

SHIN-ICHI MURAMATSU,<sup>1</sup> HIROAKI MIZUKAMI, NEAL S. YOUNG, and KEVIN E. BROWN

Hematology Branch, National Heart, Lung and Blood Institute, Bethesda, Maryland 20892-1652

Received February 9, 1996; accepted April 26, 1996

We have determined the complete nucleotide sequences of adeno-associated virus 3 (AAV-3) and generated an infectious clone. The single-stranded DNA genome of AAV-3 is 4726 nucleotides in length. The positive strand contains two large open reading frames; the left open reading frame encodes the nonstructural proteins and the right open reading frame encodes the structural proteins. The coding regions are flanked by identical inverted terminal repeat sequences containing palindromes. AAV-3 has little homology with the autonomous parvoviruses or erythroviruses but has 82% overall sequence homology with AAV-2. At the amino acid level there was 88% homology with AAV-2 nonstructural (Rep) proteins and 87% homology with AAV-2 capsid proteins. In addition, AAV-3 differed importantly from AAV-2 in the lack of a typical promoter sequence (TATA box) at p40 and the presence of the consensus sequence for adenovirus-related transcription factor E4F binding within the upstream region of the p5 promoter. These results suggest that AAV-3 not only consists of serologically distinct structural proteins but that viral propagation also may be controlled by different gene regulatory elements at the transcription level. The infectious clone confirmed the sequence and may be useful for developing new vectors for gene therapy. © 1996 Academic Press, Inc.

## INTRODUCTION

Adeno-associated viruses (AAVs) are small, nonenveloped, single-stranded DNA viruses, classified as dependoviruses because they were originally thought to replicate only in the presence of a co-infecting helper adeno- or herpes virus (Murphy *et al.*, 1995). AAVs were first discovered as contaminants of laboratory stocks of adenoviruses propagated in tissue culture (Atchison *et al.*, 1965; Hoggan *et al.*, 1966). AAVs have been isolated from a variety of different animal species, including primates, dogs, cows, horses, chickens, and sheep (Murphy *et al.*, 1995), where they appear to be nonpathogenic. To date, five primate AAVs have been described, distinguished by techniques that assay capsid protein variation, including complement fixation, immunodiffusion, neutralization, and fluorescent-antibody binding (Hoggan *et al.*, 1966; Parks *et al.*, 1967; Georg-Fries *et al.*, 1984). Isolates of AAV-2, AAV-3, and AAV-5 have been obtained directly from human clinical specimens, and man appears to be their natural host. In contrast, neither AAV-1 nor AAV-4 has been isolated from clinical specimens, but serum antibodies to AAV-1 and AAV-4 have been detected in African green monkeys, suggesting that they are simian viruses (Blacklow *et al.*, 1968; Parks *et al.*, 1970). AAV-5 appears more distantly related to the other primate AAVs, the only AAV-5 isolate was from

a penile condylomatous lesion, and, epidemiologically, transmission appears to follow acquisition of herpesviruses rather than adenoviruses (Georg-Fries *et al.*, 1984).

Of the many dependoviruses described, only AAV-2 has been characterized at the genomic organization and nucleotide sequence levels (Srivastava *et al.*, 1983; Cassinotti *et al.*, 1988; Ruffing *et al.*, 1994). The virus has a single-stranded DNA genome, with identical double-stranded hairpin structures at each end. The AAV-2 genome contains three transcriptionally active promoters, from which mRNAs for four nonstructural (Rep) proteins and three capsid (Cap) proteins are transcribed by alternative splicing (Carter *et al.*, 1990). Binding of the Rep protein to the "Rep-binding" motif in the hairpin structure appears necessary for preferential targeted integration of the AAV-2 genome into host cell chromosomal DNA (Weitzman *et al.*, 1994). The ability to integrate into human chromosome in the absence of helper viruses and lack of apparent pathogenicity have made AAV-2 a popular candidate as a viral vector for gene therapy (for review, Flotte and Carter, 1995).

Apart from serological studies on the antigenic differences between AAV-2 and AAV-3 performed in the 1960s, very little is known about AAV-3. In contrast to AAV-4, not even a restriction enzyme map for the viral genome has been determined, and there is some question as to whether AAV-3 is a serotype of AAV-2 or a distinctly different virus. DNA-RNA cross-hybridization studies suggest there is only 37% homology between AAV-2 and AAV-3 (Rose *et al.*, 1968); more recent DNA cross-hybridization studies with goose parvovirus DNA also sug-

<sup>1</sup> To whom correspondence and reprint requests should be addressed at Bldg. 10/Rm. 7C218, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892-1652. Fax: (301)-496-8396; E-mail: muramats@gwgate.nhlbi.nih.gov.

gested that there was a distinct difference between the DNA sequences of AAV-2 and AAV-3 (Brown *et al.*, 1995b). In addition, polyacrylamide gel electrophoresis of virions indicated that the capsid proteins were of different sizes (Rose *et al.*, 1971), and binding studies of the two viruses have suggested that they bind to different receptors on cellular membranes (Mizukami *et al.*, 1996).

We were interested in confirming that AAV-3 was a distinctly different virus and not a serotype of AAV-2. We have therefore cloned and sequenced the entire genome of AAV-3. Sequence analysis showed similar genomic organization and high homology to AAV-2 at both the DNA and the amino acid level, but significant divergence in the structural proteins and especially in the p5 and p40 promoter organization. Production of AAV-3 infectious clone confirmed that the sequence was correct.

## MATERIALS AND METHODS

### Cell culture and preparation of viral DNA

Seed viruses (AAV-3, AAV-2, and adenoviruses types 2 and 5) and cells (KB, 293-31, and Detroit 6 cells) were purchased from American Type Culture Collection (ATCC, Rockville, MD). Cells were maintained in Dulbecco's modified Eagle's minimal essential medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum and penicillin-streptomycin. Either KB or Detroit 6 cells were infected with AAV-3 or AAV-2 at an input multiplicity of infection (m.o.i.) of 0.3 with adenovirus helper (m.o.i. 10). Cells were incubated for 40 hr before harvesting. Virions were obtained by lysis of cells and banding in a cesium chloride gradient as previously described (Mizukami *et al.*, 1996). Replicative viral DNA was obtained by extraction of infected cells and collection of the low molecular weight DNA fraction (Hirt, 1967). Virion-associated DNA was obtained by lysis of virus particles with 0.2 mg/ml proteinase K and 0.5% SDS in 50 mM NaCl at 50° for 16 hr.

### Restriction enzyme analysis

To obtain a rough restriction enzyme map of the AAV-3 genome for cloning purposes, DNA was extracted from the supernatant of infected cell cultures and single-stranded viral DNA allowed to anneal by incubation in 50 mM NaCl at 50° for 16 hr. DNA was incubated with restriction enzymes (*Bam*HI, *Eco*RI, *Hind*III, and *Pst*I), and the fragments were analyzed by Southern hybridization using a <sup>32</sup>P-labeled AAV-2 probe excised from pAV2 (ATCC).

### Cloning and sequencing of coding sequence

Replicative AAV-3 virion DNA was digested with pairs of restriction enzymes (double digest), and the products were cloned into pUC19. Clones were screened for AAV-

3 inserts by Southern hybridization using a <sup>32</sup>P-labeled AAV-2 probe, as described above.

Clones containing AAV-3 were amplified and sequenced by the dideoxynucleotide chain termination method using either Sequenase II (Amersham, UK) for manual sequencing or *Ampli*Taq DNA polymerase (Perkin Elmer, CA) for automated DNA sequencing (Applied Biosystems, Inc.). Initially, "universal" M13 forward and reverse primers were used; subsequently primers were designed from previously obtained sequence. Inserts were completely sequenced in both directions.

Additional sequence into the right termini was obtained by PCR amplification. PCR primers were designed based on the termini sequences of AAV-2 and the obtained AAV-3 sequence. The products were cloned into a cloning vector (TA cloning; Invitrogen) and sequenced as above.

### Cloning and sequencing of the termini

AAV-3 virion-associated DNA was digested with the restriction enzyme *Sac*I and the products were analyzed on agarose gels. After the appropriate bands were excised from the gels, DNA was extracted using a gel extraction kit (Qiagen, CA). Two of the fragments were ligated into the *Sac*I/*Sma*I site of pUC18. Clones containing AAV-3 terminal sequence were amplified and sequenced by a PCR-based cycle sequencing method using *Ampli*Taq DNA polymerase (Perkin Elmer).

### Cloning of the entire viral genome

To obtain the sequence of the entire genome, AAV-3 virion-associated DNA was digested with one or both restriction enzymes *Pst*I and *Sac*I. The products were cloned into pUC18 and then subcloned into pBluescript II (KS<sup>+</sup>). The sequences of the junction regions were confirmed by direct PCR-based cycle sequencing of extracted viral DNA.

### Sequence analysis

All sequences obtained were analyzed using DNASTar (DNASTar, Inc.) and GeneWorks (IntelliGenetics, Inc.). Transcription factor recognition sequences were searched with the GeneWorks program based on the Transcription Factor database of the National Center for Biotechnology Information.

