JOURNAL OF VIROLOGY, Nov. 1997, p. 8429–8436 Vol. 71, No. 11
0022-538X/97/$04.00+0

Adeno-Associated Virus Vector Integration Junctions

ELIZABETH A. RUTLEDGE AND DAVID W. RUSSELL*

Department of Medicine and Markey Molecular Medicine Center, University of Washington, Seattle, Washington 98195

Received 29 April 1997/Accepted 31 July 1997

Vectors derived from adeno-associated virus (AAV) have the potential to stably transduce mammalian cells by integrating into host chromosomes. Despite active research on the use of AAV vectors for gene therapy, the structure of integrated vector proviruses has not previously been analyzed at the DNA sequence level. Studies on the integration of wild-type AAV have identified a common site-specific integration locus on human chromosome 19; however, most AAV vectors do not appear to integrate at this locus. To improve our understanding of AAV vector integration, we analyzed the DNA sequences of several integrated vector proviruses. HeLa cells were transduced with an AAV shuttle vector, and integrated proviruses containing flanking human DNA were recovered as bacterial plasmids for further analysis. We found that AAV vectors integrated as single-copy proviruses at random chromosomal locations and that the flanking HeLa DNA at integration sites was not homologous to AAV or the site-specific integration locus of wild-type AAV. Recombination junctions were scattered throughout the vector terminal repeats with no apparent site specificity. None of the integrated vectors were fully intact. Vector proviruses with nearly intact terminal repeats were excised and amplified after infection with wild-type AAV and adenovirus. Our results suggest that AAV vectors integrate by nonhomologous recombination after partial degradation of entering vector genomes. These findings have important implications for the mechanism of AAV vector integration and the use of these vectors in human gene therapy.

Adeno-associated virus (AAV) is a 4.7-kb single-stranded DNA virus that has been developed as a gene therapy vector (31). Only the terminal repeat (TR) sequences are required in *cis* for replication and packaging, allowing a complete replacement of viral coding sequences with foreign DNA in vectors. A major advantage of AAV vectors is their ability to stably transduce cells by integration into host chromosomes. Although integration is crucial for many gene therapy applications, the mechanism of AAV vector integration is poorly understood, and the structure of integrated proviruses has not been determined at the DNA sequence level.

Southern analysis of DNA from transduced cells maintained under selection for the vector transgene has usually demonstrated the presence of integrated AAV vector proviruses (19, 27, 30, 36, 42). However, there have also been reports of episomal vector molecules, especially with *rep*$^+$ vectors (30) or the absence of selectable markers (1, 4). Some reports of vector integration noted a predominance of concatemeric proviruses (30, 36), while others did not (11, 27, 33, 44). The reasons for these variable results are unclear and include differences in transduction protocols as well as possible effects of contaminating wild-type AAV functions.

Two-thirds of integrated wild-type AAV proviruses are found at a specific human chromosome 19 site, 19q13-qter (24, 25, 38). While this feature could prove useful in some gene therapy applications, AAV vectors have not been found to integrate at this same locus (33, 44). Site-specific AAV integration appears to be mediated by the viral Rep protein (16, 28), so the absence of the *rep* gene in most vectors can explain the lack of site-specific integration. The presence of random-sized junction fragments detected by Southern analysis of integrated proviruses suggests that vector integration sites are random (27, 30, 33, 44). However, it is possible that vector integration occurs at scattered locations within a common locus different from 19q13-qter, just as wild-type AAV integrants are found scattered throughout the chromosome 19 site-specific integration locus (23, 38).

To better understand the process of AAV vector integration, we produced an AAV shuttle vector that allowed us to recover integrated vector proviruses along with flanking human DNA as bacterial plasmids. Several integration junctions from independent, transduced HeLa cell clones were sequenced and compared. Our results show that vector integration occurs at random chromosomal sites and that none of the vector proviruses integrated as intact, full-length genomes.

MATERIALS AND METHODS

Cell culture. Human HeLa (39) and 293 cells (17) were cultured in Dulbecco's modified Eagle medium with 10% heat-inactivated (56°C for 30 min) fetal bovine serum (HyClone, Logan, Utah), amphotericin (1.25 μg/ml), penicillin (100 U/ml), and streptomycin (100 μg/ml) at 37°C in a 10% CO_2 atmosphere. Titers of vector stocks were determined on HeLa cells by selecting for G418 (GIBCO-BRL, Grand Island, N.Y.) resistance as described previously (33) except that the G418 concentration was 1 mg of active compound per ml. Transduction of HeLa cells by AAV-SNori was carried out as when titering for G418 resistance except that resistant colonies were picked and expanded to 2×10^7 cells for isolation of genomic DNA. The multiplicities of infection were 0.17 vector particles per cell for clones 2 and 6; 1.7 for clones 1, 3, 4, 10, and 11; and 170 for clones 5, 7 to 9, and 12 to 14 (see Table 1).

Preparation of virus stocks. AAV-SNori vector stocks were prepared as follows. 293 cells were plated at a density of 8×10^6 cells/dish in 12 dishes (15-cm diameter). The next day, each dish was infected with 1.2×10^8 PFU of adenovirus type 5 (ATCC VR-5; American Type Culture Collection, Rockville, Md.) and 2 hours later cotransfected with 8 μg of pASNori2 and 32 μg of pAAV/Ad (36) by the calcium phosphate method (35). After 3 days, the cells and medium were harvested and combined, subjected to three cycles of freeze-thaw lysis in a dry ice-ethanol bath, clarified by centrifugation at 5,800 × *g* (5,500 rpm) in a Sorvall HS4 rotor for 30 min at 4°C, digested with micrococcal nuclease (68 U/ml; Pharmacia, Piscataway, N.J.) at 37°C for 1 h, treated with trypsin (50 ng/ml) at 37°C for 30 min, and centrifuged through 40% sucrose in phosphate-buffered saline in a Beckman SW28 rotor at 27,000 rpm for 16 h at 4°C. The pellets were resuspended in 8 ml of a 0.51-g/ml solution of CsCl and passed twice through a 22-gauge needle. The suspension was centrifuged in a Beckman SW41 rotor at 37,000 rpm for 20 h at 4°C. The region of the gradient containing AAV virions was collected, dialyzed against Dulbecco's modified Eagle medium through a 50,000-molecular-weight-cutoff membrane (Spectrum, Houston, Tex.), and concentrated by centrifugation in Centricon 100 filters (Amicon, Inc., Beverly, Mass.). Adenovirus was inactivated by treatment at 56°C for 1 h. The final

* Corresponding author. Mailing address: Department of Medicine, Box 357720, University of Washington, Seattle, WA 98195. Phone: (206) 616-4562. Fax: (206) 616-8298. E-mail: drussell@u.washington.edu.

stock contained 6.8×10^6 genomes per μl as determined by Southern analysis (33). The level of contaminating wild-type AAV was $<2.3 \times 10^4$ genomes per μl as determined by Southern analysis using AAV coding sequences as a probe.

