# Cloning and Sequence of DNA Encoding Structural Proteins of the Autonomous Parvovirus Feline Panleukopenia Virus

JONATHAN CARLSON,<sup>1</sup>\* KEITH RUSHLOW,<sup>2</sup> IAN MAXWELL,<sup>3</sup> FRANCOISE MAXWELL,<sup>3</sup> SCOTT WINSTON,<sup>2</sup> AND WILLIAM HAHN<sup>4</sup>

Department of Microbiology, Colorado State University, Fort Collins, Colorado 80523<sup>1</sup>; Syngene Products and Research Inc., Fort Collins, Colorado 80524<sup>2</sup>; and Departments of Medicine<sup>3</sup> and Anatomy,<sup>4</sup> Health Sciences Center, University of Colorado, Denver, Colorado 80262

Received 11 January 1985/Accepted 23 April 1985

Approximately 80% of the genome of feline panleukopenia virus was cloned into pBR322. This DNA included the transcription unit for the major viral mRNA species. The nucleotide sequence of the cloned portion of the genome was determined. Comparison of the feline panleukopenia virus sequence with the sequences of the parvoviruses minute virus of mice and H-1 revealed considerable homology between the three viruses on both the nucleic acid and protein levels. Based on this homology, a model for the generation of the two size classes of viral structural proteins (VP1 and VP2') is proposed.

Feline panleukopenia virus (FPV) is an autonomous parvovirus which infects cats. The virus attacks the lymph- and blood-forming tissues and the gastrointestinal mucosa, resulting in a drop in leukocyte count and enteritis (22). In newborn kittens, the virus may attack the developing cerebellum, resulting in neurological abnormalities. FPV is interesting not only because of the disease that it causes in cats but also because of two closely related variants which have emerged in recent times. In the late 1940s, outbreaks of severe enteritis occurred in ranch mink in Canada. The agent responsible for this was shown to be an autonomous parvovirus, mink enteritis virus (22). Similarly, in the late 1970s, widespread outbreaks of enteritis and myocarditis in dogs were shown to be caused by a parvovirus, canine parvovirus (6). FPV, canine parvovirus, and mink enteritis virus are all very closely related to each other both on the protein and nucleic acid levels, as shown by antigenic cross-reaction and restriction enzyme maps (14, 17, 25).

Parvoviruses are small, DNA-containing viruses which require actively dividing cells for replication. The genomes are linear single-stranded DNA molecules of ca. 5 kilobases (kb) in length. Virions generally contain three size classes of protein. A large 80,000- to 85,000-dalton protein (VP1) comprises ca. 10 to 15% of the viral protein. The remainder of the protein is a species of ca. 64,000 to 67,000 daltons (VP2'), part of which is converted to a 60,000- to 64,000dalton species (VP2) by proteolytic cleavage (18). The amino acid sequences of the three proteins form a nested set. The entire sequence of the 60,000- to 64,000-dalton protein is contained within the 64,000- to 67,000-dalton protein, which in turn is contained within the 80,000- to 85,000-dalton

In this communication, we report the cloning of ca. 80% of the genome of FPV into bacterial plasmids. This portion of the genome includes the sequences contained in the major mRNA in infected cells as demonstrated by single-stranded specific nuclease protection experiments. The nucleotide sequence of the cloned portion of the FPV genome is presented, and comparisons of the sequences of FPV, minute virus of mice (MVM), and H-1 are discussed. The sequence includes the gene for the capsid proteins of FPV

## MATERIALS AND METHODS

Cells, virus, bacterial strains, plasmids, and enzymes. FPV was grown in Crandell feline kidney cells which were maintained in medium F12 supplemented with 5% fetal calf serum. FPV used for cloning was isolated from infected Crandell feline kidney cells grown and provided by Biologics Corporation, Inc. (Omaha, Nebr.). Restriction fragments were cloned in pBR322 and grown in Escherichia coli K-12 strain HB101. Restriction fragments were also cloned into bacteriophage M13 strains mp2, mp7, and WJ22 and grown in E. coli strain JM101. Restriction endonucleases and other enzymes used in recombinant DNA work were purchased from Bethesda Research Laboratories (Rockville, Md.), New England BioLabs (Beverly, Mass.), Boehringer Mannheim Biochemicals (Indianapolis, Ind.), or P-L Biochemicals, Inc. (Milwaukee, Wis.). All enzymes were used as recommended by the manufacturers.

Isolation of FPV viral DNA and replicative-form DNA. FPV was purified from infected cells as described by Green et al. (8). Briefly, infected cells were disrupted by three cycles of freezing and thawing, followed by sonication. The lysate was treated with DNase, RNase, trypsin, and chymotrypsin and layered onto a sucrose-cesium chloride step gradient. After centrifugation, the virus was collected from the CsCl layer. DNA was extracted from the virus by treatment with 0.2% sodium dodecyl sulfate and 0.2% proteinase K, followed by phenol extraction and ethanol precipitation. Viral DNA was made double stranded for cloning by incubating at 30°C with E. coli DNA polymerase I (Klenow fragment) and 0.02 mM each deoxynucleoside triphosphate (5). FPV replicativeform DNA was isolated from nuclei of infected cells as outlined by McMaster et al. (13). Briefly, nuclei were disrupted with 1% sodium dodecyl sulfate, NaCl was added to 1 M, and the mixture was incubated at 4°C overnight. The mixture was centrifuged at 27,000  $\times$  g, and the supernatant containing low-molecular-weight DNA was ethanol precipitated, digested with 0.5 mg of proteinase K per ml, phenol

and should be useful in the design of vaccines against FPV, canine parvovirus, and mink enteritis virus, based on recombinant DNA techniques, synthetic peptides, or other novel approaches.

<sup>\*</sup> Corresponding author.