### Transfection of cells with AAV plasmid

293-31 cells were transfected with 10 μg of plasmid DNA per 10-cm dish by using lipofectin (Gibco BRL) according to manufacturer's instructions. At 24 hr, the lipofectin medium was replaced with DMEM supplemented with 2% fetal calf serum and containing the helper adenovirus type 5 (Ad5; m.o.i. 10). At 40 hr, DNA was extracted (Tissue kit; Qiagen, Inc.) and digested with *Eco*RI. The

AAV3	TTGGCCACTC	CCTCTATGCG	CCTCGCTCG	CTCGGTGGGG	CCTGGCGACC	AAAGGTGCCC	AGACGGACGT	GCTTTGCACG	TCCGGCCCA	CCGAGCGAGC	100
AAV2	TTGGCCACTC	CCTCTCTGCG	CGCTCGCTCG	CTCACTAGAGG	CCGGGCGACC	AAAGGTGCCC	CGACGCCGG	GCTTTGCCC	GGGGCCCTCA	GTGAGCGAGC	100
									GC box		
AAV3	GAGTGGCAT	AGAGGGAGTG	GCCAACTCCA	TCACTAGAGG	-TATGGCAGT	GACGTA-ACG	CGAAGCGCC	GAAGCGAGAC	-CACGCCTA-	CCAG--CTG	193
AAV2	GAGCGCGCAG	AGAGGGAGTG	GCCAACTCCA	TCACTAGAGG	TTCTTGAGG	GGTGGAGTCG	TGA---COT	GAATTACGTC	ATAGGGTTAG	GGAGGCTCTG	196
									E4F/EivF		
AAV3	CGTCAGCAGT	CAGGTGACCC	TTTTGCACAC	GTTTGCACAC	CCACGTGGCC	GCTGAGGGTA	TATATTCTCG	AGTGAGCGAA	CCAGGAGCTC	CATTTTG-AC	292
AAV2	TATTAGAGGT	CACGTGAGTG	TTTTGCACAC	TTTTGCACAC	CCATGTGGTC	AGCTGGGTA	TTTAAGCCCG	AGTGAGC-AC	CGAGGGTCTC	CATTTTGAAG	295
		MLTF				p5	EivF				
AAV3	CGCGAAATT	GAACGAGCAG	CAGCCATGCC	GGGGTTCTAC	GAGATTGTCC	TGAAGGTCCC	GAGTGACCTG	GACGAGCGCC	TGCCGGGCAT	TTCTAACTCG	392
AAV2	CGGAGGTTT	GAACCGCEAG	CCGCCATGCC	GGGGTTTAC	GAGATTGTGA	TTAAGGTCCC	CAGCGACCTT	GACGGGATC	TGCCGGGCAT	TTCTGACAGC	395
			→ start			p5					
AAV3	TTTGTAACT	GGGTGGCCGA	GAAGGAATGG	GACGTGCCCG	CGGATTCGA	CATGGATCCG	AATCTGATTG	AGCAGGCACC	CCTGACCCTG	GCCGAAAAGC	492
AAV2	TTTGTAACT	GGGTGGCCGA	GAAGGAATGG	GAGTTGCCCG	CAGATTCGA	CATGGATCTG	AATCTGATTG	AGCAGGCACC	CCTGACCCTG	GCCGAGAAGC	495
AAV3	TTCAGCCGA	GTTCCTGGTG	GAGTGGGCC	GCGTGAGTAA	GGCCCCGGAG	GCCCTCTTTT	TTGTCCAGTT	CGAAAAGGGG	GAGACCTACT	TCCACCTGCA	592
AAV2	TGCAGCCGA	CTTCTGACG	GAATGGGCC	GTGTGAGTAA	GGCCCCGGAG	GCCCTTTTCT	TTGTGCAATT	TGAGAAGGGA	GAGAGCTACT	TCCACATGCA	595
AAV3	CGTGCTGATT	GAGACCATCG	GGGTCAAATC	CATGGTGGTC	GGCCGCTACG	TGAGCCAGAT	TAAAGAGAAG	CTGGTGACCC	GCATCTACCG	CGGGTTCGAG	692
AAV2	CGTGCTCGTG	GAAACACCAG	GGGTGAAATC	CATGGTGGTC	GGACGTTTTC	TGAGTCAGAT	TCGCGAAAAA	CTGATTCAGA	GAATTTACCG	CGGGATCGAG	695
									GC box		
AAV3	CCGCAGCTTC	CGAACTGGTT	CGCGGTGACC	AAAACGCAGAA	ATGGGCGCCG	GGCGGGGAAC	AAGGTGGTGG	ACGACTGCTA	CATCCCCAAC	TACCTGCTCC	792
AAV2	CCGACTTTGC	CAAACTGGTT	CGCGGTGACA	AAGACCAGAA	ATGGGCGCCG	AGCCGGGAAC	AAGGTGGTGG	ATGAGTGTGA	CATCCCCAAT	TACTTGTCTC	795
AAV3	CCAAGACCCA	GCCCAGCTC	CAGTGGGCGT	GGACTAACAT	GGACCAGTAT	TTAAGCGCCT	GTTTGAATCT	CGCGGAGCGT	AAACGGCTGG	TGGCCGAGCA	892
AAV2	CCAAGACCCA	GCCTGAGCTC	CAGTGGGCGT	GGACTAATAT	GGAAACAGTAT	TTAAGCGCCT	GTTTGAATCT	CAECCGAGCGT	AAACGGTTGG	TGGCCGAGCA	895
						p18			CCAAT box		
AAV3	TCTGACGCAC	GTGTGCAGAG	CGCAGGAGCA	GAACAAGAG	AATCAGAAC	CCAATTCTGA	CGCGCCGGTC	ATCAGGTCAA	AAACCTCAGC	CAGGTACATG	992
AAV2	TCTGACGCAC	GTGTGCAGAG	CGCAGGAGCA	GAACAAGAG	AATCAGAAATC	CCAATTCTGA	TGCGCCGGTG	ATCAGATCAA	AAACTTCAGC	CAGGTACATG	995
						p19				→ start	
AAV3	GAGCTGGTGC	GGTGGCTCGT	GGACCCGGGG	ATCAGCTCAG	AAAAGCAATG	GATTCAGGAG	GACCAGGCCT	CGTACATCTC	CTTCAACGCC	GCCTCCAAT	1092
AAV2	GAGCTGGTGC	GGTGGCTCGT	GGACAAGGGG	ATTACCTCGG	AGAAGCAGTG	GATTCAGGAG	GACCAGGCCT	CATACATCTC	CTTCAATGCG	GCCTCCAAT	1095
AAV3	CGCGGTCCCA	GATCAAGGCC	GCCCTGGACA	ATGCCCTCAA	GATCATGAGC	CTGACAAAAG	CGGCTCCGGA	CTACCTGGTG	GG-CAGCAAC	CCGCCGGAGG	1191
AAV2	CGCGGTCCCA	AATCAAGGCT	GCCTGGACA	ATGCCGGAAA	GATTATGAGC	CTGACTAAAA	CGGCCCCGA	CTACCTGGTG	GGCCAGCAGC	CCG-TGGAGG	1194

FIG. 1. Complete nucleotide sequences of AAV-3 and AAV-2 (Srivastava *et al.*, 1983; Cassinotti *et al.*, 1988; Ruffing *et al.*, 1994). The sequences correspond to the complementary (plus) strands. Darkened bases are matches. Dashes (-) indicate gaps introduced into the sequence to increase homology. Start and stop codons for nonstructural proteins (Rep) and capsid proteins (VP) are underlined. Transcription factor binding sites or promoters (TATA box, MLTF, E4F, EivF, GC box, CCAAT box) are double underlined. Polyadenylation and splice donor and acceptor sites are also shown (\*).

DNA was then fractionated by electrophoresis on 1% agarose gel, transferred to nylon membrane, and hybridized to <sup>32</sup>P-labeled AAV-3 full-length DNA that was excised from the AAV-3 plasmid with *EcoRI* and *XbaI*. From one dish, cells and supernatant were frozen and thawed three times, heated at 60° for 15 min to inactivate excess adenovirus, and used to infect additional cultures. AAV-3 antigen expression was analyzed by the indirect immunofluorescence technique. The first antibody was an anti-serum against AAV-3 virion purified by banding in a cesium chloride gradient (Braton Biotech, Inc., Rockville, MD) and the second antibody was fluorescein-conjugated goat anti-guinea pig immunoglobulin G (Zymed, San Francisco, CA).

## RESULTS

### Restriction map

Southern analysis of DNA extracted from AAV-3 virions produced two bands of approximately 1.7 and 4.7 kb (data not shown). After incubation with multiple restriction enzymes, the 1.7-kb band remained undigested, a

result that implied a single-stranded DNA structure. The 4.7-kb band was incised with restriction enzymes *BamHI*, *HindIII*, and *PstI*, indicating self-annealed double-stranded DNA and that, similar to AAV-2, AAV-3 virions encapsidated both positive and negative strands of DNA. AAV-3 DNA was not affected by treatment with the restriction enzyme *EcoRI*.

### Cloning and sequencing of viral DNA

Self-annealed double-stranded viral DNA was incubated with pairs of restriction enzymes *PstI/HindIII* and *PstI/BamHI*, and the products were ligated into the appropriate restriction enzyme sites of pUC19. Terminal sequences were obtained by excision of virion DNA with *SacI*, and the two small *SacI* fragments were blunt-ligated into the *SacI/SmaI* site of pUC18. The PCR fragment was cloned into a pCRII vector.

To obtain the entire genomic sequence, AAV-3 virion-associated DNA was digested with one or both restriction enzymes *PstI* and *SacI*. The products were cloned into pUC18 and then subcloned into pBluescript II (KS<sup>+</sup>). As previously, the sequences of the junction regions

AAV3	ACATTACCAA	AAATCGGATC	TACCAATCC	TGGAGCTGAA	CGGTACGAT	CCGCAGTACG	CGGCTCCGT	CTTCTGGGC	TGGGCGCAA	AGAAGTCCG	1291
AAV2	ACATTTCAG	CAATCGGATT	TATAAAATTT	TGGAACATAA	CGGTACGAT	CCCAATATG	CGGCTCCGT	CTTCTGGGA	TGGGCCACGA	AAAAGTCCG	1294
AAV3	GAAGAGAAC	ACCATCTGGC	TCITFGGGCC	GGCCACGAGC	GGTAAACCA	ACATCGCGGA	AGCCATCGCC	CACGCGTGC	CCTTCTACGG	CTGCGTAAAC	1391
AAV2	CAAGAGAAC	ACCATCTGGC	TGTTFGGGCC	TGCAACTACC	GGGAGACCA	ACATCGCGGA	GGCCATAGCC	CACACTGTGC	CCTTCTACGG	GTGCGTAAAC	1394
AAV3	TGGACCAATG	AGAACTTTCC	CTTCAACGAT	TGCGTCGACA	AGATGGTGAT	CTGGTGGGAG	GAGGGCAAGA	TGACGGCCAA	GGTCGTGGAG	AGCGCCAAGG	1491
AAV2	TGGACCAATG	AGAACTTTCC	CTTCAACGAC	TGTCGTCACA	AGATGGTGAT	CTGGTGGGAG	GAGGGGAAGA	TGACGCCCAA	GGTCGTGGAG	TCCGCCAAGG	1494
GC box											
AAV3	CCATTCTGGG	CGGAAGCAAG	GTGCGCGTGG	ACCAAAAGTG	CAAGTCATCG	GCCAGATCG	AACCCTACC	CGTGATCGTC	ACCTCCAACA	CCAACATGTG	1591
AAV2	CCATTCTGGG	AGGAAGCAAG	GTGCGCGTGG	ACCAGAAATG	CAAGTCTCG	GCCAGATAG	ACCGACTCC	CGTGATCGTC	ACCTCCAACA	CCAACATGTG	1594
AAV3	CGCGTGATT	GACGGGAACA	GCACCACCTT	CGAGCATCAG	CAGCCGTCG	AGGACCGGAT	GTTTGAATTT	GAACCTACCC	GCCGTTTGGG	CCATGACTTT	1691
AAV2	CGCGTGATT	GACGGGAAC	CAACGACCTT	CGAACACAG	CAGCCGTCG	AAGACCGGAT	GTTCAAATTT	GAACCTACCC	GCCGCTTGGG	TCATGACTTT	1694
MLTF											
AAV3	GGGAGGTCA	CCAAACAGGA	AGTAAAGGAC	TTTTTCCGGT	GGGCTCCGA	TCACGTGACT	GACGTGGCTC	ATGAGTTCTA	CGTCAGAAAG	GGTGGAGCTA	1791
AAV2	GGGAGGTCA	CCAAGCAGGA	AGTCAAAGAC	TTTTTCCGGT	GGGCAAAAGG	TCACGTGTTT	GAGGTGGAGC	ATGAATTCTA	CGTCAAAAGG	GGTGGAGCA	1794
promoter?											
AAV3	AGAAACGCC	CGCCTCAAT	GACGCGGATG	TAAGCAGGCC	AAACCGGAG	TGCACCTCAC	TTGCGCAGCC	GAAACGTC	GACGCGGAG	CACGCGCGA	1891
AAV2	AGAAAGACC	CGCCCCAG	GACGCGATA	TAAGTGAGCC	CAACCGGTG	CGGAGTCAG	TTGCGCAGCC	ATGACGTC	GACGCGGAG	ETTCATCAA	1894
p40											
* → splice											
AAV3	CTACCGGCAC	AGGTACCAA	ACAAATGTTT	TCGTCACTGT	GGCATGAATC	TGATGCTTTT	TCCTGTAAA	ACATGCGAGA	GAATGAATCA	AATTTCAAAT	1991
AAV2	CTACCGGAC	AGGTACCAA	ACAAATGTTT	TCGTCACTGT	GGCATGAATC	TGATGCTTTT	TCCTGTAGA	CAATGCGAGA	GAATGAATCA	GAATTTCAAAT	1994
AAV3	GTCTGTTTTA	CGCATGGTCA	AAGAGACTGT	GGGAATGCT	TCCTGGAAT	GTCAGAACT	CAACCCGTTT	CTGTCGTCAA	AAAGAAGACT	TATCAGAAAC	2091
AAV2	ATCTGCTTCA	CTCACGGACA	GAAAGACTGT	TTAGAGTGCT	TTCCCG---T	GTCAGAACT	CAACCCGTTT	CTGTCGTCAA	AAAG--G-GG	TATCAGAAAC	2088
AAV3	TGTGTCAAAT	TCATCATATC	CTGGGAAGGG	CACCCGAGAT	TGCTGTTCG	GCCTGCGATT	TGGCCAATGT	GGACTTGGAT	GACTGTGTTT	CTGAGCAATA	2191
AAV2	TGTGTCAAT	TCATCATATC	ATGGAAAGG	TGCC--AGA-	CGTTGCACT	GCCTGCGATC	TGGTCAATGT	GGATTTGGAT	GACTGCATCT	TTGAACAATA	2185
stop ← splice ← * → VP1 start splice ← * stop ←											
AAV3	AATGACTTAA	ACCAGGTATG	GCTGCTGACG	GTTATCTTCC	AGATTGGCTC	GAGGACAACC	TTTCTGAAGG	CATTCGTGAG	TGGTGGGCTC	TGAAACCTGG	2291
AAV2	AATGATTTAA	ATCAGGTATG	GCTGCCGATG	GTTATCTTCC	AGATTGGCTC	GAGGACATCT	TCTCTGAAGG	AATAAGACAG	TGGTGAAGC	TCAAACCTG-	2284
AAV3	AGTCCCTCAA	C-CCAAAGCC	AAC-CAACAA	CACCAGGACA	ACCCTCGGGG	TCTTGTGCTT	CCGGTTACA	AATACCTCGG	ACCCGGTAA	GGACTCGACA	2389
AAV2	-GCCACEAC	CACCAAGCC	CGCAGAGCGG	CATAAGGAGG	ACAGCAGGGG	TCTTGTGCTT	CCTGGGTACA	AGTACCTCGG	ACCCCTCAAC	GGACTCGACA	2383
AAV3	AAGGAGAGCC	GGTCAACGAG	GCGGACGCGG	CAGCCCTCGA	ACACGACAAA	GCTTACGACC	AGCAGCTCAA	GGCCGGTAC	AACCCGTACC	TCAAGTACAA	2489
AAV2	AGGGAGAGCC	GGTCAACGAG	GCAGACGCGG	CGGCCCTCGA	GCACGACAAA	GCTTACGACC	GGCAGCTCAA	CAGCGGAGAC	AACCCGTACC	TCAAGTACAA	2483
AAV3	CCACGCCGAC	GCCGAGTTTC	AGGAGCGTCT	TCAAGAAGAT	ACGTCTTTTG	GGGGCAACCT	TGGCAGAGCA	GTCTCCAGG	CCAAAAAGAG	GATCCTTGAG	2589
AAV2	CCACGCCGAC	GCCGAGTTTC	AGGAGCGCTT	TAAAGAAGAT	ACGTCTTTTG	GGGGCAACCT	CGGACGAGCA	GTCTCCAGG	CGAAAAAGAG	GGTCTTGAA	2583
→ VP2 start											
AAV3	CCTCTTGGTC	TGGTTGAGGA	AGCAGCTAAA	ACGGCTCTCG	GAAAGAAGGG	GGCTGTAGT	CAGTCTCTC	AGGAACGGG	CTCATCATCT	GGTGTGGCA	2689
AAV2	CCTCTGGGCC	TGGTTGAGGA	ACTGTITAA	ACGGCTCCGG	GAAAAAGAG	GCGGTAGAG	CACTCTCTG	TGGAGCAGA	CTCCTCTCG	GGAAACGGAA	2683
AAV3	AATCGGGCAA	ACAGCCTGCC	AGAAAAAGAC	TAAATTTTGG	TCAGACTGGA	GACTCAGAGT	CAGTCCAGA	CCCTCAACT	CTCGGAGAAC	CACCAGCAGC	2789
AAV2	AGGCGGGCAA	CGAGCCTGCA	AGAAAAAGAT	TGAATTTTGG	TCAGACTGGA	GACGACAGT	CAGTACCTGA	CECCAGCCT	CTCGGACAGC	CACCAGCAGC	2783
→ VP3 start											
AAV3	CCCCAAGT	TTGGATCTA	ATACAATGGC	TTCAAGCGGT	GGCGCACCAA	TGGCAGACAA	TAAAGAGGT	GCCGATGGAG	TGGGTAATTC	CTCAGGAAT	2889
AAV2	CCCCETGGT	CTGGAACTA	ATACGATGGC	TACAGGCAGT	GGCGCACCAA	TGGCAGACAA	TAAAGAGGCC	GCCGACGGAG	TGGGTAATTC	CTCGGGAAT	2883
AAV3	TGGCATTGCG	ATTCCCAATG	GCTGGGCGAC	AGAGTCATCA	CCACCAGCAC	GAGAACCTGG	GCCCTGCCA	CTTACAACAA	CCATCTCTAC	AAGCAAATCT	2989
AAV2	TGGCATTGCG	ATTCCCAATG	GATGGGCGAC	AGAGTCATCA	CCACCAGCAC	CCGAACCTGG	GCCCTGCCA	CTTACAACAA	CCACTCTAC	AAACAAATTT	2983
AAV3	CCAGCCAATC	AGGAGCTTCA	AACGACAACC	ACTACTTTGG	CTACAGCACC	CCTTGGGGGT	ATTTTGACTT	TAAACAGATC	CACTGCCACT	TCTACCAGC	3089
AAV2	CCAGCCAATC	AGGAGCCTCG	AACGACAATC	ACTACTTTGG	CTACAGCACC	CCTTGGGGGT	ATTTTGACTT	CAACAGATC	CACTGCCACT	TTTACCAGC	3083
AAV3	TGACTGGCAG	GACTCATTA	ACAACAAGTC	GGGATTCGCG	CCCAAGAAC	TCAGCTTCAA	GCTCTTCAAC	ATCCAAGTTA	GAGGGTTCAC	GCAGAACGAT	3189
AAV2	TGACTGGCAA	AGACTCATCA	ACAACAAGTC	GGGATTCGGA	CCCAAGAGAC	TCAGCTTCAA	GCTCTTAAAC	ATCCAAGTCA	AAGAGGTTCAC	GCAGAAATGAC	3183
AAV3	GGCAGGAGCA	CTATTGCCAA	TAACCTTACC	AGCACGGTTC	AAGTGTITAC	GGACTCGGAG	TATCAGTCC	CGTACGTGCT	CGGTCGGCG	CACCAAGGCT	3289
AAV2	GGTACGAGCA	CGATTGCCAA	TAACCTTACC	AGCACGGTTC	AGGTGTITAC	TGACTCGGAG	TACCAGTCC	CGTACGTCT	CGGTCGGCG	CATCAAGGAT	3283
AAV3	GTCTCCCGCC	GTTTCCAGCG	GACGCTTCCA	TGGTCCCTCA	GTATGGATAC	CTCACCTGA	ACAACGGAG	TCAAGCGGTG	GGACGCTCAT	CCTTTACTG	3389
AAV2	GCCTCCCGCC	GTTTCCAGCA	GACGCTTCCA	TGGTGCACCA	GTATGGATAC	CTCACCTGA	ACAACGGGAG	TCAGGCACTA	GGACGCTCTT	CATTTACTG	3383
AAV3	CCTGGAGTAC	TTCCTTCCG	AGATGCTAAG	GACTGGAAT	AACCTCAAAT	TCAGTATAC	CTTCGAGGAT	GTACCTTTTC	ACACGAGCTA	CGCTCACAGC	3489
AAV2	CCTGGAGTAC	TTTCTTCTC	AGATGCTGCG	TACCGAAAC	AACCTTACTT	TCAGTATAC	TTTTGAGGAC	GTTCTTTTC	ACACGAGCTA	CGCTCACAGC	3483
AAV3	CAGAGTTGG	ATCGCTTGT	GAATCCTCT	ATTGATCAGT	ATCTGACTA	CCTGAACAGA	ACGCAAGGAA	CAACCTCTGG	AACAACCAAC	CAATCACGGC	3589
AAV2	CAGAGTCTGG	ACCGTCTCAT	GAATCCTCTC	ATCGACCAGT	ACCTGTATTA	CTTGAGCAGA	A--CAA-ACA	CTCCAAGTGG	AACCACCAGC	CATCACGGC	3580

FIG. 1—Continued

AAV3	TGCTTTT	AG	CCAGGCTGGG	CCTCAGT-CT	ATGCTTTTG	AGGECAGAAA	TTGGCTACCT	GGGCCCTGCT	ACCGGCAACA	GAGACTTTCA	AAGACTGCTA	3688
AAV2	TTCAGTTTT	TC	TCAGGCGGGA	GC-GAGTGAC	ATTCTGGGAC	AGTCTAGGAA	CTGGCTTCT	GGACCTGT	ACCGCCAGCA	GCAGATATCA	AAGACATCTG	3679
AAV3	ACGACAACA	CAACAGTAA	TTTCCTTGA	CAGCGGCCAG	CAAATATCAT	CTCAATGGCC	CGGACTCGT	GGTGAATCCA	GGACCAGCTA	TGGCAGTCA	3788	
AAV2	CGGATAACA	CAACAGTAA	TACTCTGGA	CTGGAGGTAC	CAAGTACCAC	CTCAATGGCA	GAGACTCTCT	GGTGAATCCG	GGCCCGGCCA	TGGCAAGCCA	3779	
AAV3	CAAGGACGAT	GAAGAAAAAT	TTTTCCCTAT	GCACGGCAAT	CTAATATTTG	GCAAGAAGG	GACAACGGCA	AGTAACGCAG	AATTAGATAA	TGTAATGATT	3888	
AAV2	CAAGGACGAT	GAAGAAAAAT	TTTTCCCTCA	GAGCGGGT	CTCATCTTTG	GGAAGCAAGG	CTCAGAGAAA	ACAAATGTGG	ACATGAAAA	GGTCATGATT	3879	
AAV3	ACGGATGAAG	AAGAGATTCG	TACCACCAAT	CCTGTGGCAA	CAGAGCAGTA	TGGAAGTGTG	GCAATAACT	TGCAGAGCTC	AAATACAGCT	CCCAGGACTG	3988	
AAV2	ACAGACGAAG	AGGAAATCAG	GACAACCAAT	CCCGTGGCTA	CGGAGCAGTA	TGGTCTGTGA	TETACCAACC	TCCAGAGAGG	CAACAGACAA	GCAGCTACCG	3979	
AAV3	GAACGTCAA	T-CATCAGGG	GGCCTTACCT	GGCATGGTGT	GGCAAGATCG	TGACGTGTAC	CTTCAAGGAC	CTATCTGGGC	AAAGATTCTT	CACACGGATG	4087	
AAV2	CAGATGTCAA	CACACAAGGC	GTTCTT-CCA	GGCATGGTCT	GGCAGGACAG	AGATGTGTAC	CTTCAGGGC	CCATCTGGGC	AAAGATTCCA	CACACGGAGC	4078	
AAV3	GACACTTTCA	TCCTTCTCT	CTGATGGGAG	GCTTTGGACT	GAAACATCCG	CCTCCTGAAA	TCAATGATCAA	AAATACTCCG	GTACCGGCAA	ATCCTCCGAC	4187	
AAV2	GACATTTTCA	CCCTCTTCCC	CTCATGGGTG	GATTCGGACT	TAAACACCTT	CCTCCACAGA	TTCATATCAA	GAAACCCCG	GTACTTGGCA	ATCCTCTGAC	4178	
AAV3	GACTTTCAGC	CGGGCAAGT	TTGCTTCATT	TATCACTAG	TACTCCACTG	GACAGGTCAG	CGTGGAAATT	GAGTGGGAGC	TACAGAAAAG	AAACAGCAAA	4287	
AAV2	CACCTTCAGT	CGGGCAAAGT	TTGCTTCTTT	CATCACACAG	TACTCCACCG	GACAGGTCAG	CGTGGAGATC	GAGTGGGAGC	TGCAGAAAGG	AAACAGCAAA	4278	
AAV3	CGTTGGAATC	CAGAGATTCA	GTACACTTCC	AACTACAACA	AGTCTGTAA	TGTGGACTTT	ACTGTAGACA	CTAATGGTGT	TTATAGTGAA	CCTCGCCTA	4387	
AAV2	CGCTGGAATC	CCGAAATTC	GTACACTTCC	AACTACAACA	AGTCTGTAA	TGTGGACTTT	ACTGTGACA	CTAATGGCGT	GTATTAGAG	CCTCGCCCA	4378	
			stop ←		Poly A							
AAV3	TTGGAACCCG	GTATCTACA	CGAAACTTGT	GATCCTGGT	TAATCAATA	ACCGTTAAT	TGTTTTAGT	TGAACTTGG	CTCTTGTCA	CTTCTTATC	4487	
AAV2	TTGGCACCCG	ATACCTGACT	CGTAATCTGT	AATTGCTTGT	TAATCAATA	ACCGTTAAT	TGTTTTAGT	TGAACTTGG	TCTCTGCTA	TTTCTTT---	4475	
AAV3	TTTATCTTGT	TTCCATGGCT	ACTGCTAGA	TAAGCAGCGG	CCTCGCGCGC	TTGCGCTTCG	CGGTTTACA	CTGCTGGTTA	ATATTTAACT	CTCGCCATAC	4587	
AAV2	CTTATCTAGT	TTCCATGGCT	AC---GTAGA	TAAGTA----	-----GC	ATG-----G	CG-----	-----GGTTA	ATCATTAAC	ACAAGGA-AC	4540	
AAV3	CTCTAGTGAT	GGAGTTGGCC	ACTCCCTCTA	TGCGCACTEG	CTCGCTCGT	GGGCCCTGCG	GACCAAAGT	CGCCAGACGG	ACGTGCTTTG	CACGTCCGGC	4687	
AAV2	CCCTAGTGAT	GGAGTTGGCC	ACTCCCTCTC	TGCGCGCTCG	CTCGCTCACT	GAGGCCGGC	GACCAAAGT	CGCCGACGC	CGGGCTTTG	CCCGCCGGC	4640	
										GC box		
AAV3	CCCACCGAGC	GAGCGAGTGC	GCATAGAGGG	AGTGCCCAA							4726	
AAV2	CTCAGTGAGC	GAGCGAGCGC	GCAGAGAGGG	AGTGCCCAA							4679	

FIG. 1—Continued

were confirmed directly by PCR-based cycle sequencing of viral DNA.

### Nucleotide sequence

The genome of AAV-3 was 4726 nucleotides in length. Comparison with the genome of AAV-2 showed 82% overall sequence homology (Fig. 1). Among other parvoviruses, the closest relative of AAV-3 appeared to be the goose parvovirus which showed 54% sequence homology. There was <42% homology with the autonomous parvoviruses, including erythroviruses.

As with AAV-2, both ends of the AAV-3 genome consisted of identical inverted repeats (ITRs) containing palindromes of 146 bp and consistent with the formation of T-shaped hairpin structures. The AAV-2 nonstructural Rep protein binding motif (Weitzman *et al.*, 1994) and terminal resolution site (Snyder *et al.*, 1993) in the hairpin appeared highly conserved, with only a single base-pair substitution between the two AAVs (Fig. 2).

### Genomic organization

The location of ATG start and stop codons indicated that AAV-3 had a genomic organization similar to those of other members of the Parvoviridae, with one large open reading frame (ORF) in the left half of the genome

and a second long ORF in the right. There were no large open reading frames in the complementary strand, suggesting that it did not encode proteins.

A search for the consensus promoter sequence (TATA boxes) in the AAV-3 genome revealed two promoter regions starting at nucleotides 252 and 840, which we designated p5 and p18, corresponding to p5 and p19 of AAV-2. However, in contrast to AAV-2, there was no typical midgenome promoter sequence, with the sequence TGTA at nucleotide 1820 in the analogous position to the TATA sequence of the AAV-2 p40 promoter.

As with AAV-2, there was a single polyadenylation site (AATAAA) at the far right side (starting at nt 4433).

For transcription factor binding sites, there were 2 GC boxes (GGGCGG; nt 743 and 1499) and 13 CCAAT boxes as well as binding sites for transcription factors E4F (TGACGTAAC; nt 149), EivF (GTGACGT; nt 148), and MLTF (TCACGTGA; nt 1742). In AAV-2, GC boxes exist in ITRs (nt 80 and 4634) and EivF and MLTF binding sites are located upstream of the p5 promoter (nt 160, 206).

### Nonstructural (Rep) coding region

By analogy with other parvoviruses, the left-sided open reading frame (nt 318–2189) encoded the nonstructural or Rep protein(s). The AAV-3 left open reading frame



FIG. 2. Hairpin structure of the left ITR of AAV-3. The sequences homologous to the Rep binding motif and terminal resolution site (trs) of AAV-2 are outlined. Bases different from AAV-2 are in bold.

encoded a putative nonspliced nonstructural protein of 624 amino acids and of a predicted molecular weight of 71 kDa. The putative Rep protein showed homology to the nonstructural proteins of other parvoviruses, with highest relationship to the unspliced Rep78 protein of AAV-2 (88%) and lower homologies for nonstructural proteins of goose and duck parvoviruses (44%). Nonstructural proteins of other autonomous parvoviruses showed little homology to AAV-3 (B19 24%, simian 26%, bovine 19%). As with AAV-2 Rep78, the amino acid sequence included an ATP/GTP binding site motif (amino acids 334–341 of Rep 78) (Saraste *et al.*, 1990) and a putative zinc binding motif at the carboxyl terminus (Carter *et al.*, 1990) (Fig. 3a).

AAV-2 encodes four nonstructural proteins by alternative splicing. The acceptor/donor sequences for this alternative splicing also were conserved in AAV-3 (donor at nt 1904, acceptor at 2134) and if used would encode proteins corresponding to Rep68 and Rep40 of AAV-2.

### Capsid protein coding region

The right-sided open reading frame (nt 2209–4416) encoded a capsid protein of 736 amino acids, with a predicted molecular weight of 82 kDa. The putative capsid protein showed 87% homology to the VP1 capsid protein of AAV-2 and 54% homology to goose and duck parvoviruses' VP1, but little homology to other autonomous parvoviral capsid proteins (B19 19%, simian 18%, bovine 24%).

The AAV-2 VP2 capsid protein is believed to initiate from an ACG triplet, and there was an identical triplet at nucleotide 2620 in AAV-3. The spliced VP2 and VP3 proteins would contain 600 and 535 amino acids, respec-

tively, with predicted molecular weights of 66 and 60 kDa. All three AAV-3 capsid proteins showed high similarity to AAV-2 proteins (87% for VP1, 88% for VP2, and 89% for VP3), with the largest dissimilarity between them located in the middle of VP3, where two possible N-linked glycosylation sites exist in AAV-3, but not in AAV-2 (Fig. 3b).

### Biological activity of AAV-3 plasmid in human cells

Plasmid containing the entire AAV-3 sequence was transfected into cells, and the progeny DNA was compared with that produced by standard AAV-3 virus infection by Southern blot hybridization (Fig. 4). In the cells transfected with AAV-3 plasmid alone, only one band was detected (lane 3). Two bands were detected in cells subjected to AAV-3 plasmid transfection and Ad5 co-infection (lane 4). In this experiment, the upper band was at the position corresponding to the plasmid containing AAV-3 and the lower band was characteristic of normal AAV-3 infection (lane 1); in addition, the signal intensity of the lower band was stronger than that of the upper band, suggesting amplification of the viral sequence. Under our extraction conditions, only double-stranded viral DNA (either reannealed or replicative form) was detected in both the infection and the transfection lanes. No band was detected in cells infected by Ad5 alone (lane 2).

AAV-3-coded proteins after transfection were detected using an indirect immunofluorescence assay: an intranuclear fluorescence pattern indistinguishable from that seen in virus infected cells was observed (Fig. 5).

To confirm that infectious AAV-3 was produced, a virus stock was prepared from cultures transfected by AAV-3 plasmid by freezing and thawing the cells and supernatant of one 10-cm dish after plasmid transfection. After



**a**

AAV3	MPGFYEIVLK	VPSDLDERLP	GISNSFVNWV	AEKEWDVPPD	SDMDPNLIEQ	APLTVAEKLQ	REFLVEWRRV	SKAPEALFFV	QFEKGETYFH	LHVLJETIGV	100
AAV2	MPGFYEIVIK	VPSDLGHLF	GISDSFVNWV	AEKEWELPPD	SDMDLNLIEQ	APLTVAEKLQ	RDFLTEWRRV	SKAPEALFFV	QFEKGESYFH	MHWLVETTGV	100
AAV3	KSMVVGRRYS	QIKEKLVTRI	YRGVEPOLPN	WFAVTKTRNG	AGGGNKVDD	CYIPNYLLPK	TQPELQAWT	NMDQYLSACL	NLAERKRLVA	QHLTHVSQTQ	200
AAV2	KSMVLGRFLS	QIREKLIQRI	YRGIEPTLPN	WFAVTKTRNG	AGGGNKVVD	CYIPNYLLPK	TQPELQAWT	NMEQYLSACL	NLTERKRLVA	QHLTHVSQTQ	200
AAV3	EQNKENQNP	SDAPVIRSKT	SARYMELVGH	LVDRGITSEK	QWIQEQDQSY	ISFNAASNSR	SQIKAALDNA	SKIMSLTKTA	PDYLVGSNPP	EDITKNRIYQ	300
AAV2	EQNKENQNP	SDAPVIRSKT	SARYMELVGH	LVDRGITSEK	QWIQEQDQSY	ISFNAASNSR	SQIKAALDNA	GKIMSLTKTA	PDYLVGQPV	EDISSNRIYK	300
AAV3	ILELNGYDQP	YAASVFLGNA	QKKFGKRNTI	WLF <del>GP</del> ATTGK	TNIAEAI <del>AH</del> A	VPFYGCVNWT	NENFPFNDVCV	DKMVIWEEG	KMTAKV <del>V</del> ESA	KAILGGSKVR	400
AAV2	ILELNGYDQP	YAASVFLGNA	TKKFGKRNTI	WLF <del>GP</del> ATTGK	TNIAEAI <del>AH</del> T	VPFYGCVNWT	NENFPFNDVCV	DKMVIWEEG	KMTAKV <del>V</del> ESA	KAILGGSKVR	400
AAV3	VQKCKSSAQ	IEPTPVIVTS	NTNCAVIDG	NSTTFEHQOP	LQDRMF <del>F</del> EFL	TRRLDHD <del>F</del> GK	VTKQEV <del>K</del> DFF	RWADH <del>V</del> TDV	AHEFYVR <del>K</del> GG	AKKR <del>P</del> ASDA	500
AAV2	VQKCKSSAQ	IDPTPVIVTS	NTNCAVIDG	NSTTFEHQOP	LQDRMF <del>F</del> EFL	TRRLDHD <del>F</del> GK	VTKQEV <del>K</del> DFF	RWADH <del>V</del> VEV	EHEFYVR <del>K</del> GG	AKKR <del>P</del> ASDA	500
AAV3	DVSEPKRECT	SLAQPTSDA	EAPADYADRY	<del>Q</del> NKCSR <del>H</del> VGM	<del>N</del> LMLF <del>P</del> CKT <del>C</del>	<del>E</del> RMNQIS <del>N</del> VC	<del>F</del> TH <del>Q</del> R <del>D</del> GGE	<del>C</del> FP <del>G</del> MS <del>E</del> SQ	<del>V</del> SV <del>W</del> KK <del>T</del> YQ	<del>K</del> LC <del>P</del> I <del>H</del> HL <del>G</del>	600
AAV2	DISEPKRRE	SLAQPTSDA	EASIN <del>Y</del> ADRY	<del>Q</del> NKCSR <del>H</del> VGM	<del>N</del> LMLF <del>P</del> CR <del>Q</del>	<del>E</del> RMNQIS <del>N</del> IC	<del>F</del> TH <del>Q</del> K <del>D</del> CLE	<del>C</del> FP <del>V</del> SE <del>S</del> Q	<del>V</del> SV <del>W</del> KK <del>A</del> AYQ	<del>K</del> LC <del>Y</del> I <del>H</del> IM <del>G</del>	598
AAV3	RAPEIACSAC	DLANVDLDDC	VSE <del>Q</del>								624
AAV2	KVPD- <del>ACTAC</del>	DLVNV <del>D</del> LDDC	IFE <del>Q</del>								621

**b**

AAV3	MAADGYLPDW	LEDNLSEGR	EWALKPGVP	QPKANQQHQD	NRRGLVLPGY	KYLGPGNGLD	KGEPVNEADA	AAL <del>E</del> HDKAYD	QQLKAGDN <del>P</del> Y	LKYNHADA <del>E</del> F	100
AAV2	MAADGYLPDW	LEDLSEGR	QWMLKPGPP	PKPAERHKD	DSRGLVLPGY	KYLGPFNGLD	KGEPVNEADA	AAL <del>E</del> HDKAYD	RQLSDGN <del>P</del> Y	LKYNHADA <del>E</del> F	100
AAV3	QERLQEDTSF	GGNLGRAVFQ	AKKRILEPLG	LV <del>E</del> EAAKTAP	GKKGA <del>V</del> DOSP	QEPDSS <del>S</del> VG	KSGKQ <del>P</del> ARKR	LNF <del>Q</del> TGDSE	SVPDPQ <del>P</del> LGE	PPAAPT <del>S</del> LGS	200
AAV2	QERLKEDTSF	GGNLGRAVFQ	AKKRILEPLG	LV <del>E</del> EPVKTAP	GKKRP <del>V</del> EHSP	VEPDSS <del>S</del> GTG	KAGQ <del>P</del> ARKR	LNF <del>Q</del> TGDAD	SVPDPQ <del>P</del> PLG	PPAAP <del>S</del> GLGT	200
AAV3	NTMASGGGAP	MADNNEGADG	VGNS <del>S</del> GN <del>H</del> HC	DSQWL <del>D</del> RVVI	TTSTR <del>T</del> WALP	TYNNHLY <del>K</del> QI	SS <del>S</del> GSASNDN	HYFGY <del>S</del> TPWG	YFD <del>F</del> NR <del>F</del> HCH	FSP <del>R</del> DW <del>Q</del> R <del>L</del> I	300
AAV2	NTMATGSGAP	MADNNEGADG	VGNS <del>S</del> GN <del>H</del> HC	DST <del>M</del> GD <del>R</del> VI	TTSTR <del>T</del> WALP	TYNNHLY <del>K</del> QI	SS <del>S</del> GSASNDN	HYFGY <del>S</del> TPWG	YFD <del>F</del> NR <del>F</del> HCH	FSP <del>R</del> DW <del>Q</del> R <del>L</del> I	300
AAV3	NNW <del>G</del> FRPK	LSFKLFNIQV	<u>RGVTQNDGTT</u>	TJANNLTSTV	QVFTDSEYQL	PYVLGSAHQG	CLPPFPADVF	<u>MVPQYGYLTL</u>	<u>NNGSQAVGRS</u>	SPCYCLEYF <del>P</del> S	400
AAV2	NNW <del>G</del> FRPKR	LNFKLFNIQV	<u>KEVTQNDGTT</u>	TJANNLTSTV	QVFTDSEYQL	PYVLGSAHQG	CLPPFPADVF	<u>MVPQYGYLTL</u>	<u>NNGSQAVGRS</u>	SPCYCLEYF <del>P</del> S	400
AAV3	QMLRTGN <del>N</del> FQ	FSYTFEDV <del>P</del> F	HSS <del>Y</del> AHS <del>S</del> QL	DRLM <del>N</del> PLIDQ	YLY <del>L</del> NR <del>T</del> Q	TTS <del>G</del> T <del>N</del> QSR	L <del>F</del> SOAG <del>P</del> QS	MSL <del>D</del> ARN <del>N</del> LP	GPCYR <del>Q</del> OR <del>L</del> S	KTAN <del>D</del> NN <del>S</del> N	500
AAV2	QMLRTGN <del>N</del> F	FSYTFEDV <del>P</del> F	HSS <del>Y</del> AHS <del>S</del> QL	DRLM <del>N</del> PLIDQ	YLY <del>L</del> SR <del>T</del> N-	T <del>P</del> SG <del>T</del> TQSR	L <del>O</del> FSOAG <del>A</del> SD	TRD <del>S</del> RN <del>N</del> LP	GPCYR <del>Q</del> OR <del>V</del> S	KTSA <del>D</del> NN <del>N</del> SE	499
AAV3	FPWTAA <del>S</del> KYH	LNGR <del>D</del> SLV <del>N</del> P	GPAMASH <del>K</del> QD	EEK <del>F</del> FP <del>M</del> H <del>G</del> N	LIFG <del>K</del> EGTTA	SNAEL <del>N</del> VMI	TDE <del>E</del> EIR <del>T</del> TN	PVATE <del>Q</del> Y <del>G</del> TV	ANN <del>L</del> QSS <del>N</del> TA	PTTG <del>V</del> T <del>N</del> H <del>Q</del>	600
AAV2	YSWTGAT <del>K</del> YH	LNGR <del>D</del> SLV <del>N</del> P	GPAMASH <del>K</del> QD	EEK <del>F</del> FP <del>Q</del> S <del>V</del>	LIFG <del>K</del> Q <del>G</del> SEK	T <del>N</del> V <del>D</del> IE <del>K</del> VMI	TDE <del>E</del> EIR <del>T</del> TN	PVATE <del>Q</del> Y <del>G</del> SV	ST <del>N</del> L <del>Q</del> R <del>G</del> NR <del>Q</del>	AA <del>T</del> AD <del>V</del> T <del>N</del> Q <del>G</del>	599
AAV3	ALPGM <del>V</del> WQDR	DVYLQ <del>G</del> PIWA	KIPHTD <del>G</del> H <del>F</del>	PSPLM <del>G</del> G <del>F</del> GL	KHPP <del>P</del> QIMIK	NTPV <del>P</del> ANPPT	T <del>F</del> SPAK <del>F</del> ASF	ITQ <del>Y</del> ST <del>G</del> Q <del>V</del> S	VEI <del>E</del> WEL <del>Q</del> KE	NS <del>K</del> R <del>R</del> N <del>P</del> EIQ	700
AAV2	VLP <del>G</del> M <del>V</del> WQDR	DVYLQ <del>G</del> PIWA	KIPHTD <del>G</del> H <del>F</del>	PSPLM <del>G</del> G <del>F</del> GL	KHPP <del>P</del> QLIK	NTPV <del>P</del> ANPST	T <del>F</del> SAAK <del>F</del> ASF	ITQ <del>Y</del> ST <del>G</del> Q <del>V</del> S	VEI <del>E</del> WEL <del>Q</del> KE	NS <del>K</del> R <del>R</del> N <del>P</del> EIQ	699
AAV3	<u>YTSN<del>Y</del>NK<del>S</del>VN</u>	<u>VDFTV<del>D</del>TNGV</u>	<u>YSEPR<del>P</del>IGTR</u>	<u>YLTR<del>N</del>L</u>							736
AAV2	<u>YTSN<del>Y</del>NK<del>S</del>VN</u>	<u>VDFTV<del>D</del>TNGV</u>	<u>YSEPR<del>P</del>IGTR</u>	<u>YLTR<del>N</del>L</u>							735

FIG. 3. Amino acid homologies of the nonstructural proteins and capsid proteins between AAV-3 and AAV-2. AAV-2 sequence as described in the legend to Fig. 1. Darkened amino acids are matches. (a) Amino acid sequence of AAV-3 nonstructural protein to AAV-2 Rep78 sequence. The ATP/GTP binding site motif is underlined. The cysteines and histidines of a putative zinc binding motif at the carboxyl terminus are indicated (\*). (b) Amino acid sequence alignment of capsid proteins of AAV-3 and AAV-2. The first amino acids of VP2 and VP3 are indicated by arrows. The regions predicted to be exposed on the surface of the capsid are underlined. Possible N-linked glycosylation sites are marked (#).

heat treatment to inactivate excess adenovirus we used this preparation to infect additional 293-31 cells with fresh Ad5 helper (m.o.i. 10). At 38 hr, immunofluorescence with an anti-AAV-3 antibody showed an intranuclear fluorescence pattern typical of AAV-3 infection.

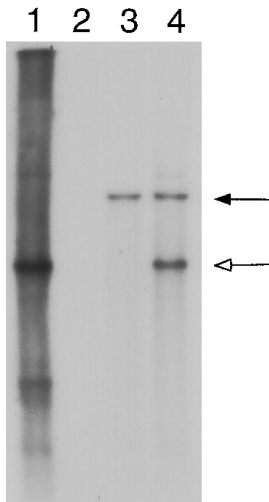
## DISCUSSION

We have cloned and sequenced the entire genome of a second human dependovirus, AAV-3. We confirmed that AAV-3 encapsidates DNA strands of both plus and minus polarity and that both strands are encapsidated with equal frequency into individual virions (Rose *et al.*, 1969), allowing double-stranded DNA to be obtained by

annealing under low-salt condition. A similar approach was used for cloning of AAV-2 (Samulski *et al.*, 1982; Senapathy and Carter, 1984), bovine parvovirus (Shull *et al.*, 1988), and simian parvovirus (Brown *et al.*, 1995a).

The genome of AAV-3 (4726 bp) was longer than that of AAV-2 (4679 bp). This difference in length can be mainly attributed to short deletions in the regions between the inverted terminal repeats and open reading frames on both sides of AAV-2 genome. Cross-hybridization studies of AAV DNA-RNA interactions performed in the 1960s indicated a homology of approximately 37% between AAV-2 and AAV-3 (Rose *et al.*, 1968). Our results show that the homology between them is much greater.

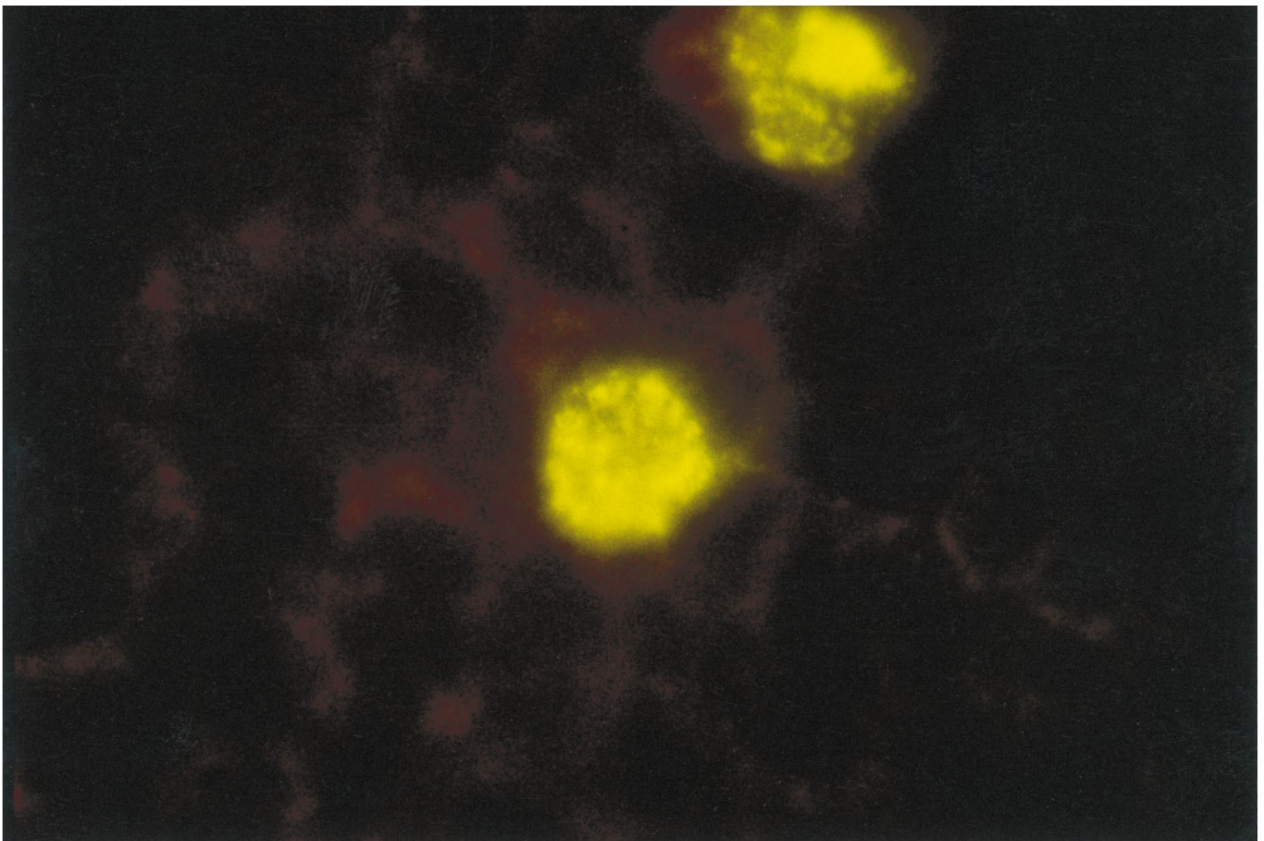
AAV-3 had a genomic organization similar to the AAV-



**FIG. 4.** Southern blot analysis of rescue and amplification of AAV-3 DNA from infectious clone. DNA was isolated from 293-31 cells that had been infected or transfected with AAV-3 virion and Ad5 (lane 1), Ad5 alone (lane 2), AAV-3 plasmid alone (lane 3), and AAV-3 plasmid and Ad5 (lane 4). The DNA was digested with *EcoRI*, fractionated on a 1% agarose gel, transferred to a nylon membrane, and hybridized to  $^{32}\text{P}$ -labeled AAV-3 DNA. Black arrow indicates bands corresponding to the AAV-3 plasmid, and white arrow indicates bands corresponding to AAV-3 virus replication.

2. The internal portion of the genome divided into two open reading frames that encode nonstructural and structural proteins. In AAV-2, nonstructural proteins Rep78 and Rep68 control DNA replication and also can mediate the formation of a complex between the ends of the AAV genome and the chromosome 19 integration locus. If the reduced number of GAGC repeats in the putative Rep binding motif of AAV-3 is significant for Rep binding, it might be expected to affect site-specific integration as well as DNA replication.

Whether the TGTA sequence at nucleotide 1820 functions as a promoter remains to be elucidated. However, it is unlikely that this sequence is a cloning or sequencing artifact because five different clones, all directly derived from either replicative form or self-annealed form of virus DNA and sequenced by both manual and automated methods, showed the same results. In addition, the corresponding region of AAV-1 also lacks a TATA box (K. E. Brown *et al.*, manuscript in preparation). Upstream from this sequence is a GC box (starting at nt 1499) and binding sites for the major late transcription factor (MLTF; nt 1742). In parvovirus B19, there is no functionally active midgenome TATA box and the GC box is a major controlling sequence for *in vitro* transcription (Blundell and



**FIG. 5.** Immunofluorescent staining with polyclonal guinea pig serum antibodies directed against AAV-3 capsids in 293-31 cells transfected with AAV-3 plasmid co-infected with Ad5 as a helper. Cells were prepared 3 days after transfection with AAV-3 plasmid.



Astell, 1989). In AAV-2, a MLTF binding site exists in the upstream region of p5 promoter and reduces the basal activity and enhances the E1A-induced activity of the p5 control region (Chang *et al.*, 1989). Thus, AAV-3 and AAV-2 may be controlled by different gene regulatory elements at the transcription level. We are now determining the transcription map of AAV-3.

It is important to note that within the upstream region of the AAV-3 p5 promoter there was a consensus sequence for transcription factor E4F or EivF binding. These cellular factors are involved in E1A-responsive transactivation of the adenovirus E4 promoter and function in the initiation of adenovirus replication (Raychaudhuri *et al.*, 1987; Cortes *et al.*, 1988; Rooney *et al.*, 1990). E4F binds to a sequence (TGACGTAAC) within the adenovirus E4 promoter overlapping two activating transcription factor (ATF) binding sites, and EivF binds to a sequence (GTG-ACGT) closely related to the ATF sequence (TGACGT). An adenovirus E4 gene is required as part of the AAV helper function during infection of human cells with adenovirus particles (Carter *et al.*, 1983), and recent studies have shown that plasmid expression of the E4 open reading frame 6 protein in cells infected with recombinant AAV increases gene transduction up to 1000-fold (Ferrari and Samulski, 1995). The EivF binding sequence overlaps the M26 motif, an enhancer element for recombination, which is also found in AAV-2 and in its integrated cellular sequence (Giraud *et al.*, 1994). The E4F binding sequence is not found in the AAV-2.

The most divergent regions for the VP3 polypeptides were located in the middle region of these proteins. Based on the atomic structure determined by X-ray crystallography of canine parvovirus, these regions may be exposed at the surface of capsid proteins (Tsao *et al.*, 1991; Chapman and Rossmann, 1993). Different capsid proteins presumably contribute to the distinct antigenicities of AAV-2 and AAV-3 and may explain the difference in binding to cell membranes of the two viruses (Mizukami *et al.*, 1996). Differences in cell binding may also reflect different cell tropisms for AAV-2 and AAV-3.

Intact AAV-3 genome can be rescued from a recombinant clone when transfected cells are superinfected with adenovirus. The rescued genome is functional in that AAV-3 protein is expressed in the nuclei of cells and infectious virus is produced. The infectious clone helps to validate the sequence that we report here. An infectious clone will allow study of AAV-3 at a molecular level comparable to AAV-2 and it may be possible to develop novel gene therapy vectors based on AAV-3.

## REFERENCES

- Atchison, R. W., Casto, B. C., and Hammon, W. M. (1965). Adenovirus-associated defective virus particles. *Science* **149**, 754–756.
- Blacklow, N. R., Hoggan, M. D., and Rowe, W. P. (1968). Serologic evidence for human infection with adenovirus-associated viruses. *J. Natl. Cancer Inst.* **40**, 319–327.
- Blundell, M. C., and Astell, C. R. (1989). A GC-box motif upstream of the B19 parvovirus unique promoter is important for in vitro transcription. *J. Virol.* **63**(11), 4814–4823.
- Brown, K. E., Green, S. W., O'Sullivan, G., and Young, N. S. (1995a). Cloning and sequencing of the simian parvovirus genome. *Virology* **210**, 314–322.
- Brown, K. E., Green, S. W., and Young, N. S. (1995b). Goose parvovirus—an autonomous member of the dependovirus genus? *Virology* **210**, 283–291.
- Carter, B. J., Marcus-Sekura, C. J., Laughlin, C. A., and Ketner, G. (1983). Properties of an adenovirus type 2 mutant, Ad2dl807, having a deletion near the right-hand genome terminus: Failure to help AAV replication. *Virology* **126**, 505–516.
- Carter, B. J., Trempe, J. P., and Mendelson, E. (1990). Adeno-associated virus gene expression and regulation. In "Handbook of Parvoviruses" (P. Tijssen, Ed.), Vol. 1, pp. 227–254. CRC Press, Boca Raton, FL.
- Cassinotti, P., Weitz, M., and Tratschin, J. D. (1988). Organization of the adeno-associated virus (AAV) capsid gene: Mapping of a minor spliced mRNA coding for virus capsid protein 1. *Virology* **167**, 176–184.
- Chang, L. S., Shi, Y., and Shenk, T. (1989). Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. *J. Virol.* **63**(8), 3479–3488.
- Chapman, M. S., and Rossmann, M. G. (1993). Structure, sequence, and function correlations among parvoviruses. *Virology* **194**, 491–508.
- Cortes, P., Buckbinder, L., Leza, M. A., Rak, N., Hearing, P., Merino, A., and Reinberg, D. (1988). EivF, a factor required for transcription of the adenovirus EIV promoter, binds to an element involved in E1A-dependent activation and cAMP induction. *Genes Dev* **2**(8), 975–990.
- Ferrari, F. K., and Samulski, R. J. (1995). The role of Ad E4 in the lifecycle of AAV. Vllth Parvovirus Workshop Abstract, p. 45.
- Flotte, T. R., and Carter, B. J. (1995). Adeno-associated virus vectors for gene therapy. *Gene Ther.* **2**(6), 357–362.
- Georg-Fries, B., Biederlack, S., Wolf, J., and zur Hausen, H. (1984). Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. *Virology* **134**, 64–71.
- Giraud, C., Winocour, E., and Berns, K. I. (1994). Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc. Natl. Acad. Sci. USA* **91**(21), 10039–10043.
- Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**(2), 365–369.
- Hoggan, M. D., Blacklow, N. R., and Rowe, W. P. (1966). Studies of small DNA viruses found in various adenovirus preparations: Physical, biological and immunological characteristics. *Proc. Nat. Acad. Sci. USA* **55**, 1467–1471.
- Mizukami, H., Young, N. S., and Brown, K. E. (1996). Adeno-associated virus type 2 binds to a 150-kilodalton cell membrane glycoprotein. *Virology* **217**, 124–130.
- Murphy, F. A., Fauquet, C. M., Mayo, M. A., Jarvis, A. W., Ghabrial, S. A., Summers, M. D., Martelli, G. P., and Bishop, D. H. L., Eds. (1995). "The Classification and Nomenclature of Viruses: Sixth Report of the International Committee on Taxonomy of Viruses." Archives of Virology, Springer-Verlag, Vienna.
- Parks, W. P., Green, M., Pina, M., and Melnick, J. L. (1967). Physicochemical characterization of adeno-associated satellite virus type 4 and its nucleic acid. *J. Virol.* **1**(5), 980–987.
- Parks, W. P., Boucher, D. W., Melnick, J. L., Taber, H., and Yow, M. D. (1970). Seroepidemiological and ecological studies of the adenovirus-associated satellite viruses. *Infect. Immun.* **2**, 716–722.
- Raychaudhuri, P., Rooney, R., and Nevins, J. R. (1987). Identification of an E1A-inducible cellular factor that interacts with regulatory sequences within the adenovirus E4 promoter. *EMBO J.* **6**(13), 4073–4081.
- Rooney, R. J., Raychaudhuri, P., and Nevins, J. R. (1990). E4F and ATF,

- two transcription factors that recognize the same site, can be distinguished both physically and functionally: A role for E4F in E1A trans activation. *Mol. Cell. Biol.* **10**(10), 5138–5149.
- Rose, J. A., Berns, K. I., Hoggan, M. D., and Koczot, F. J. (1969). Evidence for a single-stranded adenovirus-associated virus genome: Formation of a DNA density hybrid on release of viral DNA. *Proc. Natl. Acad. Sci. USA* **64**(3), 863–869.
- Rose, J. A., Hoggan, M. D., Koczot, F., and Shatkin, A. J. (1968). Genetic relatedness studies with adenovirus-associated viruses. *J. Virol.* **2**(10), 999–1005.
- Rose, J. A., Maizel, J. V., Jr., Inman, J. K., and Shatkin, A. J. (1971). Structural proteins of adenovirus-associated viruses. *J. Virol.* **8**(5), 766–770.
- Ruffing, M., Heid, H., and Kleinschmidt, J. A. (1994). Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: Lack of an RGD integrin-binding motif. *J. Gen. Virol.* **75**(Pt 12), 3385–3392.
- Samulski, R. J., Berns, K. I., Tan, M., and Muzyczka, N. (1982). Cloning of adeno-associated virus into pBR322: Rescue of intact virus from the recombinant plasmid in human cells. *Proc. Natl. Acad. Sci. USA* **79**(3), 2077–2081.
- Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990). The P-loop: A common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434.
- Senapathy, P., and Carter, B. J. (1984). Molecular cloning of adeno-associated virus variant genomes and generation of infectious virus by recombination in mammalian cells. *J. Biol. Chem.* **259**(7), 4661–4666.
- Shull, B. C., Chen, K. C., Lederman, M., Stout, E. R., and Bates, R. C. (1988). Genomic clones of bovine parvovirus: Construction and effect of deletions and terminal sequence inversions on infectivity. *J. Virol.* **62**(2), 417–426.
- Snyder, R. O., Im, D. S., Ni, T., Xiao, X., Samulski, R. J., and Muzyczka, N. (1993). Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. *J. Virol.* **67**(10), 6096–6104.
- Srivastava, A., Lusby, E. W., and Berns, K. I. (1983). Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* **45**(2), 555–564.
- Tsao, J., Chapman, M. S., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T. J., Rossmann, M. G., Compans, R. W., *et al.* (1991). The three-dimensional structure of canine parvovirus and its functional implications. *Science* **251**(5000), 1456–1464.
- Weitzman, M. D., Kyostio, S. R., Kotin, R. M., and Owens, R. A. (1994). Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. USA* **91**(13), 5808–5812.
- Zadori, Z., Stefancsik, R., Rauch, T., and Kisary, J. (1995). Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. *Virology* **212**, 562–573.