Plasmid constructions. To construct pASNori2 (see Fig. 1), a *Bam*HI-*Esp*3I fragment of pSV2neo (41) containing an *Ssp*I-*Bst*1107I origin fragment from pACYC184 (7) in the *Bst*BI site (end filled with the Klenow fragment of DNA polymerase I) downstream of the *neo* (neomycin phosphotransferase) gene was inserted in the *Bgl*II sites of the AAV vector backbone of pTR (34) after attachment of *Bam*HI linkers to the pSV2neo *Esp*3I site. The pACYC184 fragment contains the p15A bacterial plasmid origin (10), with the direction of leading-strand DNA synthesis opposite that of *neo* gene transcription. pASNori2 also contains the pMB1 plasmid origin (5) from pBR322 (6). pASNori1, a deletion derivative of pASNori2 lacking the pMB1 origin, was constructed by end filling and circularizing a *Bsa*I-*Bst*1107I fragment of pASNori2.

Isolation of HeLa genomic DNA. Two confluent 10-cm-diameter dishes of each HeLa clone (2×10^7 cells) were lysed in genomic DNA lysis buffer (10 mM Tris [pH 8], 1 mM EDTA, 200 mM NaCl, 0.5% sodium dodecyl sulfate, 200 μg of proteinase K per ml) at 37°C overnight. The samples were then extracted with phenol and chloroform, extracted with butanol, and precipitated with 2 volumes of ethanol overnight at −20°C. The DNA was pelleted at 6,000 rpm in a Sorvall HS-4 rotor at 4°C for 25 min, washed with 70% ethanol, and air dried briefly. The DNA was resuspended in 500 μl of TE (10 mM Tris [pH 8], 1 mM EDTA), digested with 10 μg of RNase A (Sigma, St. Louis, Mo.) at 37°C for 3 h, extracted with phenol and chloroform, and precipitated with 50 μl of 3 M sodium acetate and 1 ml of ethanol at −20°C. Each pellet was resuspended in 200 μl of TE.

Provirus recovery in bacteria. To recover integrated AAV-SNori proviruses, 10 μg of HeLa genomic DNA was treated with 20 U of calf intestinal phosphatase (Boehringer Mannheim, Indianapolis, Ind.) to prevent ligation of free ends in the sample, heat inactivated at 65°C for 1 h, extracted with phenol and chloroform, and precipitated with ethanol. The resuspended DNA was digested with 20 U of *Eco*RI, which does not cut in the SNori vector, at 37°C for 4 h, heat inactivated at 65°C for 30 min, extracted with phenol and chloroform, and precipitated with ethanol. The resulting DNA fragments were resuspended and circularized with 200 U of T4 DNA ligase in 400 μl at 14°C overnight. The DNA was precipitated and one-fifth of the sample (approximately 2 μg) was electroporated into supercompetent *Escherichia coli* XL1Blue MRF′ cells (Stratagene, La Jolla, Calif.).

Provirus excision and amplification. HeLa cells and each of the nine HeLa clones from which plasmid had been recovered (see Table 1) were plated at 10^6 cells per 35-mm-diameter dish (Corning, Corning, N.Y.). The next day, cells were infected with wild-type AAV type 2 at 10 replication-competent particles/cell and adenovirus type 5 at 10 PFU/cell or were left uninfected as indicated in Fig. 6. Forty-four hours after infection, episomal DNA was isolated by the method of Hirt (21), with an additional proteinase K digestion, extraction with phenol and chloroform, and precipitation with ethanol. One-fourth of the episomal DNA from each dish was separated by alkaline agarose gel electrophoresis and transferred to Hybond-N+ membranes (Amersham, Arlington Heights, Ill.) according to standard procedures (35). Standards were prepared by *Bsm*I digest of pASNori2 to produce a 2.6-kb fragment. DNA was detected by Southern analysis using a *neo* gene probe.

DNA techniques. Restriction enzymes, T4 DNA ligase, and DNA polymerases were from New England BioLabs, Beverly, Mass. Proteinase K was from Boehringer Mannheim. Enzyme reactions were performed by using the manufacturer's recommended conditions. DNA manipulation and Southern blot analysis were performed by standard procedures (35). Southern blots were quantitated using a PhosphorImager 400S (Molecular Dynamics, Sunnyvale, Calif.). Plasmids were prepared by using Qiagen (Chatsworth, Calif.) columns. Dye terminator cycle sequencing was carried out with an AmpliTaq FS polymerase sequencing kit (Perkin-Elmer, Foster City, Calif.) and analyzed on an Applied Biosystems Inc. (Foster City, Calif.) sequencer. Oligonucleotides for sequencing are P1 (5′-dTACAAATAAAGCAATAGCATCAC-3′) and P2 (5′-dCCTCTGACAC ATGCAGCTC-3′). Sequences were analyzed by the Wisconsin version of the Genetics Computer Group program, using FASTA with default parameters and searching against GenBank/EMBL. HeLa flanking sequences were compared with the human *Alu* repetitive sequence humalurp7 (GenBank accession no. M57427), using FASTA with default parameters; a sequence with more than 70% identity over at least 100 nucleotides was considered a match. The probe used for detecting vector genomes by Southern analysis contained internal AAV-SNori sequences including the *neo* gene. The probe used for detecting site-specific integration was a 1.7-kb *Eco*RI-*Bam*HI fragment from the chromosome 19 integration locus cloned in pRE2 (38).

RESULTS

Transduction of HeLa cells by AAV-SNori. We designed an AAV shuttle vector (AAV-SNori) that could be recovered as a bacterial plasmid along with flanking human DNA after integration into host chromosomes. The recovered proviral plasmid could then be propagated in bacteria, and the junction fragments could be sequenced. The AAV-SNori vector contains the *neo* gene under the control of both the simian virus 40 (SV40) early promoter and the transposon 5 (Tn*5*) promoter for expression in human or bacterial cells, as well as the p15A bacterial replication origin packaged between AAV2 TRs (Fig. 1). These elements can support replication and confer kanamycin resistance in *E. coli*. The p15A origin was chosen because it replicates at the relatively low rate of 12 to 15 copies per *E. coli* chromosome (10), and we have found that plasmids containing the AAV TRs are more stable if they replicate at lower copy numbers (data not shown).

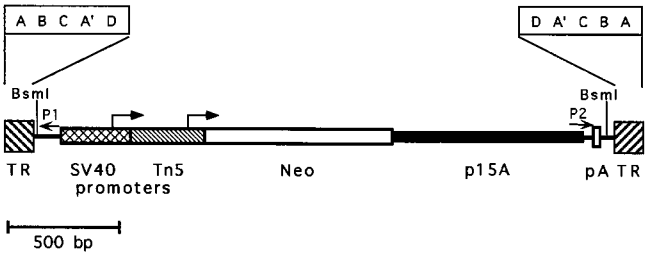

FIG. 1. AAV-SNori shuttle vector. The structure of the packaged genome of the AAV-SNori vector is shown; features include the neomycin phosphotransferase gene (Neo), the SV40 and Tn*5* promoters with their transcription start sites (arrows), the p15A replication origin, the AAV TRs, and the polyadenylation site (pA). An expanded diagram of the TR structure is shown to indicate the repeat domains. P1 and P2 represent binding sites for sequencing primers. The positions of *Bsm*I restriction sites are shown.

HeLa cervical carcinoma cells were transduced with purified AAV-SNori. One day after infection, the cells were treated with trypsin, and dilutions were plated in culture medium without selection. Selection was begun 24 h later with the antibiotic G418. Approximately 3×10^4 intact vector particles were required to produce a single, stable, G418-resistant colony. As all of these colonies contain integrated proviruses (see below), this represents the minimal vector integration frequency. The transient transduction rate was higher, as many of the colonies visible at earlier time points did not survive continued selection in G418, perhaps due to episomal vector gene expression. Fourteen G418-resistant colonies were isolated and expanded to approximately 2×10^7 cells under selection. Genomic DNA was isolated from each of 14 clones, digested with *Bsm*I or *Eco*RI, and analyzed by Southern blots probed with the *neo* gene (Fig. 2). *Bsm*I digests inside each of the TRs, so intact, integrated vector genomes should contain a 2.6-kb *Bsm*I fragment (Fig. 1). Figure 2A shows that 10 of the 14 clones had a fragment of the expected size. Clones 2, 6, 8, and 9 had deletions or rearrangements. The high-molecular-weight vector band in clone 8 DNA is faint, presumably due to a partial deletion of vector sequences. Clone 10 had a second, smaller fragment containing at least a portion of the *neo* gene. Quantitative analysis of the blots was consistent with single-copy integrated proviruses for all the clones except 8, 12, and 14, with the latter two clones containing four to five copies/cell.

Digestion of the transduced HeLa clones with *Eco*RI produced bands of various sizes ranging from 4 to 23 kb (Fig. 2B). As the AAV-SNori vector genome does not contain an *Eco*RI site, this finding is consistent with random integration. Although two *Bsm*I vector fragments were present in the DNA from clone 10 (Fig. 2A), only a single *Eco*RI fragment was detected, suggesting that both *Bsm*I fragments were located at a common integration site.

Provirus recovery as plasmids. Integrated vector proviruses were recovered from the transduced HeLa clones by digestion with *Eco*RI, circularization with DNA ligase, and transfer into *E. coli*. In this strategy, the *Eco*RI fragments containing intact

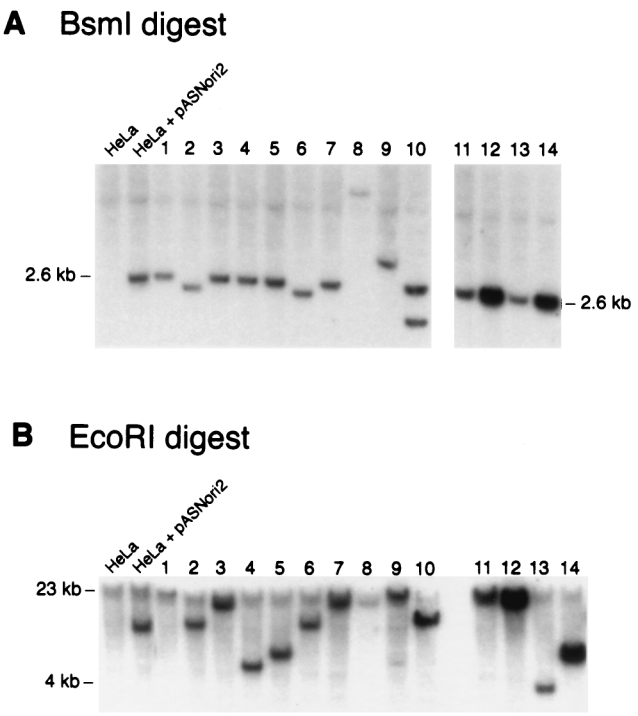

FIG. 2. Southern blots of transduced HeLa clones. Ten micrograms of genomic DNA from each HeLa clone (lanes 1 through 14) was digested with either *Bsm*I (A) or *Eco*RI (B), electrophoresed through a 1.2% agarose gel, transferred to a nylon membrane, and probed with internal vector sequences to detect integrated AAV-SNori proviruses. Lane HeLa, 10 μg of nontransduced HeLa DNA; lane HeLa + pASNori2, 10 μg of nontransduced HeLa DNA with 10 pg of pASNori2 plasmid DNA. This image was prepared by using Adobe Photoshop (Mountain View, Calif.) software.

vector proviruses also contain flanking chromosomal DNA, and their expected sizes can be predicted from the Southern analysis in Fig. 2B. Ligation at low DNA concentration enhances circularization of individual *Eco*RI fragments. The circularized fragments containing vector proviruses are then propagated as bacterial plasmids after electroporation of bacteria and selection with kanamycin. We recovered proviral plasmids from 9 of 14 HeLa transductants. Plasmids could not be recovered from the five other HeLa clones, even after repeated transformations (Table 1). This could be due to deletion or rearrangement of a region essential for plasmid function in bacteria but not required for G418 resistance in mammalian cells, such as the p15A origin or Tn*5* promoter. Four of the five HeLa clones containing proviruses that could not be recovered had an altered internal vector *Bsm*I fragment by Southern analysis (Fig. 2A), consistent with vector rearrangements.

Recovered plasmids were isolated from several of the kanamycin-resistant bacterial colonies and digested with *Eco*RI (Table 1). Most of these contained a single *Eco*RI site and were the size expected from Southern analysis (Fig. 2B). Transformation efficiencies varied considerably and did not appear to correlate with plasmid size. Only those plasmids with a single *Eco*RI fragment of the predicted size were considered correct. Of the plasmids that did not appear correct, 12 of 13 contained one or more extra *Eco*RI fragments in addition to the predicted fragment, which were presumably acquired by intermolecular ligation. Multiple proviral *Eco*RI fragments recovered from each cell line had the same restriction pattern, suggesting that major DNA rearrangements had not occurred

TABLE 1. Recovered provirus plasmids[a]

HeLa clone	Size of *Bsm*I fragment (kb)	Size of *Eco*RI fragment (kb)	Transformation efficiency/μg of DNA	No. of plasmids analyzed	No. of plasmids correct
1	2.6	19	<0.17	0	0
2	2.3	8	<0.17	0	0
3	2.6	16	2.5	5	5
4	2.6	4.7	101	5	5
5	2.6	5.5	68	5	4
6	2.3	7	<0.17	0	0
7	2.6	17	18	5	3
8	6.5	11	<0.17	0	0
9	3.3	16	<0.12	0	0
10	2.6, 1.9	9	2	4	2
11	2.6	20	9.5	5	5
12	2.6	18	4	8	1
13	2.6	4	1.5	3	3
14	2.6	5.5	4.5	5	4

[a] Integrated proviruses were recovered from transduced HeLa clones and transformed into bacteria. The sizes of fragments from Southern blots (Fig. 2), the transformation efficiency, and the number of plasmids that had a single *Eco*RI fragment (no. of plasmids correct) are listed for each HeLa clone.

during the transfer to bacteria. Further restriction digests confirmed that each recovered plasmid contained a single vector provirus copy (data not shown).

Because the TR sequences might be unstable during provirus recovery and replication in bacteria, we performed control transformations to ensure that our assay could recover plasmids with intact TRs. Plasmid pASNori1 contains the entire AAV-SNori vector genome including both TRs and depends on the p15A replication origin in the vector sequences for replication, making it similar in structure to the recovered plasmids. pASNori1 was linearized with *Eco*RI, gel purified, mixed with nontransduced, *Eco*RI-digested HeLa DNA, and subjected to the same recovery procedure as integrated proviruses. Of 10 recovered plasmids, 8 contained the expected unique *Eco*RI site. Two recovered plasmids contained extra *Eco*RI fragments in addition to the expected pASNori1 fragment. The restriction digest patterns of all eight correct plasmids were identical to the original pASNori1 plasmid after digestion with four restriction enzymes that have sites in the TRs, confirming that the TRs remained intact during our recovery procedure (data not shown).

Sequence analysis of integration junctions and flanking genomic DNA. One example of each correct recovered plasmid was sequenced by using primers P1 and P2 (Fig. 1). These primers bind to locations in the vector genome about 70 nucleotides inside each of the TRs. The sequenced regions begin at internal vector sequences and then proceed outward through the TR region, the integration junction site, and flanking human DNA. Of the nine plasmids analyzed, flanking human sequence information of 400 or more nucleotides was obtained from 14 of the 18 possible integration junctions. Three other junctions (12L, 12R, and 14L) contained nearly complete TRs that stalled the sequencing polymerase and one (10R) contained TR sequences joined to Tn*5* sequences (see below). The sequences around each junction site are shown in Fig. 3 and are designated as from the left TR (L) or right TR (R) of each recovered provirus. Figure 3A displays junction sequences that could be aligned in the "flip" orientation; Fig. 3B shows a right junction sequence that was in the "flop" orientation. In the flop orientation, the B and C regions are inverted compared to the flip orientation (D-A′-B-C-A instead of D-A′-C-B-A) (29). Figure 3C shows a left junction sequence

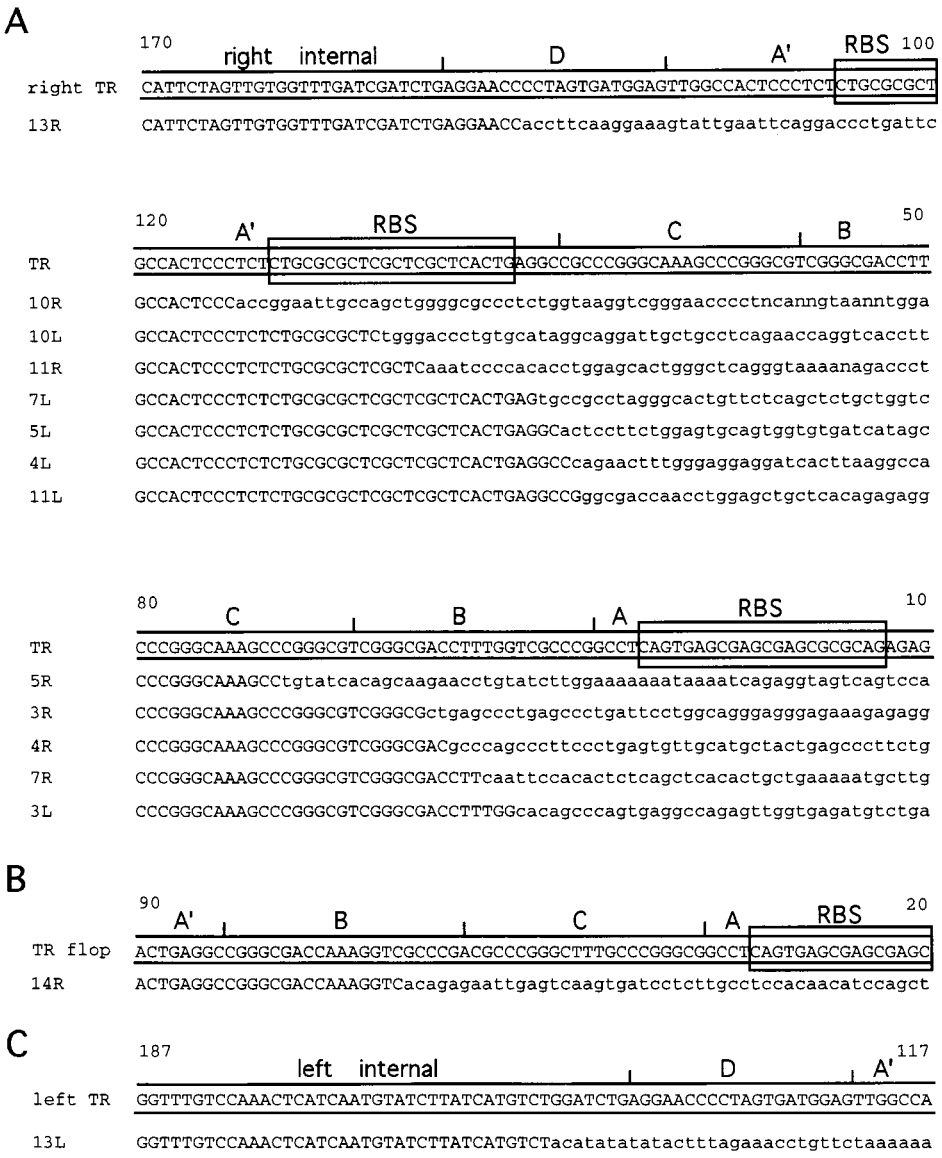

FIG. 3. AAV vector integration junctions. Portions of AAV vector sequences are shown in uppercase. Up to 62 nucleotides of flanking sequences are shown for each junction in lowercase and are aligned beneath the intact vector sequence (TR). Nucleotide positions are numbered beginning at the end of the intact TR. The sequences are designated as from the left (L) or right (R) of each numbered provirus clone. The Rep binding site (RBS) (9, 34) is boxed. Subdomains of the TRs (A, B, C, A′, and D) are indicated above the sequences. The flanking sequence is HeLa DNA, except for junction 10R, where it is Tn5 DNA (see text). (A) Sequences aligned to the right TR in the flip orientation and internal vector DNA. (B) Sequence aligned to the flop orientation of the TR. (C) Sequence aligned to the left internal portion of the vector.

with the recombination site internal to the TR. Of the six junctions containing part of the B or C region in which the orientation could be determined, five were in the flip orientation.

The sequence data show that no two proviruses had the same junction site. Junctions were dispersed throughout the TRs and internal vector sequences. We could not identify a common sequence motif in vector or chromosomal DNA at the junction sites. None of the flanking sequences were similar to each other, and there were no matches with any gene in the GenBank or EMBL database. The only known sequence element found was a single *Alu* repeat in the flanking human DNA of junction 4R.

Comparison of the flanking sequences with two sequences from the chromosome 19 integration locus, AAVS1 (GenBank accession no. S51329) and pRE2 (provided by R. J. Samulski), produced no matches, indicating that the clones did not integrate into the site-specific integration locus of wild-type AAV2. This observation was confirmed by probing Southern blots of HeLa clone DNAs with a portion of the site-specific integration locus, which showed no rearrangements at this locus in any of the 14 transductants analyzed (data not shown). In addition, analysis of the flanking sequences revealed no binding sites for the AAV Rep protein (9, 34), in contrast to the chromosomal site-specific integration locus (16).

Two of the junctions (4L and 5R) had structures that were the result of rearrangements. A schematic representation of these proviruses is shown in Fig. 4. The rearrangements in sequences 4L and 5R were in the regions just inside each TR where a 44-nucleotide region is repeated in the left and right

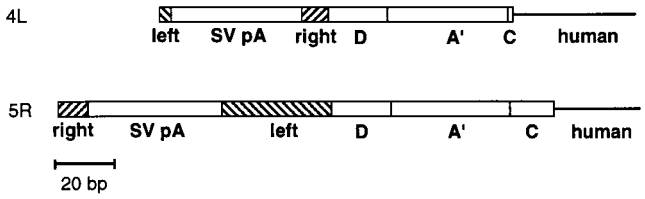

FIG. 4. Structures of rearranged sequences. The 4L and 5R provirus junction sequences are diagrammed to show where internal vector sequences recombined. The positions of left and right internal vector sequences, the common SV40 polyadenylation site found in either end (SV pA), the TR domains (D, A′, and C), and flanking human DNA are shown.

portions of the vector. This region contains the SV40 polyadenylation signal and was duplicated as a result of vector construction. In the case of junction 4L, sequencing from the left P1 primer (Fig. 1) read a portion of left internal sequences, the common SV40 sequence, and then right internal vector sequences followed by a portion of the right TR and flanking chromosomal DNA. A similar finding was observed in junction 5R reading from the right P2 primer. Apparently homologous recombination occurred between these sites and produced an inverted internal portion of the vector. Recombination could have happened during the production of the AAV-SNori stocks or the transduction of HeLa cells.

Junction (10R) was not joined to human DNA. This junction contained the A′ region of the right TR joined to a second copy of vector Tn*5* DNA. The sequence around the junction site is shown in Fig. 3. This explains the two *neo*-hybridizing bands detected in Fig. 2A for HeLa clone 10. Flanking HeLa DNA sequence was not obtained for this junction and must lie beyond the region sequenced.

When the TR is drawn in its predicted secondary structure, the recombination sites of the integrated proviruses are scattered throughout the repeat (Fig. 5). Except for some potential clustering of recombination events between the C and A′ repeat domains, our data set does not suggest that secondary structure in the TRs directs a specific integration process. Additional junctions will need to be sequenced to determine if the four junctions observed between C and A′ represent a recombination hotspot.

Flanking DNA sequence was not obtained for the remaining 3 of the 18 possible integration junctions recovered (12L, 12R, and 14L). The sequences obtained for these junctions ended abruptly in the A′ region of the TR, where the sequencing polymerase apparently stalled. Based on our sequencing of intact AAV TRs, this is consistently observed when the secondary structure of the inverted repeat remains intact, and the polymerase stalls at the first base that can pair in the secondary structure (data not shown). If we assume that this also occurred while we were sequencing these three junctions, then we can place the end of the vector provirus DNA at the last base of the predicted secondary structure that would have stalled the sequencing polymerase, in the corresponding positions of the A domain. These junctions are indicated with asterisks in Fig. 5.

Provirus rescue by wild-type infection. Integrated wild-type AAV proviruses can be excised and amplified by infection with adenovirus (3, 8, 18, 26), and a similar rescue of vector proviruses can occur by infection with wild-type AAV and adenovirus (19, 30, 36, 42). We screened each of the nine transduced HeLa clones that contained proviruses recovered in bacteria to see if their TR structures would correlate with the potential for excision and amplification. Only clones 12 and 14 contained proviruses that could be rescued by coinfection with wild-type AAV and adenovirus, and none of the clones could be rescued by adenovirus alone, as expected if no *rep* gene was present in the cells. Figure 6 shows representative results from this experiment for clones 12, 13, and 14. The two rescuable proviruses had the most intact TRs of all the proviruses that we analyzed (Fig. 5), and the nearly complete repeats on both ends of clone 12 correlated with more efficient rescue than those of clone 14.

DISCUSSION

We have developed a shuttle vector system to study the structure of integrated AAV vector proviruses and analyze their flanking DNA. Fourteen independent, G418-resistant HeLa clones were isolated after transduction with an AAV vector containing the *neo* gene, and integrated vector proviruses were recovered as bacterial plasmids from nine of these lines. Sequence analysis of the proviral junction sites showed that each integration event occurred at a different chromosomal site, and none of these sites were at the chromosome 19

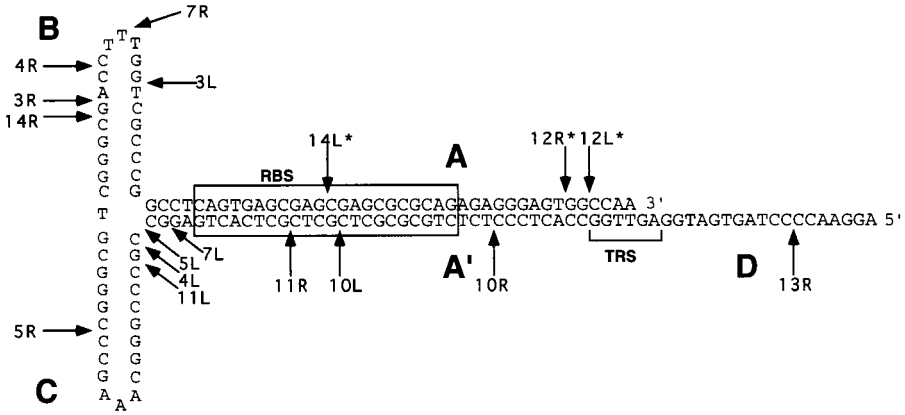

FIG. 5. Junction sites and TR secondary structure. Junction recombination sites within the TRs of recovered proviruses are shown in relation to the secondary structure of the TR. All sequences were adapted to be displayed in the flip orientation. Junctions 12L, 12R, and 14L are presumed junction sites because the actual sequences ended in the corresponding positions in the A′ region, suggesting that the TR was intact up to that point. The positions of the Rep binding site (RBS) (9, 34) and terminal resolution site (TRS) (40) are shown.

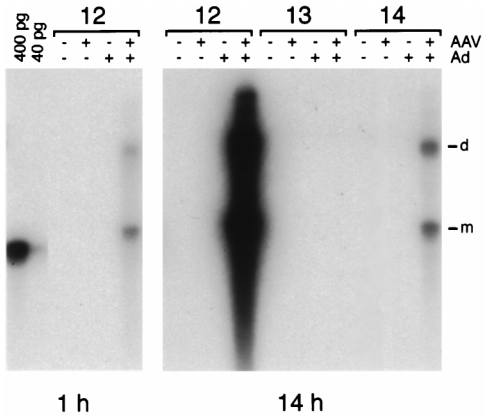

FIG. 6. Provirus rescue. Transduced HeLa clones were infected with wild-type AAV2 (AAV), and/or adenovirus (Ad) or left uninfected as indicated. Episomal DNA was isolated from the cells 44 h later, electrophoresed through an alkaline agarose gel, and subjected to Southern analysis using a *neo* gene probe. The positions of the monomer (m) and dimer replicative (d) vector forms are indicated. The standards are 400 and 40 pg of the 2.6-kb *Bsm*I fragment of pASNori2. The results for HeLa clones 12 to 14 are shown (14-h exposure), with HeLa clone 12 also shown at a shorter exposure (1 h). This image was prepared by using Adobe Photoshop software.

site-specific integration locus of wild-type AAV. Each recombination event was at a different position in the vector genome, and no fully intact vector proviruses were recovered. There was no sequence homology between the vector and flanking human DNA and none between the different flanking sequences recovered. These results demonstrate that AAV vector integration occurs at random chromosomal locations and that each integrated provirus contains variable amounts of the complete vector genome.

The recovered proviruses were joined to human DNA sequences at nucleotide positions throughout the vector TRs and at internal vector sequences in one case. Because our system required the presence of functional promoters, *neo* gene, and bacterial origin, the ends of the vector were the only portions that could have been disrupted by integration and still produced a functional provirus capable of being recovered. Southern analysis of the proviruses that could not be recovered showed that four of five contained rearrangements within internal vector sequences, suggesting that the junctions occurred inside the *Bsm*I sites (Fig. 2A). While our results are consistent with random recombination throughout the vector genome, we cannot exclude a partial preference for TR recombination sites, since more junctions should have been observed in the approximately 170 nonessential internal vector nucleotides if a completely random representation had been recovered.

Based on Southern analysis and restriction mapping of recovered plasmids, 11 of 14 HeLa transductants contained one integrated vector genome per cell, and clone 10 contained a partially duplicated vector genome. Two clones (12 and 14) had a stronger hybridization signal consistent with four to five vector copies per cell. In these clones, the vector signal was present in unique bands after digestion outside of vector DNA with *Eco*RI, and their recovered plasmids contained one copy of the vector provirus within the *Eco*RI fragment. This finding suggests that amplification of the provirus and flanking DNA contained in the *Eco*RI fragment occurred after integration of the vector, perhaps during the selection process in G418. It is possible that proviral sequences were responsible for this amplification process, as the AAV TR has been shown to function as a replication origin in carcinogen-treated mammalian cells (46), and these two proviruses contained the most complete TR structures of all those recovered.

Both concatemeric and single-copy vector proviruses have been described in prior studies of vector integration based on Southern analysis of transduced, cultured cells (11, 27, 30, 33, 36, 44). More recently, animal studies have demonstrated the presence of concatemeric vector sequences in high-molecular-weight DNA by PCR after in vivo vector administration (14, 20, 45). The conditions favoring concatemeric versus single-copy integration are unknown and could include cell-type-specific functions, multiplicity of infection, contaminating rep^+ helper virus particles, and differences in selection conditions. None of the G418-resistant HeLa cell transductants that we analyzed contained concatemeric vector proviruses, over a range of infection multiplicities of 0.17 to 170 vector particles/cell, with $\leq$0.3% wild-type AAV contamination (wild-type genomes per vector genome). Our results are in agreement with previous studies suggesting that concatemeric vector proviruses are preferentially rescued by coinfection with wild-type AAV and adenovirus (30, 36), as most of the integrants that we analyzed could not be rescued. Although the two transductants that could be rescued were not present as concatemers joined at the TRs, they were present at multiple copies per cell. These proviruses also had the smallest deletions in their TRs, with at least one repeat containing a complete copy of each repeat domain sequence. This proviral structure would be expected to regenerate a fully intact TR by gene conversion as previously shown for wild-type AAV (37).

There were both differences and similarities between vector integration in our system and integration by wild-type AAV2. Unlike the random vector integration that we observed, two-thirds of wild-type proviruses studied in latently infected human cells were found in the site-specific integration locus on chromosome 19 (25, 38). Although these wild-type integration events occurred at a common locus, the junctions were at different nucleotide positions within the locus and at different positions in the AAV TRs (23, 38). This heterogeneity in junction sites within the TRs is similar to our findings with vector integration. Site-specific integration into chromosome 19 requires the presence of a Rep binding site in the chromosomal target DNA (16, 28), and no Rep binding sites were identified in the human flanking sequences that we recovered. This finding suggests that vector integration is not mediated by the Rep protein, including Rep protein molecules that might be contained in the entering vector particle. Presumably, either Rep expression in the transduced cell is necessary for site-specific integration or wild-type virions contain functional Rep molecules that are absent from vector virions. A related observation may be that integrated wild-type proviruses are often found as concatemers (8, 22, 26), which we did not observe for vector proviruses. This observation supports the hypothesis that site-specific integration involves Rep-dependent replication of the chromosome and virus during integration (28, 43), while Rep-independent vector integration occurs by another, nonreplicative mechanism.

Our data suggest that AAV vector integration is a nonhomologous recombination reaction. While no clear recombination hotspots were identified in the vector genome, there may have been a clustering of junctions between the C and A′ regions of the TR, suggesting a possible role for Holliday-like recombination intermediates in the reaction. Although the exact mechanism has not been determined, the recombination event is presumably mediated by cellular proteins. Cleavage of the chromosomal preintegration site may be due to host nucleases or DNA damage. The requirement for chromosomal breaks could in part explain why DNA-damaging agents in-

crease transduction by AAV vectors (2, 32). The lack of specific vector recombination enzymes also suggests why vector integration is such an inefficient process, requiring hundreds to thousands of vector particles per integration event (15, 19, 33, 36). One explanation for the consistent recovery of incomplete proviruses is that the entering, linear vector genomes are partially degraded before the integration event. Despite our determination of the structure of integrated proviruses, key steps in the recombination process have yet to be elucidated, including the role of vector second-strand synthesis, which is also associated with increased transduction rates (12, 13), and whether the recombination reaction involves single-stranded or double-stranded vector molecules. The nature of the chromosomal preintegration site is completely unknown, and there could be large deletions or even interchromosomal crossover events associated with vector integration.

The uncertain nature of the AAV vector integration process has important consequences for gene therapy. In contrast to retroviral vectors, which integrate in a precise and predictable reaction mediated by the viral integrase protein, each AAV vector integration event is likely to produce a different, incomplete provirus, with unknown effects on the host chromosomes involved. Many of the integrated vectors could contain deletions of the transgene being delivered, with resulting decreases in transduction efficiencies. As our understanding of the process improves, methods for increasing the efficiency and predictability of vector integration that improve the prospects for gene therapy by AAV vectors may be developed.

ACKNOWLEDGMENTS

We thank Roli Hirata and Jaclynn Mac for expert technical assistance.

This work was supported by grants from the Cystic Fibrosis Foundation, the American Society of Hematology, and the National Institutes of Health (P01 HL53750 and K08 HL03100).

REFERENCES

1. **Afione, S. A., C. K. Conrad, W. G. Kearns, S. Chunduru, R. Adams, T. C. Reynolds, W. B. Guggino, G. R. Cutting, B. J. Carter, and T. R. Flotte.** 1996. In vivo model of adeno-associated virus vector persistence and rescue. J. Virol. **70:**3235–3241.
2. **Alexander, I. E., D. W. Russell, and A. D. Miller.** 1994. DNA-damaging agents greatly increase the transduction of nondividing cells by adeno-associated virus vectors. J. Virol. **68:**8282–8287.
3. **Berns, K. I., T. C. Pinkerton, G. F. Thomas, and M. D. Hoggan.** 1975. Detection of adeno-associated virus (AAV)-specific nucleotide sequences in DNA isolated from latently infected Detroit 6 cells. Virology **68:**556–560.
4. **Bertran, J., J. L. Miller, Y. Yang, A. Fenimore-Justman, F. Rueda, E. F. Vanin, and A. W. Nienhuis.** 1996. Recombinant adeno-associated virus-mediated high-efficiency, transient expression of the murine cationic amino acid transporter (ecotropic retroviral receptor) permits stable transduction of human HeLa cells by ecotropic retroviral vectors. J. Virol. **70:**6759–6766.
5. **Betlach, M., V. Hershfield, L. Chow, W. Brown, H. Goodman, and H. W. Boyer.** 1976. A restriction endonuclease analysis of the bacterial plasmid controlling the EcoRI restriction and modification of DNA. Fed. Proc. **35:**2037–2043.
6. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
7. **Chang, A. C., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:**1141–1156.
8. **Cheung, A. K., M. D. Hoggan, W. W. Hauswirth, and K. I. Berns.** 1980. Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J. Virol. **33:**739–748.
9. **Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. Kyöstiö, R. M. Kotin, and B. Safer.** 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. **68:**7448–7457.
10. **Cozzarelli, N. R., R. B. Kelly, and A. Kornberg.** 1968. A minute circular DNA from Escherichia coli 15. Proc. Natl. Acad. Sci. USA **60:**992–999.
11. **Einerhand, M. P., M. Antoniou, S. Zolotukhin, N. Muzyczka, K. I. Berns, F. Grosveld, and D. Valerio.** 1995. Regulated high-level human beta-globin gene expression in erythroid cells following recombinant adeno-associated virus-mediated gene transfer. Gene Ther. **2:**336–343.
12. **Ferrari, F. K., T. Samulski, T. Shenk, and R. J. Samulski.** 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. **70:**3227–3234.
13. **Fisher, K. J., G. P. Gao, M. D. Weitzman, R. DeMatteo, J. F. Burda, and J. M. Wilson.** 1996. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J. Virol. **70:**520–532.
14. **Fisher, K. J., K. Jooss, J. Alston, Y. Yang, S. E. Haecker, K. High, R. Pathak, S. E. Raper, and J. M. Wilson.** 1997. Recombinant adeno-associated virus for muscle directed gene therapy. Nat. Med. **3:**306–312.
15. **Flotte, T. R., R. Solow, R. A. Owens, S. Afione, P. L. Zeitlin, and B. J. Carter.** 1992. Gene expression from adeno-associated virus vectors in airway epithelial cells. Am. J. Respir. Cell. Mol. Biol. **7:**349–356.
16. **Giraud, C., E. Winocour, and K. I. Berns.** 1995. Recombinant junctions formed by site-specific integration of adeno-associated virus into an episome. J. Virol. **69:**6917–6924.
17. **Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. **36:**59–74.
18. **Handa, H., K. Shiroki, and H. Shimojo.** 1977. Establishment and characterization of KB cell lines latently infected with adeno-associated virus type 1. Virology **82:**84–92.
19. **Hermonat, P. L., and N. Muzyczka.** 1984. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA **81:**6466–6470.
20. **Herzog, R. W., J. N. Hagstrom, S. H. Kung, S. J. Tai, J. M. Wilson, K. J. Fisher, and K. A. High.** 1997. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. Proc. Natl. Acad. Sci. USA **94:**5804–5809.
21. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26:**365–369.
22. **Kotin, R. M., and K. I. Berns.** 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. Virology **170:**460–467.
23. **Kotin, R. M., R. M. Linden, and K. I. Berns.** 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. **11:**5071–5078.
24. **Kotin, R. M., J. C. Menninger, D. C. Ward, and K. I. Berns.** 1991. Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. Genomics **10:**831–834.
25. **Kotin, R. M., M. Siniscalco, R. J. Samulski, X. D. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns.** 1990. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA **87:**2211–2215.
26. **Laughlin, C. A., C. B. Cardellichio, and H. C. Coon.** 1986. Latent infection of KB cells with adeno-associated virus type 2. J. Virol. **60:**515–524.
27. **Lebkowski, J. S., M. M. McNally, T. B. Okarma, and L. B. Lerch.** 1988. Adeno-associated virus: a vector system for efficient introduction and integration of DNA into a variety of mammalian cell types. Mol. Cell. Biol. **8:**3988–3996.
28. **Linden, R. M., E. Winocour, and K. I. Berns.** 1996. The recombination signals for adeno-associated virus site-specific integration. Proc. Natl. Acad. Sci. USA **93:**7966–7972.
29. **Lusby, E., K. H. Fife, and K. I. Berns.** 1980. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. J. Virol. **34:**402–409.
30. **McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J. Virol. **62:**1963–1973.
31. **Muzyczka, N.** 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. Curr. Top. Microbiol. Immunol. **158:**97–129.
32. **Russell, D. W., I. E. Alexander, and A. D. Miller.** 1995. DNA synthesis and topoisomerase inhibitors increase transduction by adeno-associated virus vectors. Proc. Natl. Acad. Sci. USA **92:**5719–5723.
33. **Russell, D. W., A. D. Miller, and I. E. Alexander.** 1994. Adeno-associated virus vectors preferentially transduce cells in S phase. Proc. Natl. Acad. Sci. USA **91:**8915–8919.
34. **Ryan, J. H., S. Zolotukhin, and N. Muzyczka.** 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J. Virol. **70:**1542–1553.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
36. **Samulski, R. J., L. S. Chang, and T. Shenk.** 1989. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. J. Virol. **63:**3822–3828.
37. **Samulski, R. J., A. Srivastava, K. I. Berns, and N. Muzyczka.** 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. Cell **33:**135–143.
38. **Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter.** 1991. Targeted integration of adeno-associated virus

(AAV) into human chromosome 19. EMBO J. **10:**3941–3950. (Erratum, **3:**1228, 1992.)
39. **Scherer, W. F., J. T. Syverton, and G. O. Gey.** 1953. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. J. Exp. Med. **97:**695–709.
40. **Snyder, R. O., D. S. Im, T. Ni, X. Xiao, R. J. Samulski, and N. Muzyczka.** 1993. Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J. Virol. **67:**6096–6104.
41. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1:**327–341.
42. **Tratschin, J. D., I. L. Miller, M. G. Smith, and B. J. Carter.** 1985. Adeno-associated virus vector for high-frequency integration, expression, and rescue of genes in mammalian cells. Mol. Cell. Biol. **5:**3251–3260.
43. **Urcelay, E., P. Ward, S. M. Wiener, B. Safer, and R. M. Kotin.** 1995. Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. J Virol. **69:**2038–2046.
44. **Walsh, C. E., J. M. Liu, X. Xiao, N. S. Young, A. W. