of BPV RNA using BPV-specific probes: Lanes A and B contain identical poly-A RNA preparation. A. Northern probed with ^{32}P -labeled BPV *Nhe*I fragment containing, BPV nucleotide sequences 1-352. B. Northern probed with full-length ^{32}P -labeled BPV DNA.

Analysis of cDNA fragments

Amplification of BPV-specific cDNA ends from total and poly-A mRNA were obtained by the polymerase chain reaction (PCR). The 5' and 3' ends of each transcript were amplified separately. Each amplification resulted in a number of DNA fragments (Fig. 25) which hybridized to ³⁵S-labeled BPV DNA (Fig. 26, 27). The sizes assigned to the amplified fragments represent estimates based on the mobility of a 1 kb DNA ladder. Amplifications with P4 and 5'LT resulted in two fragments of 1995 and 251 base pairs (bp). Each fragment was cloned into the KpnI-SalI site of pUC19. Sequence analysis showed that these fragments were actually 2119 and 240 bp in length (Fig 28, clones 1 and 2). Clone 1 contained uninterrupted BPV nucleotide sequences from nt 296 to 2409. Clone 2 contained BPV sequences from nt 296 to 325 and from nt 2198 to nt 2409 (splice A, Fig. 28). When translated using the Pustell sequence program, the nucleotide sequence of both clones was open in reading frames 2 (left ORF) and 3 (mid ORF). Amplifications with 3'LT and T₁₇ resulted in four fragments of 631, 468, 390 and 355 bp.

Amplification with P4 and 5'MID-I resulted in five fragments of 2239, 708, 501, 398, and 224 bp. Two fragments were cloned into the SstI-SalI site of pUC19 (Fig. 28, clones 3 and 4). The insert in clone 3 is 467 bp in length, and contains a splice identical to that observed in clone 2. The fourth clone contains a 270 bp insert, and in addition to the large splice found in clones 2 and 3, it contains a smaller splice that spans nt 2337 to nt 2534 (splice B) (Fig. 28) downstream of the large splice. If translated, the nucleotide sequence contained in these two clones was open in reading frames 2 (left ORF) and 3 (mid ORF). Both splice A and B have donor and acceptor sequences that agree with suggested consensus sequences (Fig. 29) (16). Amplification with 3'MID-I and T₁₇ resulted in three fragments with sizes of 501, 447 and 281 nts.

To overcome the ambiguity due to the annealing of the MID-I primers to both the left and mid ORFs, primers with sequence unique to the mid ORF were synthesized (5' and 3'MIDII). The 5' primer annealed to nt 3111 and the 3' primer annealed to nt 3038. Amplification with P4 and

5'MIDII yielded fragments of 344 and 700 bp, where as amplification with T₁₇ yielded a fragment of 344 (Fig. 25).

Amplifications with P4 and 5'RT resulted in fragments of 955, 708 and 562 nts and amplification with 3'RT and T₁₇ in fragments of 1259, 1122, 708, 427 and 339 nts.

BPV ORFs	P4 5' END cDNAs		T ₁₇ 3' END cDNAs	
	<i>Left</i>	5'-LT	1995, 251	3'-LT
<i>Mid</i>	5'-MID-I	2239, 708, 501, 398, 224	3'-MID-I	501, 447, 281
	5'-MID-II	700, 344	3'-MID-II	344
<i>Right</i>	5'-RT	955, 708, 562	3'-RT	1259, 1122, 708, 427, 339

Figure 25. DNA fragments generated by PCR from amplifications of BPV cDNA ends: Sizes of the fragments are given in basepairs. Abbreviations: LT, left; RT, right.

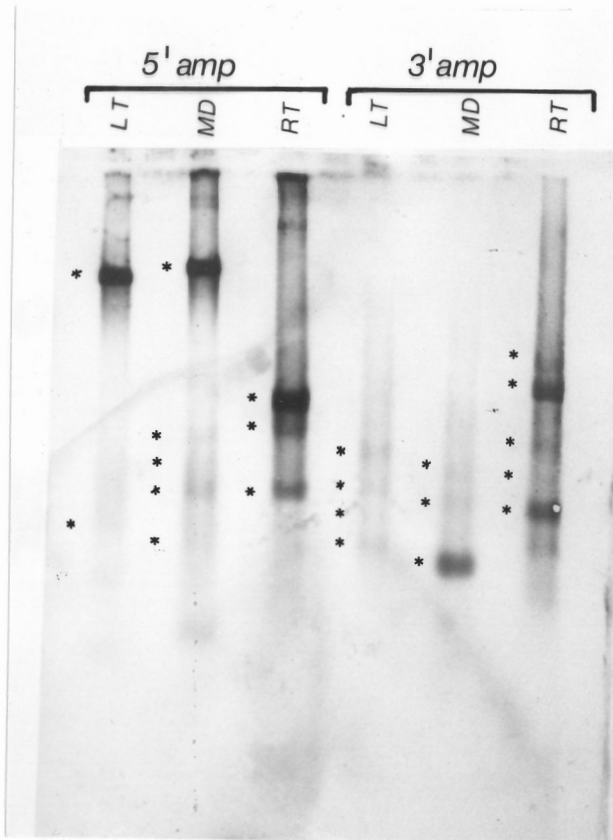


Figure 26. Southern blot analysis of BPV cDNA fragments generated by PCR: PCR products from amplifications of BPV cDNA ends were run on a 2% agarose gel, transferred to a Zeta-Probe membrane and hybridized with full length 32 -P BPV DNA. Abbreviations: LT, left; MD, Mid; RT, Right.

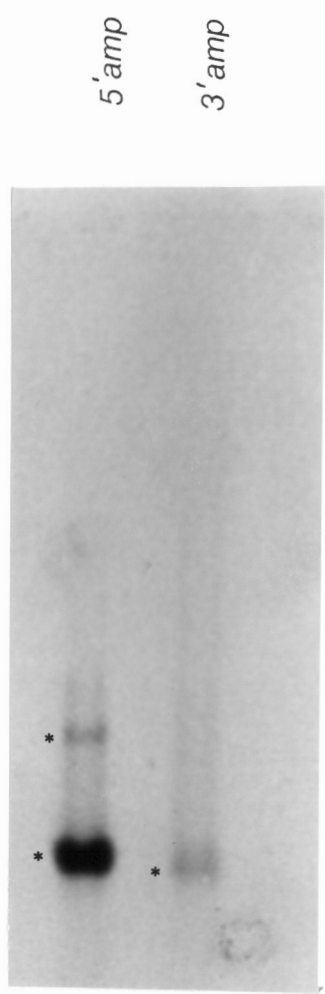


Figure 27. Southern blot analysis of BPV cDNA fragments generated with MID-II primers by PCR: PCR products from amplifications of BPV cDNA ends were electrophoresed on a 2% agarose gel, transferred to a Zeta-Probe membrane and hybridized with full length ³²-P BPV DNA. Abbreviation: MD, Mid.

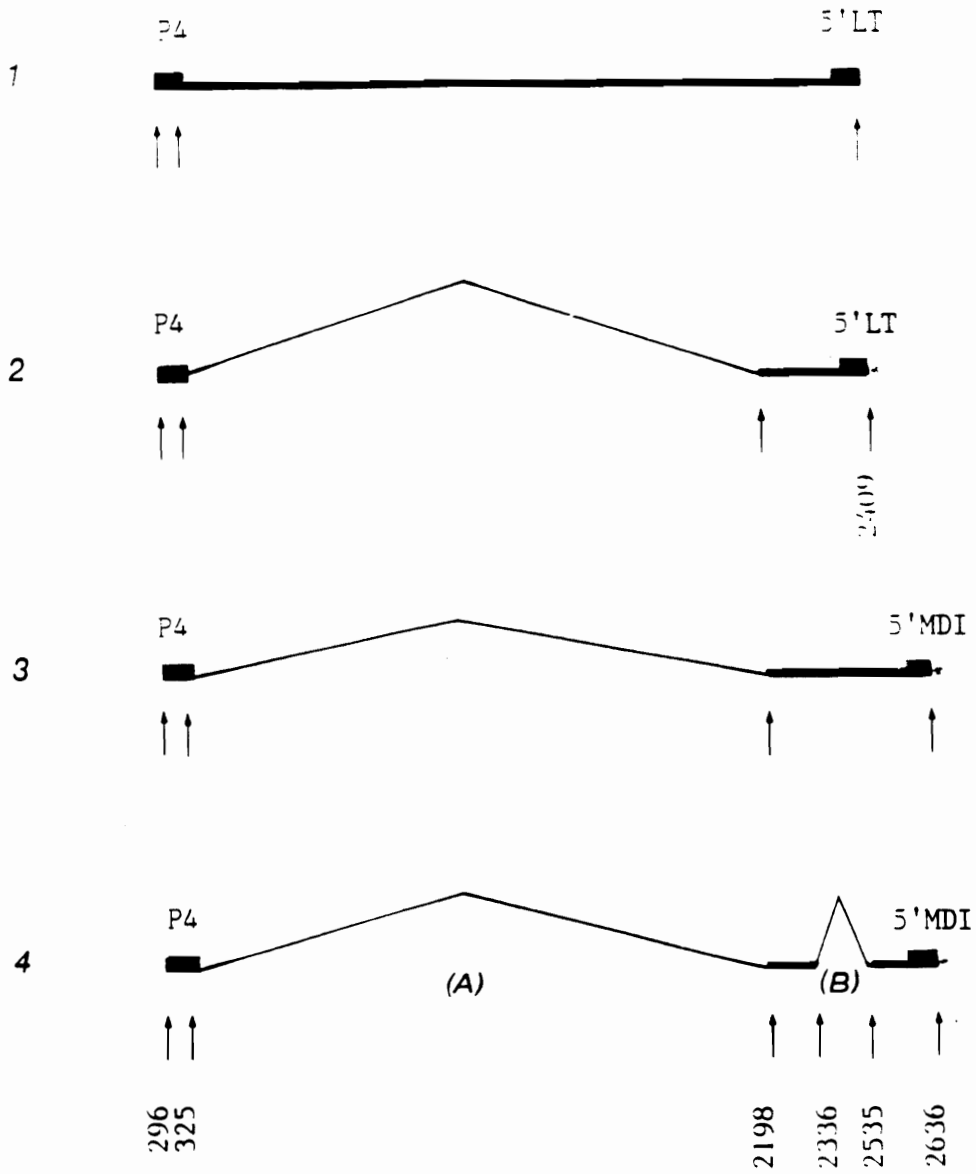


Figure 28. cDNA recombinant clones generated by the RACE protocol: Small black boxes represent primers used to generate the amplified fragment. Heavy lines represent sequences present in the clones while thin lines represent sequences present in the genome but absent in the clones. (A) and (B) represent putative splices. The nucleotide numbers of the cloned sequences are given.

SPICE	DONOR SITE	ACCEPTOR SITE	ABSENCE OF AG @ -15 TO -5 OF ACCEPTOR SITE
A	C A G G T A G T (8) (9) (10) (11) (12) (13) (14) (15)	(T) _n N A G G (8) (9) (10) (11) (12) (13) (14) (15)	YES
B	(316) C T C T G A G T G G G C C A C A G A (2198)	YES
	(2334) A C A G T G A G C G G G A C T A G A T C (2537)	YES

Figure 29. Sequence of splice junctions illustrated in Fig. 28: Splice donor-acceptor consensus sequences (15) are given at the top of the figure. Numbers given for consensus sequences represent the probability of having the respective nucleotide at the specific location. The sequences of splices A and B (see Fig. 28) are given and the locations of the splices in the BPV genome are indicated by the numbers in parenthesis. All nucleotide sequences are plus strand BPV sequence.

DISCUSSION

Previous studies of BPV RNA resulted in a preliminary transcription map for this virus. It was our goal to determine a detailed transcription map by cloning cDNAs. As this was unsuccessful (see below), we attempted to create the detailed map by using the polymerase chain reaction to amplify the 5' and 3' termini of BPV cDNAs.

We must demonstrate that amplified BPV-specific cDNA fragments generated using primers specific for each ORF correspond in size to that expected based on the sizes of known BPV proteins. BPV codes for two large non-structural proteins (two forms of NS1) with molecular weights of 83 and 75 kDa. This is unlike the rodent parvoviruses for which a single NS1 protein has been identified. Lederman et al. (12) found that *in vitro* translation of BPV RNA resulted in both NS1s suggesting that both proteins are coded for by independent transcripts. A small nonstructural protein, NP1, with a molecular weight of 28 kDa is also coded for by BPV. A nonstructural protein (NS2) of similar size is observed for other parvoviruses. BPV codes for three capsid proteins, VP1, VP2 and VP3 with molecular weights of 82, 72 and 60 kDa, respectively. A fourth protein, VP4, is a proteolytic product of VP3. Lederman et al. (13) found that immunoprecipitation with preimmune IgG and anticapsid IgG precipitated the capsid proteins as well as NP-1. Chymotryptic digestion of VP1 and NP-1 resulted in two bands of similar sizes. This suggests that NP1 and VP1 share sequences in common.

Other non-characterized protein may be encoded by the BPV genome. Once a complete transcription map has been produced, mRNAs may be found which do not code for known viral proteins. These proteins could be searched for in infected cells and their functions established.

BPV RNA species

Early RNA studies using S1-nuclease digestion and two dimensional gel electrophoresis showed that the BPV genome codes for four size classes of RNAs. Hybridization with BPV restriction fragments mapped the location of the 2.6 kb class to the BPV genome. Short leader sequences from the left end of the genome were joined to a 2.25 kb main body from the right end of the genome.

Northern blot analysis of BPV poly A-mRNA transferred to nitrocellulose and probed with ³²P-BPV DNA showed RNA species of 7, 5.8, 5.1, 4.2, 3.8, 3.0, 2.4, 1.8 and 1.6 kb (Fig. 21). With the exception of the 4.2 kb transcript, BPV-specific RNA species observed by Northern analysis agreed with those observed earlier (3). Electrophoresis of RNA preparations containing BPV-specific RNA on alkaline gels showed that only the 5 kb band remained after electrophoresis (Fig. 23). Those species of genome length or greater (5.1, 5.8, 7.0 Kb) likely represent contaminating DNA molecules generated by replication of the viral DNA. This contaminating DNA probably resulted from the annealing of minus strand BPV DNA with plus strand mRNA. Isolation of total RNA by precipitation or isolation of poly-A mRNA by affinity to oligo dT-cellulose would result in RNA preparations containing BPV-specific RNA-DNA hybrids. However, the possibility of the existence of a full length BPV transcript cannot be excluded since some parvoviruses like AAV and the rodent parvoviruses MVM and H-1 make a full length or close to full length transcript.

The appearance of BPV transcription products was greatest at 30 hr postinfection (Fig. 22). This disagrees with earlier studies in which these transcripts were most abundant at 20 hr postinfection. The discrepancy is likely due to the different cell types used in the two sets of experiments. In the early work (3), RNA was extracted from BPV-infected, SV-40 transformed bovine fetal lung (BFL-T) cells. Although these cells are not fully permissive for BPV as no progeny virus particles are detected, the limited replication of BPV proceeds more quickly. Larger amounts of

BPV RNA and replicative form DNA are produced than in the nontransformed BFL cells used in the present study.

All BPV transcripts initiate from P4

Northern blots containing identical RNA samples probed with ³²P-full length BPV DNA and a ³²P-BPV DNA fragment containing nucleotides 1-350 (NheI site) resulted in identical patterns. This suggests that most if not all BPV transcripts initiate at promoter sequences localized at map unit 4. Earlier studies (3) proposed that all BPV transcripts initiated at promoter sequences localized at map unit 14. BPV DNA restriction fragments were labeled at the 3' terminus by T4 DNA polymerase-mediated nucleotide replacement synthesis and hybridized to BPV RNA. The hybrids were digested with S1 nuclease and the products were observed by neutral and alkaline gel electrophoresis. Mapping of BPV RNA with the PstI/KpnI restriction fragment (nt 287 - 2404) did not result in any hybrids. The PstI site is 67 nts upstream of the NheI site. If, in fact, most BPV transcripts initiate at map unit 4, S1 nuclease treatment of the PstI/KpnI-BPV RNA hybrids would remove the 3' label and these hybrids would go undetected. I propose that the cap site of BPV transcripts lies between the PstI (nt 287) and the NheI site (nt 350) of BPV.

BPV infection results in minimum amounts of BPV transcripts

Construction of BPV cDNAs by conventional cDNA methods (15) was virtually impossible due to the low yields of BPV RNA from infected cell lysates. This is in contrast to the rodent parvoviruses which produce substantial amounts of DNA as well as RNA during their infections. A commercial company attempted to construct a cDNA library from RNA isolated from BPV-infected BFL cells, but they were not successful. The cDNA library they prepared contained too few clones to be of use.

BPV transcripts must be translated very efficiently since the low amounts synthesized are sufficient for a productive infection. Previous experiments in our laboratory showed that BPV replicates much more efficiently when co-infected with bovine adenovirus. This suggests that BPV is not replicating at its maximum potential in BFL cells and that some function required for BPV replication is enhanced by replication of adenovirus. This function could be provided by an adenovirus viral protein or a host protein enhanced by adenovirus replication.

Analysis of PCR generated cDNA fragments

Amplification of BPV cDNA ends resulted in more BPV-specific fragments than predicted from earlier studies on BPV RNAs and proteins. These fragments could represent truncated cDNAs or portions of transcripts for yet uncharacterized BPV proteins. The discussion below includes only those fragments that could account for known BPV-coded proteins.

cDNA fragments of mRNAs for the BPV non-structural protein, NS-1

The majority, if not all, of the amino acid sequence of characterized parvovirus large non-structural proteins is coded for by the left ORF (18). 5' amplification of the left ORF resulted in DNA fragments of approximately 1995 and 251 bp. Estimates of the fragment sizes were based on a 1 kB ladder. When cloned and sequenced (Fig. 28, clones 1 and 2), the actual sizes of the fragments was 2119 bp and 240 bp. Clone 1 contains an insert with uninterrupted BPV nucleotide sequence from nucleotide 296 to 2409. This 2119 bp fragment, capable of coding for a 77 kDa protein, most likely codes for the major portion of both BPV NS1s (M_r of 72 and 83). The MID-I primers are not specific to the mid ORF since these can anneal to cDNAs generated from transcripts made from the left as well as the mid ORF. The 5'LT and 5'MID-I primers anneal at positions 2414 and 2640 of the BPV genome respectively. If the nucleotide sequence spanned by these

primers is part of the transcripts coding for the two NS1 proteins, then amplification with 5'MID-I should result in a fragment identical to that obtained with 5'LT but 200 nt longer. The 2239 bp fragment obtained with 5'MID-I likely represents a fragment derived from the 5' end of mRNA(s) for BPV NS1(s). Amplification of the 3' end of the left ORF resulted in a minimum of three fragments. Either of the two smaller fragments could represent the 3' terminus of the transcripts coding for the two NS1s (Fig. 30).

cDNA fragments of mRNAs for the BPV non-structural protein, NP-1

The insert in clone 2, produced after amplification using the P4 and 5'LT primers (annealing at nt 296 and 2392, respectively), with coding capacity of 8.8 kDa, most likely codes for the amino terminus of the small BPV nonstructural protein, NP-1. A large splice (splice A) spanning nt 326 to nt 2198 of the BPV genome is present in this clone. When translated, the sequence of this clone is open in reading frames 2 (left ORF) and 3 (Mid ORF). As described above, amplification with 5'MID-I should result in a fragment 200 nucleotides longer than that in clone 2, also containing splice A. A fragment produced with P4 and 5'MID-I, 467 bp long, was cloned (clone 3). Sequence analysis revealed a splice identical to that seen for clone 2 (splice A) and this fragment very likely represents the same transcript as that for clone 2.

The second clone obtained with fragments generated with P4 and 5'MID-I (Fig. 28, clone 4) contains splice A found in clones 2 and 3 as well as a smaller splice (splice B) downstream of splice A. Splice B spans nts 2337 to nt 2534 of the BPV genome. The third nucleotide of the mid ORF is the first nucleotide of the acceptor site of this splice. We propose that clones 2, 3 and 4 represent clones generated from the transcripts coding for the small (28 kDa), BPV non-structural protein, NP-1. Clones 2 and 3 may derive from immature transcripts for NP-1 while clone 4 may derive from mature transcripts for this protein. Since nts 2198 (acceptor of the splice A) through 2336 (donor of splice B) and 2535 (acceptor of splice B) through 2910 (end of the left ORF) would code for a protein with a MW of 19 kDa, a significant portion of the NP-1 transcript is proposed

to be coded for by the mid ORF (Fig. 30). The transcript coding for the small nonstructural protein NS2 in MVM has two splices, similar to those observed in clone 4 (7). In MVM, the large splice contained in the NS2 transcript causes the reading frame to shift from frame 3 (left ORF) to frame 2 (small ORF present in MVM). When translated the nucleotide sequence contained in BPV clones 2, 3 and 4 are open in reading frames 2 (left ORF) and 3 (mid ORF). With the information available we cannot identify the frame used to generate NP1 but since the carboxyl end of the protein is expected to come from the Mid ORF (ORF 2) we propose that at least this portion of the protein is read in frame 2. In MVM, the large splice contained in the NS2 transcript causes the reading frame to shift from frame 3 (left ORF) to frame 2 (small ORF present in MVM) (7). Since fragments with splice A were obtained with primers specific only to the left ORF, the mRNA for NP-1 may contain sequences from this ORF although this sequence may not be translated into protein.

We propose that both the 631 bp fragment generated with T₁₇ and 3'LT primers and the 447 bp fragment generated with T₁₇ and 3'MIDI primers represent the cDNA for the carboxyl terminus of NP-1. These fragments, if combined with the inserts of clones 2 and 4 have a coding capacity for a 28 kDa protein.

This suggestion is corroborated by the data obtained with the MIDII primers. Amplification with P4 and 5'MID-II resulted in a 700 bp fragment. This fragment likely represents BPV sequences between annealing site of 5'MID-II (3134) and 5'MID-I (2640) linked to the BPV insert contained in clone 4. Amplification with 3'MID-II and T₁₇ resulted a major 344 bp fragment. This fragment likely represents BPV sequences between the annealing site of 3' MID-II (3038) and the first internal poly A signal at 3385. If these fragments resulted from the same transcript, that transcript could code for NP-1 (Fig. 30).

The poly A signal used by these transcripts could be in the middle (nt 3385 or nt 3549) or at the far right end of the genome (nt 5403). Earlier data from this laboratory (3) suggest that these transcripts use one of the internal poly-A sites since after hybridization with the BglII fragment (nt 2954 through 5517) and S1 nuclease digestion two small fragments of 320 and 360 bp were seen, and, no hybrids mapping to the right end of the genome were observed.

cDNA fragments of the mRNAs for the BPV structural proteins

In earlier studies from our laboratory (3), it was proposed that the capsid proteins, VP1, VP2 and VP3 were coded for by a 2.6 kb RNA class which contained a 2.25 kb main body. This 2.25 kb body would contain nucleotide sequences from position 2954 (BglII site) to nucleotide 5204. Chen et al. (5) proposed that translation for VP2 and VP3 initiated at positions 3286 and 3697, respectively. Assuming that translation products from nucleotide sequences between positions 3286 and 5204 and between 3697 and 5204 are fully contained in mature VP2 and VP3 proteins, respectively, these proteins would have an M_r of 70 and 55. Lederman et al. (12) showed that VP1, VP2 and VP3 migrated on SDS gels as proteins with M_r s of 80, 72 and 62, respectively. Amplification of the right ORF with the T₁₇ primer and the 3'RT primer, which anneals at position 3743 of the BPV genome, resulted in a fragment too small to account for VP3. Based on earlier studies, this amplification should have resulted in a fragment of approximately 1461 bp. The smaller size of the fragment could result from amplification of a truncated cDNA end. When the 5' end of the right ORF was amplified, fragments of 955, 708, and 562 were observed. Numerous amplifications with different cDNA preparations consistently produced these fragments.

One interesting finding concerning the capsid proteins resulted from this data. Lederman et al. (13) found that immunoprecipitation with preimmune IgG and anticapsid IgG precipitated the BPV capsid proteins as well as NP-1. Chymotryptic digestion showed that VP1 and NP-1 shared fragments of similar sizes. If NP-1 is coded for by the mid ORF and NP-1 and VP1 share amino acid sequences, it was proposed that the 5' end of the transcript coding for VP1 was generated from the mid ORF. If this hypothesis were correct, amplifications with either 3' primer for the mid ORF (3'MIDI or 3'MID-II), should result in a large fragment representative of the VP1 transcript. In neither case was a large fragment observed. These data suggest that VP1 is not coded for in part by the mid ORF.

I propose that NP-1 is associated with one or more of the capsid proteins and that immunoprecipitation with capsid antibodies precipitated NP-1 as a contaminant. Electrophoresis of purified BPV capsids on an SDS polyacrylamide denaturing gel resulted in the four expected

BPV capsid proteins and on occasion NP-1. Chymotrypsin is a degenerate cutter, and incomplete digestion of NP-1 and VP1 with this enzyme could yield fragments unrelated in sequence but of similar sizes. Analysis of the protein sequences for all three BPV ORFs, searching for recognition sites for chymotrypsin, showed that digestion of any putative BPV protein with this enzyme could produce fragments of similar sizes but absolutely no sequence similarities. Therefore, the fragments of similar sizes observed by Lederman et al. (13), after chymotryptic digestion of NP-1 and VP1, may not contain amino acid sequences in common.

Since the BPV right ORF is too small to code for all of VP1, the amino terminus of this protein must be coded for by sequence upstream of the right ORF. Since the data described above suggest that the mid ORF does not code for this portion of VP1, we propose that the 5' end of the VP1 transcript results from one or both small ORFs present upstream of the right ORF (ORFs a and b, Fig. 20) in the same reading frame. Since the MID-I and MID-II primers annealed to sequences upstream and downstream of ORF b respectively, any transcripts containing sequences from these ORFs would not have been detected. Since splice A would delete the most upstream of these ORFs (orf a), we do not expect the transcripts for the capsid proteins to contain this splice. The leader sequence present in the capsid protein transcripts is therefore expected to be longer than that present in transcripts containing splice A.

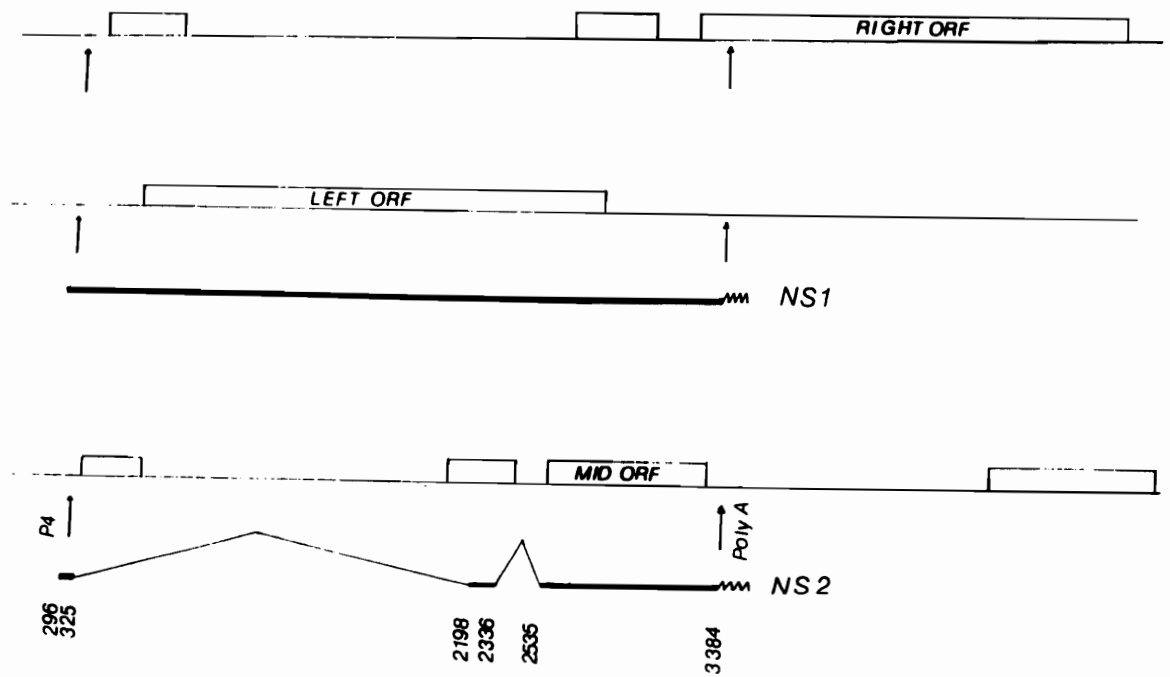


Figure 30. Schematic diagram of potential transcripts for BPV non-structural proteins, NS1 and NP-1 (NS2): Heavy lines represent sequences contained in the transcripts while thin lines represent sequences present in the BPV genome but not contained in the transcript. The nucleotide numbers of the BPV sequences contained in the transcripts are given.

LITERATURE CITED

1. Bergoin, M., M. Jourdan, M. Gervais, F. X. Jousset, S. Skory, and B. Dumas. Molecular cloning, nucleotide sequence and organization of an infectious genome of the *Junonia coenia* Densovirus (JcDENV). 1989. Abstract. EMBO Workshop: Molecular biology of parvoviruses. Kibbutz Ma'ale Hachamisha, Israel.
2. Bradley, J. E., G. A. Bishop, T. St. John, and J. A. Frelinger. A simple, rapid method for the purification of Poly-A RNA. *Biotechniques*, Vol. 6 No. 2, p 114-116.
3. Burd, P. 1982. Characterization and localization of in vivo bovine parvovirus transcription products. Dissertation, Virginia Polytechnic Institute and State University.
4. Carmichael, G. G. and McMaster, G. K. 1980. The analysis of nucleic acids in gels using glyoxal and acridine orange. *Methods Enzymol.* 65:380-390.
5. Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates. 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. *J. Virol.* 60:1085-1097.
6. Chomczynski, P. and N. Sacchi. 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry.* 162. 156-159.
7. Cotmore, S. F.. 1990. Gene expression in the autonomous parvoviruses, in *Handbook of parvoviruses*, Vol. I. Tijssen, P., Ed. CRC Press, Boca Raton, Fl.
8. Doerig, C., B. Hirt, J. Antonietti, and P. Beard. 1990. Nonstructural protein of parvoviruses B19 and Minute Virus of Mice controls transcription. *J. Virol.* 64:387-396.
9. Feinberg, A.P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.
10. Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci, USA.* Vol. 85:8998-9002.
11. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166:557-580.
12. 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.
13. Lederman, M. L., 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.
14. Liu, J. M., H. Fujii, S. W. Green, N. Komatsu, N. S. Young, and T. Shimada. 1991. Indiscriminate activity from the B 19 parvovirus P6 promoter in nonpermissive cells. *Virology* 182:361-364.
15. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring, N.Y.

16. Mount, S. M., 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459-472.
17. 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.
18. Rhode, S. L., and Iversen, P., 1990. Parvovirus genomes: DNA sequences, in (Tijssen, P., Ed.), *Handbook of parvoviruses*, Vol. I. CRC Press, Boca Raton, Fl.
19. Rodriguez, R. L. and R. C. Tait. 1983. Recombinant DNA techniques: an introduction. Addison-Wesley Publishing Company, Reading, MA.
20. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
21. Shade, R. O., B. 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.
22. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA.* 77:5201-5204.
23. Tijssen, P. Nucleotide sequence and organization of genome of *Galleria mellonella* Densovirus (GmDNV). Abstract. EMBO Workshop: Molecular biology of parvoviruses. Kibbutz Ma'ale Hachamisha, Israel.

Chapter V

SUMMARY

The mechanism(s) that determines the ratio of flip to flop terminal sequence orientations and the ratio of plus to minus DNA encapsidated has been a long-standing question in parvovirus replication. The encapsidation of DNA strands of both polarities with equal frequency is thought to result from the presence of identical terminal palindromic sequences. The work done on LuIII detailed in this dissertation clearly demonstrates that identical ends are not required for the equal encapsidation of plus and minus DNA strands.

LuIII *HindIII* fragments containing either map units 1-50 or map units 51-100 were cloned separately into pUC18 and pUC19 vectors and sequenced to obtain the nucleotide sequence of both LuIII termini. The left hairpin consists of 122 nt which can assume a T-shaped intra-strand base-paired structure and the right end consists of 211 nts which can assume a U-shaped intra-strand base-paired structure. All clones containing the intact 3' terminus had the sequence conformation designated flip by Astell while the 5' terminus had both sequence conformations, flip and flop. As suggested by the heteroduplex analysis, the palindromic nucleotide sequences and the secondary structure assumed by the LuIII termini are virtually identical to those of the rodent parvoviruses MVM and H-1 with only minor nucleotide differences. The sequence identity between the termini

of LuIII, MVM and H-1 parvoviruses, the conservation of secondary structure and the uniquely flip conformation at the left terminus of the minus strand of LuIII suggests that this virus is closely related to the rodent parvoviruses, even though it was originally isolated from a human cell line. It is very unlikely that the signals necessary for strand selection during replication are defined solely by the termini. If they were, encapsidation of only minus strand LuIII DNA is expected, based on its virtual sequence identity with MVM DNA, unless the minor nucleotide differences found between the terminal palindromes of LuIII and the rodent parvoviruses are sufficient to account for the different encapsidation patterns observed for these viruses. These findings clearly indicate that identical ends are not required for equal encapsidation of plus and minus strands and that there are nucleotide sequences other than those in the termini that influence the encapsidation pattern. Since the signals for encapsidation do not appear to reside in the termini, the complete nucleotide sequence of LuIII was determined and compared to those of MVM and H-1 in search of a putative encapsidation signal(s). A LuIII genomic clone was constructed using the LuIII *HindIII* subgenomic clones. This proved to be a difficult task since the first clones obtained contained deletions either internally or in the terminal palindromes. Digestion of the ligation products and subsequent purification of a LuIII full length molecule from the ligation product prior to its insertion into a vector resulted in a number of full length, infectious genomic clones of LuIII. The majority of the deleted clones contained identical deletions whose biological significance is not known in the 5' hairpin,

The complete nucleotide sequence of LuIII was determined by the dideoxy method using as template clones containing a nested set of deletions created by Exonuclease III digestion of the genomic clone. The LuIII genome is 5135 bases. The organization of the LuIII genome is similar to that of all mammalian parvoviruses sequenced to date. There are two large open reading frames (ORF), designated left and right, and two small ORFs in the plus strand. No ORFs of significant sizes were found in the minus strand. Promoter sequences are present at map unit 4 (nt 181) and at map unit 38 (nt 1982). LuIII shares over 80% sequence identity with MVM and H-1. Once again, this suggests that LuIII, although initially isolated from a human cell line, is a very close relative of the rodent parvoviruses. The sequence and location of P4 are virtually identical to those

of MVM and H-1. The sequence at P38 differs slightly but the regulatory sequences of P38 characterized in H-1 are present in a similar location in LuIII. The similarities of P4 and P38 among the three virus suggests that both LuIII promoters are likely functional. There are three polyadenylation signals at map units 90, 94 and 95.

Given the sequence identity of LuIII with MVM, the transcription map of LuIII is probably similar to that of MVM and H-1 since the sequence of the characterized regulatory regions and the splice donor-acceptor sites in these viruses are virtually identical.

Based on the conserved amino acid sequences of mammalian parvoviruses, we assign the left ORF of LuIII to code for the nonstructural proteins NS1 and NS2. The GKRN and WVEE amino acid sequences conserved in NS1 among parvoviruses of known sequence are also conserved in LuIII. The major parts of the two capsid proteins are thought to result from the right ORF. When translated, the majority of the amino acids changes for both the non-structural and structural proteins are localized to the carboxyl end and most of the changes are conservative. The amino acid sequence of VP2 differs the most between LuIII and the rodent parvoviruses. These changes could account for the lack of immunological cross-reactivity between these viruses.

Two regions of the LuIII sequence differ significantly from those of MVMp and H-1. At the right end of the MVMp and H-1 genomes is a sequence present in tandem. LuIII has only one copy of this sequence. Deletion of one of these copies from the MVMp genome resulted in 10 and 100 fold reductions in replication in A-9 and COS-7 cells, respectively. Although these repeats have been suggested to function as an internal origin of replication, the absence of one copy of this sequence from the LuIII genome is not expected to influence its encapsidation pattern since the immunosuppressive strain of MVM (MVMi) encapsidates primarily minus strand DNA and also has only one copy of this sequence.

Downstream of the right ORF, at map unit 89, LuIII has an A-T rich region of 47 nt which is not present in MVMp or H-1. Its location near the right terminus and its nucleotide sequence suggests that this A-T rich region could play an important role during LuIII replication. This sequence could be responsible for the characteristic virion DNA distribution of LuIII. Transfection studies with mini-genomes containing this A-T rich region and with an H-1/LuIII chimeric clone,

carried out by Solon Rhode, showed that a determinant of encapsidation is located within the last 848 nt of the LuIII genome. The A-T rich stretch is the only sequence in this region that differs significantly from MVMP and H-1. It should be possible to determine the effect of this A-T rich region on encapsidation by its deletion from the LuIII genome. This would result in a genome virtually identical to that of MVMi. Transfection with this recombinant clone is expected to result in the encapsidation of primarily minus strands. Insertion of this A-T rich sequence into the infectious clone of MVMP would result in a genome virtually identical to that of LuIII.

Encapsidation of both plus and minus DNA strands is expected after transfection. Other studies will be required to define the specific step(s) in the replication cycle at which this signal acts.

This dissertation details several approaches used to define the transcription map for BPV. Northern blot analysis of BPV RNA resulted in RNA species similar to those observed in earlier studies. RNA species in the range of 2.3 to 4.2 kb appeared as doublets that could represent immature and mature transcripts for each species. Yields of BPV RNA were greatest at 30 hours post-infection, with the 2.4 kb species being the most abundant. Alkaline gel analysis and DNase I treatment demonstrated that the 5 kb and larger RNA species likely represent contaminating DNA.

Northern blots containing identical RNA samples were probed with full length ^{32}P -labeled BPV DNA and separately with a ^{32}P -labeled BPV fragment (nt 1-352) containing BPV promoter-like sequences present at map unit 4. Identical RNA species were observed on both blots. This suggests that most, if not all, BPV transcripts initiate from promoter sequences at map unit 4 and that the promoters at map units 12 and 38 are not functional. Based on these and earlier RNA studies we propose that the cap site of most if not all BPV transcripts lies between nt 287 (PstI site) and nt 350 (NheI site) of the BPV genome.

Earlier studies of BPV RNA resulted in a preliminary transcription map for BPV. Our goal was to construct full length cDNAs by conventional methods. Due to the low yields of BPV-specific RNA in RNA preparations this was unsuccessful. Since BPV is capable of replicating efficiently the low amounts of BPV transcripts synthesized must be translated very efficiently. BPV-specific cDNA ends from total and poly-A mRNA were amplified by the polymerase chain

reaction. These amplifications resulted in a number of BPV-specific fragments. Four of these fragments were cloned. The BPV inserts in Clones 1 and 2 were generated with P4 and 5'LT. Clone 1 contained uninterrupted BPV nucleotide sequence from nt 296 to 2409. Clone 2 contained BPV sequences from nt 296 to 325 and from nt 2198 to 2409. The BPV inserts in clones 3 and 4 were generated with P4 and 5'MIDI. The BPV insert in clone 3 contains a splice identical to that in clone 2. The insert in clone 4 contains the large splice found in clones 2 and 3, and a smaller splice that spans nts 2337 to nt 2534. Translation of the nucleotide sequence of all four clones using the Pustell sequencing program showed that all four sequences were open in reading frames 2 (left ORF) and 3 (mid ORF).

The cDNA insert in clone 1, most likely codes for the major portion of both BPV NS1s. Either of the two smaller fragments generated with T₁₇ and 3'LT could represent the 3' terminus of these transcripts. Clones 2, 3 and 4 very likely represent clones generated from the transcripts coding for the small nonstructural protein, NP-1. Clones 2 and 3 may derive from immature transcripts while clone 4 may derive from mature transcripts for NP-1. Since the coding capacity of the insert in clone 4 cannot account for a 28 kDa protein we propose that a significant portion of the NP-1 transcript is coded for by the mid ORF. With the information available we cannot identify the frame used to generate NP-1 but since the carboxyl end of the protein is expected to come from the mid ORF (ORF 2) we propose that at least this portion of the protein is read in frame 2. The 631 bp fragment generated with T₁₇ and 3'LT or the 447 bp fragment generated with T₁₇ and 5'MIDI could represent the cDNA for the carboxyl terminus of NP-1. These fragments if combined with the inserts in clones 2 and 4 have a coding capacity for a 28 kDa protein.

Amplification of transcripts generated from the right ORF with T₁₇ and 3'RT resulted in fragments too short to code for even the smallest of the capsid proteins, VP3. Further work would require the synthesis of a different right ORF primer to possibly obtain a fragment representative of VP3. Since MID-I and MID-II primers annealed to sequences upstream and downstream of ORF b respectively, any transcript containing sequences from these ORFs would not have been detected. We propose that the amino terminus of the VP1 transcript results from one or both small ORFs (ORFs a and b) present upstream of the right ORF, in the same frame. To test this hy-

pothesis, amplifications using a primer specific to ORF b can be done. If our hypothesis is correct, amplifications of the right end should result in a large fragment representative of the 5' end of the VP1 transcript.

VITAE

Nanette Diffoot

April 1992

PERSONAL DATA

Birthdate: October 2, 1960
Birthplace: New York, New York
Marital Status: Married

CURRENT POSITION AND ADDRESS

Graduate Research Assistant
Department of Biology
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 231-7084 (Lab)

EDUCATION

<u>Date</u>	<u>Place</u>	<u>Major Field</u>	<u>Degree</u>
1982	University of Puerto Rico Mayaguez Campus	Biology	B.A.
1986	Virginia Polytechnic Institute and State University Thesis: <i>Corydoras aeneus</i> : A Diploid-Tetraploid Fish Species Complex	Zoology	M.S.
1986-Present	Virginia Polytechnic Institute and State University	Genetics	Ph.D.

PROFESSIONAL EXPERIENCE

1987-Present: Graduate Research Assistant, Department of Biology, Virginia Polytechnic Institute and State University.
Summer 1987: Graduate Teaching Assistant, Governor's School Camp, Virginia Polytechnic Institute and State University.

- 1986-1987: Graduate Research /Teaching Assistant, Department of Biology, Virginia Polytechnic Institute and State University.
- 1984-1986: Graduate Teaching Assistant, Department of Biology, Virginia Polytechnic Institute and State University.
- 1983: Graduate Research Assistant, Department of Philosophy (Genetics), Virginia Polytechnic Institute and State University.

HONORS AND AWARDS

National Dean's List, University of Puerto Rico, 1981 and 1982
National Hispanic Student Award, 1985-1986

MEMBERSHIP IN PROFESSIONAL SOCIETIES

Virginia Academy of Science, 1985-present
The Society of Sigma Xi

CURRENT RESEARCH INTERESTS

Research in progress includes the cloning and sequencing of the parvovirus LuIII genome, identification of any possible signals at the termini of the LuIII genome regulating encapsidation and refinement of the transcription map of BPV by construction and characterization of cDNAs made from BPV RNAs.

RESEARCH SUPPORT

- 1990: American Society for Virology Travel Grant, \$250, for attendance and presentation of research at the annual meeting.
- 1987: American Society for Virology Travel Grant, \$250, for attendance and presentation of research at the annual meeting.

PUBLICATIONS

Diffoot, N.; Shull, B.C.; Chen, K.; Stout, E.R.; Lederman, M.; Bates, R.C.; Identical ends are not required for the equal encapsidation of plus and minus-strand parvovirus LuIII DNA. *Journal of Virology*, Vol. 63, No. 7, p. 3180-3184.

MANUSCRIPTS IN PREPARATION

Diffoot, N.; Chen, K.; Lederman, M.; Bates, R.C.; Construction of an infectious genomic clone of parvovirus LuIII and its complete nucleotide sequence.

Turner, B.; Diffoot, N.; Rasch, E.M.; The callichthyid catfish *Corydoras aeneus* is an unresolved diploid-tetraploid sibling species complex.

PAPERS PRESENTED

1. Diffoot, N., Shull, B. C., Bates, R. C., and Stout, E. R. The Molecular Cloning of Parvovirus LuIII. Joint Meeting of Virginia Biochemists and Microbiologists, October, 1986; Third Annual Graduate Research Symposium at VPI&SU, November 1986 and Choices and Challenges Forum at VPI&SU, November 1986.

2. Diffoot, N., Shull, B. C., Bates, R. C., and Stout, E. R. Genomic Cloning of LuIII and Sequence Analysis of the Termini. Annual Meeting of The American Society for Virology, University of North Carolina, Chapel Hill, North Carolina, June 1987.
3. Diffoot, N., Shull, B. C., Bates, R. C., and Stout, E. R. Sequence Analysis of Genomic Clones Demonstrates Nonidentical Ends in LuIII. Parvovirus Workshop at Oxford, England, July 1987.
4. Diffoot, N., Bates, R.C., and Lederman, M.; Complete nucleotide sequence and genome organization of parvovirus LuIII. American Cancer Society, VPI&SU, March 1990.
5. Diffoot, N., Bates, R.C., and Lederman, M.; Complete nucleotide sequence and genome organization of parvovirus LuIII. Annual Meeting of The American Society for Virology, University of Utah, Salt Lake City, Utah, July 1990.
6. Diffoot, N., Bates, R.C., and Lederman, M., Characterization of bovine parvovirus transcripts by the polymerase chain reaction. Annual meeting of the Virginia Academy of Science, VPI&SU, Blacksburg, Virginia, May, 1991.
7. Diffoot, N., Lederman, M., and Bates, R.C.; Characterization of bovine parvovirus (BPV) transcripts by the polymerase chain reaction; transcripts for BPV nonstructural and structural proteins initiate from promoter sequences at map unit 4. Parvovirus Workshop at Elinore, Denmark, August, 1991.
8. Diffoot, N., R. C. Bates, and M. Lederman; A preliminary transcription map of BPV as generated by amplification of cDNA ends by PCR. Virginia Branch of the American Society for Microbiology Meeting, November 1991.

Nanette Diffoot-Carlo