FIG. 1. Regions of the FPV genome cloned into plasmids. The scale of the genome is shown in map units (m.p.) and kilobases (kb) in the top line. The positions of several restriction sites used in cloning are shown in the second line. The portion of the genome contained in the plasmids pEP19, pPH120, and pEH20 are indicated below the restriction map. The DNA between the EcoRI site (m.p. 20) and the PstI site (m.p. 58) was cloned between the EcoRI and PstI sites of pBR322 to form pEP19. The DNA between the PstI site (m.p. 58) and the HaeIII site (m.p. 96) was cloned between the PstI site and filled in EcoRI site of pBR322 to form pPH120. pEH20 was constructed by inserting the pEP19 and pPH120 inserts into the EcoRI site of pBR322. The portion of the genome contained in the major FPV mRNA species is also indicated.

extracted, and ethanol precipitated. The FPV replicativeform DNA was freed from contaminating cellular DNA by boiling and phenol extraction in 0.5 M NaCl. Because of its hairpin structure, a portion of the replicative-form DNA renatured immediately and was recovered in the aqueous phase as double-stranded DNA.

Single-stranded specific nuclease mapping of FPV-specific transcripts. Cytoplasmic RNA was isolated from uninfected or FPV-infected cells by suspending and lysing cells in 0.14 M NaCl-10 mM Tris (pH 8.2)-10 mM MgCl<sub>2</sub>-0.2% Nonidet P-40. Nuclei were removed by centrifugation at 1,000  $\times$  g. The supernatant was extracted with phenol-chloroform (1:1), ethanol precipitated, incubated with proteinase K at 100 µg/ml, phenol extracted, and ethanol precipitated. Mapping of the regions of DNA which were transcribed into RNA was performed essentially as described by Berk and

Sharp (4). Approximately 0.5 µg of the appropriate DNA was mixed with ca. 5 µg of total cytoplasmic RNA and ethanol precipitated. The precipitate was dissolved in 20 µl of 80% formamide-0.4 M NaCl-0.4 M PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid)] (pH 6.4)-1 mM EDTA. The solution was sealed in a glass capillary, heated to 85°C for 5 min, and incubated at 45°C for hybridization. After hybridization overnight, the reaction mixture was diluted 50-fold into 50 mM NaCl-50 mM sodium acetate (pH 5.0)-1 mM zinc acetate-0.015 mg of denatured herring sperm DNA per ml. Mung bean nuclease was added to a final concentration of 143 U/ml, and the mixture was incubated at 30°C for 1 h. Ten micrograms of yeast tRNA was added as a carrier, and the reaction was stopped by the addition of EDTA to 2 mM, sodium acetate to 0.2 M, and 2.5 volumes of ethanol. The precipitated nucleic acids were redissolved in 10 mM Tris-1



FIG. 2. Strategy for sequencing of pEH20. Restriction sites used in the determination of the sequence are indicated. All restriction fragments were labeled at the 5' terminus. The direction and extent of the sequence obtained are indicated by arrows. Abbreviations used are as follows: RI, EcoRI; RII, EcoRII; X, XbaI; HD, HindIII; N, NcoI; T, TaqI; HP, HpaI; AL, AluI; HC, HincII; P, PstI; PV, PvuII; B, BglII; S, Sau3A; HF, Hinfl; and A, AvaII. The sequence from plasmid pBR322 is indicated by jagged lines at either end.

	10	20	30	40	50	óů	70	80	90	100	110	120
GAATTCA	AACTAAAAA	+ GGAAGTGTC	*	ACTTTECE	GACTIGGTIA	+ STAAAAGAGT/	ACATCACCTO	AAGACTGGAT	GATGTTACA	+	• ATATTGAAA1	GATEG
A A						^ Heh I						
	130	140	150	140	170	180	180	200	210	224	27.4	240
	*	*	*	*	*	*	•	*	*	•	230	
LACAACC	AGGAGGTGI	MAATCTTT	AAAAAATACI	ACTIGAAATI	TGTACTTTGA	CTTTAGCAAG	RACAAAAACAI	SCATTIGAATI	ATACTTGA	RAAAGCAGAT	ATACTAAAC	TAALTA
		•	Malll									
	250	260	270	280	290	* 200	310	320	330	340	350	260
ACTTEA	ATCTTGCAA	TTCTAGAA	CATGTCAAAT	TTTTAGAATI	SCACEGATEGA	ATTEGATTAA	AGTTTGTCAC	GCTATASCAT	GTGTTTTAAA	TAGACAAGGT	GGTAAAAGAA	ATACAG
		Xbal							Ahal	11		
	370	380	390	400	410	420	430	440	450	460	470	480
TTCTTT	TTCATGGAC	* CAGCAAGTA	CAGGAAAATC	TATTATTGC	TCAAGCCATAG	*	GGGTAATGTI	GGTTGTTATA	ATECASCAA	TGTAAATTTI	CCATTTAATO	ACTETA
	490	500	510	520	570	540	550	560	570	580	590	600
CCAATA	*	•	*	*	*		*	*	* *			GTAAGC
CLIVITY		TTGGATTG	ANGANGC 166	IMACTING	1CAACAAGTTA		AGCAATTIGI	I L I BGALAAA	CARTINGART	^		
					HincII	Aha	111			BCII		70%
	610 •	620 •	¢30	640 •	650 *	660 *	670 +	680 *	690 •	700 •	*	•
AAATTG	ACCAACTC	CAGTAATTA	TGACAACTAA	TGAAAATAT	ACAATTGTAA	GAATTGGATG	TGAAGAAAGA	CCTGAACATA	CACAACCAAT	AAGAGACAGA	ATGCTGAACA	TTAAGT
	730	740	75Ú	760	770	780	790	800	810	820	820	840
TAGTAT	GTAAGCTTC	CAGGAGACT	TTGGTTTGGT	TGATAAAGA	AGAATGGCCTT	TAATATGTGC	ATGGTTAGTT	AAACATGGTT	ATGAATCAAC	CATESCTARC	TATACACATC	ATTGGG
	A Hindlll									Ncol		
	850	860	870	880	<b>8</b> 90	900	910	920	930	940	950	960
GAAAAG	TACCAGAAT	GGEATGAAA	+ ACTEGECEGA	GCCTAAAAT		TAAATTCAC			ACAAGCGGC	AAGCAATCE	*	AECACG
						. 1						
	874											1/180
	*	*	*	*	1010	*	1030	1040	1050	1060	•	1080
TICTAA	ACTECTET64	CTEEGGAEG	TAGTGGACCT	TGCACTGGA	ACCGTGGAGTA	CTECAGATAC	GCCTATTGCA	GAAACTGCAA	ATCAACAATC	AAACCAACTT	GGCGTTACTC	ACAAAG
					Sc	al						
	1090 •	1100 •	1110	1120	1130	1140	1150	1160	1170	1190	1190	1200
ACGTGC	AAGCGAGTC	CEACATEET	CCGAAATAGA	GGCAGACCT	AAGAGCCATCT	TTACTTCTGA	ACAATTGGAA	GAAGATTTTC	GAGACGACTT	GGATTAAGGT	ACGATGGCAC	CTCCGG
		Tth11	11					Xan I			Ban I	
	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320
CAAAGA	GAGCCAGGA	+ 6AGETAAGG	GTGTGTTAGT	AAAGT6666	GAGEGGAAAA	ATTTAATAAC	TTAACTAAGT	ATGTGTTTTT	TTATAGGACT	TETECCTCCA	GGTTATAAAT	ATCTTG
	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440
GGCCTG	* GGAACAGTC	*	* 6060000000	*						TERTADAAAC		•
56CC 75		TIGHCCHHO	GAGAALLAAL		^ ^	GANADAAGA				100110000	CONTRACTOR	
				BOA1	HgiDI		BOAT HIUGI					
	1450	1460	1470	1480	1440	1500	1510	•	1530	1540	1550	1560
CGCCAG	CAGATCAAC	GCTTTATAG	ATCAAACTAA	GGACGCTAC	AGATTGGGGGG	GGAAAATAGG	ACATTATTT	TTTAGAGETA	AAAAAGCAAT	TGCTCCAGTA	TTAACTGATA	CACCAG
											Fok1	
	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680
ATCATC	CATCAACAT	CAAGACCAA	CAAAACCAAC	TAAAAGAAG	TAAACCACCAC	CTCATATTT	CATCAATCTT	GCAAAAAAAA	AAAAAGCCGG	TGCAGGACAA	GTAAAAAGAG	ACAATC
	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1000
AAGCAC	AABCACCAATGAGTGATGGAGCAGTTCAACCAGACGGTGGTCAACCTBCTBTCAGAAATGAAAAGCTACABBATCTBBBAACBBOTCTGBAGGCBGGBGTBGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTG											
				- 	nell							
	1810	1870	1074								1814	1974
TRACEA		• •			• • •	1990	1870	1880	1840	1400	4	•
		STALITICA	MIAATCAGAC	GGAATTTAA	ATTTTTGGAAA	ACGGATGGGT	GGAAATCACA	GCAAACTCAA	IGCAGACTTG1		HIGLCAGAAA	
				Aha	111					Ahall	1	

FIG. 3. Nucleotide sequence of the portion of the FPV genome contained in pEH20. The positions of a number of restriction enzyme cleavage sites are shown.

1930	1940	1950	1960	1970	1980	1990	2000	2010	2020	2030	2040
ATTATAAAABAG	TAGTTETAAATA	ATATGGATAA	ACTECAETT	AAABGAAACA	TGGCTTTAGAT	TEACACTCATE	TACAAATTEI	AACACCTTGG	TCATTGETTO	GATECAAATG	TTGGG
Pst I Sfani						faNI					
2050	2060	2070	2080	2090	2100	2110	2120	2130	2140	2150	2160
	+ ATCCASSAGATT	BBCAACTAAT	TBTTAATACT/	TEASTBAST	TOCATT TAST	TABTTTTBAAG	AABAAATTTI	TAATETTETT	TTAAABACTI	# BTTTCAGAAT	CTECTA
BatXI AhaIII											
2170	2180	2190	2200	2210	2220	2230	2240	2250	2260	2270	2280
	 # Сталаститата	#			A TABATAR					+ GAGACATTGG	*
									~ ^ D-111		
								0770	2790	2380	2400
2240	2300	2310	2320	2330	2340	2350	2360	2370	2380	2370 *	*
ATCCATGGAAAC	CAACEATAEEAA	CTCCATEGAG	ATATTATTT	CAATBEGATA	GAACATTAAT	ACCATCTCAT	NCT 66AACTAI	ST GBCACACCA	ACAAATATA	TATCATEGTA	LAGATL
Ncol		BstXI NcoI									
2410	2420	2430	2440	2450 *	2460	2470	2480	2490	2500 *	2510	2520
CABATGATGTTC	AATTTTATACTA	TTGAAAATTC	TETECCAETA	CACTTACTAA	GAACAG BTGA	TGAATTTECT	CASGAACATI	TTTTTTEAT	TETAAACCA	TETAGACTAA	CACATA
										Acc1	
2530	2540	2550	2560	2570	2580	2590	2600	2610	2620	2630	2640
CATEGCAAACAA	ACABAGCATTES	GCTTACCACC	ATTTCTAAAT		ATCTGAAGG	ASCTACTAACI	TTEETEATAI	ABBASTTCAA	CAAGATAAAA	GACGTEGTE	TAACTC
2650	2660	2670	2680	2690	2700	2710	2720	2730	2740	2750	2760
*		+ +			TTRATTATAS					*	TATTE
	CASACIAIAIIA		in in in onen			A	~ 	Acci		Aballi	
						HgiAl	ngel				
<b>27</b> 70	2780	27 <b>9</b> 0 •	2800	2910	2820	2830	2840	2850	2860	2870	2880
CASCAGGACGGG	GGGBABCGCAAA	CAGATGAAAA	TCAAGCAGCA	BATSSTGATC	CAAGATATEC	ATTTGGTAGAC	AACATEETC	MAAAACTACT	ACAACAGGAG	SAAACACCTG/	AGAGAT
						AccI					
2890	2900	2910	2920	2930	2940	2950	2960	2970	2980	2990	<b>3000</b>
TTACATATATA	SCACATCAAGAT	ACAGGAAGAT	ATCCASCASS	ABATTEGATT	CAAAATATTA	ACTITAACCTI	CCTBTAACA	ATBATAATST	ATTECTACC	AACABATCCAA	TTGGAG
		E	coitV								
3010	3020	3030	3040	3050	3060	3070	3080	3090	3100	3110	3120
870000006600			TACTTATEST	*	CATTAAATAA	* TGTACCACCA	STTTATCCAA	ATESTCAAAT	TTEEGATAA	GAATTTGATA	CTGACT
GINNERCHOOM											
					7100	3190	3200	3210	3220	3230	3240
3130	3140	3150	3160	\$170	4	*	*	*	*		
TAAAACCAAGAC	TTCATGTAAATG	CACCATTET	TTGTCAAAAT	AATTGTCCTG	GTCAATTATT	TGTAAAAGTI	BEBEETAMIT	(MILINA) BA			
									STANI	3350	3340
3250	3260	3270	3280	3290	3300	3310	3320	*	3340	*	*
TETCAAGAATTE	TAACTTACTCAB	ATTTTTOSTG	BAAAGETAAA	TTABTATTTA	AABCTAAACT	AAGAGCATCT	CATACTTOBA	ATCCAATTCA	ACAAATGAGT	ATTAATSTAG	AIMALL
				Ah	aIII						
3370	3380	3390	3400	3410	3420	3430	3440	3450	3460	3470	3480
*		*	* BAAAATTETA	* TATBAAAAAT	TCAACTABC	ACCTABAAAA	TTATATTAA		TBTTTTTAT	BETTATTACAT	ATCAAC
MATTINACTAL											
	7500	3510	3520	3530	3540	3550	3540	3570	3580	3590	3600
3490	*	*	*						TACTTOTAT	TTBATATABG	ATTTAGA
TAGCACCTAGA	AATTATATTAA										
							-	7400	3700	3710	3720
3610	3620	3630	3640	3650	*	3670	3000	*	*	*	
AGETTTETTAT	TESTATACAAT	ACTETAABAA	ATASAAGAAC	ATTTAGATCA	TEGTTABTAT	GETATACAAI	AACTOTAAB	watasay	GRI I (ABATI	uni de l'Ind li	
	Accl					AccI	_				
3730	3740	3750	3760	3770	37 <b>9</b> 0 *	3790 *	<b>38</b> 00	<b>38</b> 10 +	3820	* 2820	3840
TTTATAAAA TETAATETAAACTATTAATETATETATETTATEETETEEEEEEEE											
										Ahall	I
3850	3860	3870	3880	3890	3900	3910	3920	3930	3940		
GTCTCGTATAC	TGTCTATAAGGT	BAACTAACCT	+ TACCATAAGT	ATCAATCTGT	CTTTAAGGGGG	GEGETEGETE	6GAGATGCAC	AATATCAGTA	GACTGACTG		
^ ^						Sfani		Ac	cI		





FIG. 4. Translational reading frames. The positions of the two long open translational reading frames in the sequence of pEH20 are indicated along with the portions of the sequence contained in the major FPV-specific mRNA species.

mM EDTA (pH 7.4), fractionated on an alkaline agarose gel, blotted onto diazobenzyloxymethyl-paper, and annealed to appropriate <sup>32</sup>P-labeled, nick-translated DNA probes (26).

### RESULTS

Cloning of FPV. The 3' end of DNA isolated from parvovirus virions contains inverted repeat sequences which form hairpin-like structures (1). Because of this structure, parvoviral DNA can function as a primer-template molecule for DNA polymerase. DNA from CsCl-purified FPV was converted to double-stranded DNA with E. coli DNA polymerase (5, 15) and used to clone the fragment between the *Eco*RI site at map position (m.p.) 20 and the *Pst*I site at m.p. 58 (plasmid pEP19, Fig. 1). Synthesis of full-length, doublestranded DNA with polymerase and single-stranded viral DNA proved difficult. DNA between the PstI site at m.p. 58 and the HaeIII site at m.p. 96 was cloned from doublestranded, replicative-form DNA isolated from the nuclei of infected cells (plasmid pPH120, Fig. 1). A plasmid containing the entire FPV sequence between the *Eco*RI site at m.p. 20 and the HaeIII site at m.p. 96 was constructed with the FPV inserts from pEP19 and pPH120. To have a convenient source of single-stranded DNA for single-stranded specific nuclease mapping of FPV mRNA, several fragments were cloned into appropriate bacteriophage M13 derivatives. The EcoRI-PstI fragment from pEP19 was cloned into M13mp7. The HindIII fragment between m.p. 34 and m.p. 48 was cloned into M13WJ22. The FPV fragment from pEH20 was cloned into M13mp2.

Sequence of pEH20. The sequence of the FPV portion of

TABLE 1. Nucleic acid and protein sequence homology among parvoviruses"

	% Homology for:							
Virus poir	Fra	me I	Frame II					
virus pan	Nucle- otide	Amino acid	Nucle- otide	Amino acid				
FPV-MVM	73	78	55	51				
FPV-H-1	72	78	55	52				
MVM-H-1	88	91	72	71				

" The homology between pairs of viruses for the two long open reading frames (frame I and frame II) is shown for both nucleotide and predicted amino acid sequences.

pEH20 was determined by the method of Maxam and Gilbert (11) according to the strategy shown in Fig. 2. The entire sequence of 3,948 base pairs (bp) is shown in Fig. 3. The sequence shown is that of the C strand which is complementary to the V strand packaged in the virion. The C strand contains the sequences present in the mRNAs. Restriction endonuclease cleavage sites predicted from the sequence generally agree with the published maps for FPV and mink enteritis virus (14, 25). The numbering of the pEH20 sequence begins at the EcoRI site at m.p. 20, ca. 1 kb from the left end of the FPV genome. The nucleotide sequences of H-1 and MVM are numbered from the left ends of the genomes and contain EcoRI sites at nucleotides 1087 and 1084, respectively (2, 21). If the three sequences are aligned at their EcoRI sites, then there is considerable base sequence homology among the three viruses. Therefore, comparison of the FPV sequence with the H-1 and MVM sequences is facilitated by the addition of ca. 1,100 to the number in the pEH20 sequence. The nucleotide sequence of pEH20 contains two long open translational reading frames (Fig. 4). As in MVM and H-1, one of these reading frames (frame I) includes most of the left half, and the other (frame II) includes most of the right half of the genome. Frame I includes the EcoRI site and terminates at the TAA codon at base 1181. Since the DNA to the left of the EcoRI site at ca. m.p. 20 was not cloned, the precise position of the beginning of frame I on the FPV genome is unknown. Frame II begins at base 1269 and terminates with a TAA codon at base 3443. Homology among the viruses is greater in frame I than in frame II (Table 1). FPV is about equally homologous with MVM and H-1, which are substantially more homologous to each other than to FPV in both reading frames. Electron microscopic studies on heteroduplex DNA from a number of parvoviruses have shown that the DNA sequences in the left half of the genome are highly conserved among H-1, MVM, LuIII, and Kilham rat virus. The right half of the genome is much less conserved (3). These results are consistent with the direct sequence comparison of FPV, H-1, and MVM.

The long reading frame from the left-hand portion of the genome of MVM codes for a nonstructural protein of unknown function. This protein reacts with antisera from convalescent animals of several species infected with various parvoviruses (7). The amino acid sequences of FPV and MVM (frame I) differ in only 89 of 397 positions. Fifty-one of these differences are in the carboxy-terminal 104 amino acids. For comparison, H-1 and MVM differ in only 35 of 397 HAPPAKRARGKOVLVKWGEGKNLIT HAPPAKRARGKOLRDGWLVGY HAPPAKRARGKOLRDGWLVGY

a

b

LSHICFF JOLVPPGYKYLGPGNSLDQGEPTNPSDAAAKEHDEAYAAM DWGGKILGH FEGKNPYL Y ADORF I DOTKDA PPGYKYLGPGNSL YFSAADORFIDOTKDA DWGGI GH GPGNSLDQGEPTNPSDAAAKEHDEAYDQMI **degknpy** VP2' DPDLCD--PAV GAGQVKRDNQAFMSDCA YFFG FERPTKP (KK DTDTFD GITGGUERA 6644006501 MSDG HSA YEE GGG NOT EFISFLIENGWVEITA USBRU **BEN** GGGGG **I** di 16 T G GOGWVE DT EGG PWSLVDANAWGVW ENVU KTVEEBATQPPT-YNND VKGNMA D I DOGAGODA NND UKGNM HEDDLGGQ-A INNND **1RSETLGFYPWKP** LIPSHTGTS-GTPTNIYHETDPDDV DF LTAGLIVILDSNN DQQ RYYE SSNSARG I T VTYENQ-EGTVEHNVHET LTAC ETLGFYPWKPT <u>attyk</u>vynk Mν DSNN FYTPAA A :vogolj Ð KGIFDFF OCKPCRLTHTWOTNRALGLPPFLNSLPDSEGATNFGDIGVOODNA SVTOMGNTDYI **IENCVPVHLLRTGDEF** TGT ACLOGITOL IEI RIGDE TSDTATASL TANGE THTWOTNE IONVNYV IENT001 LATEDER reth VEDTNSVK THTWOTNRELGOPPLLSTFPEADTDAGTLTAOGS **FHEIT**I CMG -VNWV YSFEASTOSPAKIPIAAGRGGADTDENGAADGDPRVAFGRONGOKTTTTGETPERTTYIAHOD **B**rs IDANGA I RE KOHGEDWAKOG DAIDS FEASRABPE ARE KVPADITOG VEKEANGSVRYSMEKOHGENWASHGPA PERMIWDETSP T-GR YPAGDWIENINFNLF VTNDNYLLPTDPEGGKTGINYT DTDLEPRLH YGPL GQIWDKE REDA NONO GRTN 'GP GSGROTKDGF IDSAPLV PLNGILTNANPIGTKNDIHF SMENSYGPL THESHPSPAYPEGQIWDKE TWNFIGDMSINVDNQFNYLPN APFVD PGC EYDPDASANHERIVTYEDF AKLRASI MI AK G ATTOSVANSYMNUK THEFYC GATL GRIVTYG **IDF** SAEDNGNSYMSVT NIGAME IVYER BOLAPRKLY KWLPSATGNMHEDPLICRPVPHMTY KWLPTATGNMOLYFLITRPVARNTY

FIG. 5. Comparison of amino acid sequences of the capsid proteins of FPV, H-1, and MVM. The amino acid sequences from the short reading frames postulated to code for the amino termini of VP1 are shown in (a). The sequences of FPV, H-1, and MVM are the top, middle, and bottom lines, respectively. The position of the putative 5'-splice junction at nucleotide 1218 is indicated by a vertical arrow. The amino acid sequences from the long open translational reading frames are shown in (b). The position of the putative 3'-splice junction at nucleotide 1291 is indicated by a vertical arrow. Spaces were introduced in the sequences to maximize homology. Regions of identical sequence in the three proteins are outlined in boxes. The postulated amino termini for VP2' are also indicated.

positions, and 23 of these differences are clustered in the carboxy-terminal 104 amino acids.

The right half of the genome in MVM and H-1 has been shown to code for the structural proteins of the virion (7, 16, 21). Figure 5 compares the amino acid sequence predicted from frame II with H-1 and MVM. Approximately the first 100 amino acids near the amino terminus are highly conserved among the three viruses. In general, regions which are homologous between FPV and H-1 are also homologous between FPV and MVM. This suggests that these regions may have some critical functional significance in the structural proteins of parvoviruses. It is interesting that despite extensive amino acid homology among parvoviruses, there is little immunological cross-reactivity among them (23).

Single-stranded specific nuclease mapping of FPV transcripts. The mRNA species produced in MVM infection have been mapped in detail (19). To determine whether the major message was similar in FPV, single-stranded specific nuclease mapping experiments were performed. When cytoplasmic RNA from FPV-infected cells was hybridized to the sequences contained in pEH20, fragments of 270 and 2,500 bp were protected (data not shown). When RNA was hybridized to the DNA contained in pEP19, fragments of 270 and 660 bp were protected (Fig. 6, lanes 5 and 6). When the DNA between the HindIII sites at m.p. 34 and m.p. 48 (nucleotides 729 through 1390) was used, a fragment of 270 bp was protected (Fig. 6, lanes 1 and 2). These results are consistent with a major mRNA species composed of a 270-bp exon mapping between the HindIII sites which is spliced to an uninterrupted stretch of 2,500 bp complementary to the 5' half of the genome. A similar result is observed for the major mRNA in MVM- or H-1-infected cells. No attempt was made to map the minor viral mRNA species.

#### DISCUSSION

Transcriptional control sites. Transcription of the major virus-specific mRNA in MVM and H-1 begins at a promoter at ca. m.p. 39. An intervening sequence mapping at ca. m.p. 46 is removed from the primary transcript to produce the mature mRNA. Consistent with this, the nucleotide sequence of MVM and H-1 contains the sequence TATAAAT at m.p. 39. This sequence, the so-called TATA box, is characteristic of eucaryotic promoters and is situated ca. 30 bp upstream from the transcriptional initiation site. This region in MVM has been shown to function as a promoter in the in vitro transcription system developed by Manley, with transcription initiating ca. 30 bp downstream from the TATA box (19). Figure 7 shows the DNA sequences of FPV, H-1, and MVM aligned beginning at this sequence. In FPV the sequence (beginning at nucleotide 893) is changed to TGTAAAT, an unusual but not unprecedented change. A similar sequence, TAAAATA (not conserved in H-1 or MVM), occurs 15 bases upstream at nucleotide 878. Either or both of these sequences might function in determining the position of the transcription initiation site in FPV. There is an AATAAA at nucleotide 3816 which probably marks the end of the mature transcript in FPV. This sequence is characteristic of polyadenylation sites for eucaryotic mRNAs and is also found in multiple copies at approximately the same position in both the H-1 and MVM sequences. A transcript beginning around nucleotide 920 and ending near nucleotide 3816 would contain most if not all of the sequences coding for the viral structural proteins VP1 and VP2'.

Upstream from the postulated polyadenylation site in FPV are two sets of direct repeat sequences. The first set, a

587-434-267 -184 -FIG. 6. Mapping of RNA sequences by single-stranded specific nuclease protection experiments. Cytoplasmic RNA from uninfected or FPV-infected CFK cells was hybridized with singlestranded DNA from M13 phage containing various FPV inserts. The hybrids were digested with mung bean nuclease, fractionated on alkaline agarose gels, blotted onto diazobenzyloxymethyl-paper and probed with <sup>32</sup>P-labeled FPV DNA. An autoradiogram of a blot is shown. M13 phage DNA (0.5 µg) containing the HindIII fragment (m.p. 34 to m.p. 48) was hybridized with 1  $\mu$ g of infected cell RNA (lane 1), 5 µg of infected cell RNA (lane 2), and 5 µg of uninfected cell RNA (lane 3). Lane 4 is a HaeIII digest of pBR322 DNA used as molecular weight markers. M13 phage DNA (0.5 µg) containing the EP19 insert (m.p. 20 to m.p. 58) was hybridized to 1 µg of

60-base repeat (nucleotides 3417 to 3476 and 3477 to 3537). overlaps the long rightward open reading frame. There is a 1-base mismatch between the repeats (base 3565 is a G and base 3525 is a T). The other set is a 51-base repeat (nucleotides 3611 to 3661 and 3662 to 3712). The significance of these direct repeats is uncertain, but it is interesting to note that the H-1 sequence contains a set of 55-base repeats in a similar location (20).

infected cell RNA (lane 5), 5 µg of infected cell RNA (lane 6), and

5 µg of uninfected cell RNA (lane 7). The positions of four molecular

weight markers are indicated in the left margin.

Coding sequences for capsid proteins. The similarity in size and relative abundance of the capsid proteins (VP1 and VP2') among parvoviruses makes it likely that the transcription and translation signals will be conserved. Initiation of translation at the ATG at nucleotide 1689 would result in a protein of 584 amino acids with a calculated molecular weight of 64,661, based on amino acid sequence. This agrees quite well with the molecular weight of VP2' (64,000 to 67,000) as determined by gel electrophoresis (18). Furthermore, this ATG codon and the three codons immediately following it are conserved among FPV, H-1, and MVM (Fig. 5b). This codon has recently been found to initiate synthesis of VP2' in H-1 (16).

The initiation of translation of VP1 could begin at the ATG codon at position 1275. This would yield a protein of 722 amino acids with a calculated molecular weight of 79,845, which agrees reasonably well with estimates of 82,000 to 83,000 as measured by gel electrophoresis. However, this initiation codon is not found in H-1 and MVM. Alternatively, translation of VP1 could begin at the ATG at position 1188. This codon begins an open reading frame whose amino acid sequence is highly conserved among FPV, MVM, and H-1 for 13 amino acids and then terminates shortly thereafter (Fig. 5a). Comparison of the sequences of the three viruses (Fig. 7) shows that the sequence AGGTAAG at position 1217



	900	920	94	0	960
FPV H1 MVM	TgTAAATTCACcA66TT TATAAATTCgCTA66TT TATAAATTCgCTA66TT TATAAATTLACTA66TT	gcAaAgaCTtAgagaCaC. CaAtgCGCTCACCATCtC CggcACGCTCACCATtca	aagCggC AAGcAa tGACtCCGAgAAGTAC cGACaCCGAaAAGTAC	tCCTCagA8tCAAgAC GCCTCTCA8CCAAAAC GCCTCTCA8CCAgAAC GCCTCTCA8CCAgAAC	CACGETCTAACTC TACGCTCTEACTC TAEGCaCTAACTC
	980	1000	102	0	1040
FPV H1 MVM	CtCTgaCtcCGGACgTa CACTTGCATCGGACCTt CACTTGCATCGGAtCTc	GtGGACCTtGCaCTgGAa BcGGACCTaGCTCTAGAG GaGGACCTgGCTtTAGAG	CCgTGGAGŁACŁCCAg CCTTGGAGCACACCAA CCTTGGAGCACACCAA	ATACgCCTaTTGCaGa ATACTCCTGTTGCGGG ATACTCCTGTTGCGGG	aACTGCAaAtcaa CACTGCAGcAAgC CACTGCAGAAAcC
	1060	,	1080	1100	1120
FPV H1 MVM	CAAtcaAaccaactt CAAAACACT6666Ag6C CAgAACACT6666Aa6C	- GGcgttACtcaCaA TGGTTCCACAGCCTGCCA TGGTTCCAaAGCCTGCCA	AGAcGtgCAAgcGAGt AGgTGcTCAACgGAGC AGATGgTCAACtGAGC	CCgACaTGGTCCGAaA CCAACcTGGTCCGAGA CCAACtTGGTCaGAGA	TaGAGGCaGAccT TCGAGGCGGATTT TCGAGGaGGATTT
	. 114	.0 1	1160	1180	
FPV H1 MVM	aAGAGCcatCTTtAcTt GAGAGCtTGCTTCAGTc GAGAGCgTGCTTCgGTg	CtGAACAaTTGGAagAAG aaGAACAGTTGGAGAgcG CgGAACcGTTGaAGAAAG	AtTTtcGaGAcGA ACTTCAaCGAGGAGCT ACTTCAGCGAG¢cGCT	CTTGGALTAAGGT GACCTTGGACTAAGGT GAACTTGGACTAAGGT	ACGATGGCACCTC ACAATGGCACCTC ACGATGGCgCCTC
	1200	1220	1240	1260	
FPV H1 MVM	CgGCaAAgAGAGCCAgg CAGCTAAAAGAGCTAAA CAGCTAAAAGAGCTAAA	AG <mark>AGGTA</mark> AGGGTgTgtta AGAGGTAAGGGgcTAAGG AG <mark>AGGTA</mark> AGGGTtTAAGG	GtaaagTGGggGGaGG GATGGTTGGTTGGTGG GATGGTTGGTTGGTGG	GGAAAAALTLAATAAC GGTACTAATGTATGAC GGTALTAATGTLTAAL	Ttaa TACCTGTTTTACA TACCTGTTTTACA
	1280	1300	<b>,</b>	1320	1340
FPV H1 MVM	-ctaaGtAtgtgtTTtt GGCCTGAAATCACTTGG GGCCTGAAATCACTTGG	TTATAGGACTEGTGCCTC TTETAGGTTGGGTGCCTC TTETAGGTTGGGTGCCTC	CaGGtTAtAAaTAtCT CTGGCTACAAGTACCT CTGGCTACAAGTACCT	tGGgCCtGGGAACAGt GGGACCAGGGAACAGC GGGACCAGGGAACAGC	CTTGACCAAGGAG CTTGACCAAGGAG CTTGACCAAGGAG

FIG. 7. Comparison of nucleotide sequences of FPV, H-1, and MVM in the mRNA promoter and splicing regions of the genomes. The sequences are aligned at the TATAAA boxes. Spaces are added to the FPV sequence to maximize homology. The numbering system is from the pEH20 sequence. The H-1 sequence shown begins at nucleotide 1979 (12). The MVM sequence shown begins at nucleotide 1976 (14). Nucleotides which are identical in at least two of the three sequences are shown as capital letters. Nucleotides which differ from the consensus are shown in lowercase. Conserved sequences which may function as 5'-splice junctions are shown in a box at nucleotide 1181 and 1217. Conserved sequences which may function as a 3'-splice junction are shown in a box at nucleotide 1285. The termination codon for the leftward open reading frame is underlined at nucleotide 1179. The conserved ATG codon postulated to initiate VP1 is underlined at nucleotide 1188.

in FPV is conserved in the three viruses. This sequence agrees well with the consensus sequence for 5'-splice junctions  $(C/A)AG \parallel GT(A/G)AGT$  (12). This putative splice junction is very near the end of the amino acid homology in this reading frame between FPV and the other two viruses. Further downstream at position 1285 is the sequence TTATAGG, which conforms well to the consensus sequence for 3'-splice junctions  $(T/C)nN(C/T)AG \parallel G$  (12). The sequence is also largely conserved in MVM and H-1. It is interesting to note that this postulated 3'-splice junction is very close to the position in which the amino acid homology between the long right-hand open reading frames of the three viruses begins. If these are the splice junctions used, then a short stretch of 10 codons (beginning with the ATG at position 1188 and ending with the G at position 1218) would be fused at the G at position 1291 to create an open reading frame of 727 codons. This open reading frame would code for a protein with a calculated molecular weight of 80,366. This agrees reasonably well with the observed molecular weight of VP1.

For the initiation of the smaller of the two virion proteins (VP2') to occur at the ATG at position 1689, it might be necessary to remove the proposed ATG for VP1 at position

1188. The presence of a second donor consensus sequence just upstream from this ATG suggests a mechanism for this removal. The sequence AAGGTA at position 1180 is conserved among the three parvoviruses and agrees well with the 5'-splice junction consensus sequence. If this splice junction at 1180 were joined to the 3'-splice junction at 1291, the ATG at 1188 would be removed, allowing the use of the next ATG at position 1689 for translation initiation of VP2'. If transcription initiates ca. 30 bp downstream from the TGTAAAT sequence at position 893 and the splice event joining nucleotide 1180 to 1291 is used, then the resulting mRNA would consist of a 5' exon of ca. 260 bp spliced to a 3' exon of ca. 2.5 kb. This is consistent with the singlestranded specific nuclease experiments described above. If, on the other hand, the 5'-splice junction at 1217 is joined to the 3'-splice junction at 1291, then translation would begin at the ATG at position 1188 and VP1 would be synthesized. This mRNA would consist of a 5' exon of ca. 290 bp spliced to a 2.5-kb exon. MVM and H-1 (and probably FPV) also produce doubly spliced RNA which is transcribed from a promoter at ca. m.p. 4. Introns of ca. 1.5 kb, located between m.p. 8 and m.p. 38, and 100 bases located around m.p. 45 are removed from this species. Either one of the

splicing schemes described above could be used to remove the intron at m.p. 45 from this RNA as well.

The translational scheme outlined above would generate proteins of reasonable molecular weights and involves features highly conserved among the three viruses. However, translational initiation of the coat proteins would probably not occur at the first AUG in the FPV mRNAs. If transcription begins ca. 30 bp downstream from the TGTAAAT sequence (at nucleotide 993), then the first ATG in the mRNA would be at nucleotide 1100. This reading frame is terminated 15 codons downstream by a TGA at position 1142. This ATG is not present in H-1 or MVM. Although translation initiates at the first AUG in many eucaryotic mRNAs, numerous exceptions have been described. Recent results suggest that initiation at a downstream AUG can occur efficiently, provided that the upstream AUG is followed by an in-frame termination codon (9, 10).

FPV has many features in common with the more thoroughly studied parvoviruses, H-1 and MVM. The major mRNA species maps in the same region of the genome in all three viruses. Comparison of the DNA sequences of FPV, H-1, and MVM in the putative promoter and splicing regions (Fig. 7) allowed us to propose a model for generating mRNAs coding for either VP1 or VP2'. This model proposes an alternate 5'-splice junction joining to a common 3'-splice junction. The proposed splice junctions conform reasonably well to consensus sequences, and all are conserved among the three parvoviruses.

Although the region of the FPV genome between m.p. 0 and ca. m.p. 20 was not cloned, it is evident that the protein coded for by the long leftward open reading frame exhibits extensive homology to that coded by H-1 and MVM. This is not surprising, since antisera from animals infected with various parvoviruses are able to immunoprecipitate this protein from in vitro translation products of MVM mRNA (7). The portion of the genome which codes for the structural proteins of FPV (the long rightward open reading frame) exhibits less homology with H-1 and MVM. Nevertheless, comparison of the predicted protein sequences for the viruses shows that there is substantial conservation of sequences between the viral proteins. During maturation of parvoviruses, a portion of VP2' is converted to VP2 by a proteolytic cleavage which removes ca. 30 amino acids from the amino terminus of VP2'. If VP2' initiates with the ATG at position 1689, then the cleavage point would be in an unusual sequence composed mainly of glycine residues with a few serines interspersed. This glycine-serine-rich region is conserved among all three parvoviruses.

## LITERATURE CITED

- 1. Astell, C., M. Smith, M. Chow, and D. Ward. 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at origins of DNA replication. Cell 17:691-703.
- 2. Astell, C. R., M. Thomson, 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.
- 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: 257-262.
- Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonucleasedigested hybrids. Cell 12:721-732.
- 5. Bourguignon, G. J., P. J. Tattersall, and D. C. Ward. 1976.

DNA of minute virus of mice: self-priming, nonpermuted, single-stranded genome with a 5'-terminal hairpin duplex. J. Virol. **20:**290–306.

- 6. Carmichael, L. E., and L. N. Binn. 1981. New canine enteric viral infections. Adv. Vet. Sci. Comp. Med. 25:1-37.
- 7. Cotmore, S. F., L. J. Sturzenbecker, and P. Tattersall. 1983. The autonomous parvovirus MVM encodes two nonstructural proteins in addition to its capside polypeptides. Virology 129: 333-343.
- Green, M. R., R. M. Lebowitz, and R. G. Roeder. 1979. Expression of the autonomous parvovirus H1 genome: evidence for a single transcriptional unit and multiple spliced polyadenylated transcripts. Cell 17:967–977.
- 9. Kozak, M. 1984. Selection of initiation sites by eukaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucleic Acids Res. 12: 3873-3893.
- 10. Liu, C. C., C. C. Simonsen, and A. D. Levinson. 1984. Initiation of translation at internal AUG codons in mammalian cells. Nature (London) 309:82-85.
- 11. Maxam, A. M., and W. Gilbert. 1977. A new method of sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. 74:560-564.
- 12. McCumber, M., and R. L. Ornstein. 1984. Molecular model for messenger RNA splicing. Science 224:402–405.
- 13. McMaster, G. K., P. Beard, H. D. Engers, and B. Hirt. 1981. Characterization of an immunosuppressive parvovirus related to the minute virus of mice. J. Virol. 38:317–326.
- McMaster, G. K., J.-D. Tratschin, and G. Siegl. 1981. Comparison of canine parvovirus with mink enteritis virus by restriction site mapping. J. Virol. 38:368–371.
- 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.
- Paradiso, P. R., K. R. Williams, and R. L. Constantino. 1984. Mapping of the amino terminus of the H-1 parvovirus major capsid protein. J. Virol. 52:77-81.
- 17. Parrish, C. R., and L. E. Carmichael. 1983. Antigenic structure and variation of canine parvovirus type-2, feline panleukopenia virus, and mink enteritis virus. Virology 129:401-414.
- Peterson, J. L., R. M. K. Dale, R. Karess, D. Leonard, and D. C. Ward. 1978. Comparison of parvovirus structural proteins: evidence for post-translational modification, p. 431-445. *In* D. C. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 19. Pintel, D., D. Dadachanji, C. R. Astell, and D. C. Ward. 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic Acids Res. 11:1019–1038.
- Rhode, S. L., III, and R. Klaassen. 1982. DNA sequence of the 5' terminus containing the replication origin of parvovirus replicative form DNA. J. Virol. 41:990-999.
- Rhode, S. L., III, and P. R. Paradiso. 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybridarrested translation. J. Virol. 45:173–184.
- Siegl, G. 1976. The parvoviruses, p. 52-67. Springer-Verlag, Vienna.
- 23. Siegl, G. 1984. Biology and pathogenicity of autonomous parvoviruses, p. 297-362. In K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
- Tattersall, P., A. J. Shatkin, and D. C. Ward. 1977. Sequence homology between the structural polypeptides of minute virus of mice. J. Mol. Biol. 111:375-394.
- 25. Tratschin, J. D., G. K. McMaster, G. Kronauer, and G. Siegl. 1982. Canine parvovirus: relationship to wild-type and vaccine strains of feline panleukopenia virus and mink enteritis virus. J. Gen. Virol. 61:33-41.
- Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzoxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. 76:3683-3687.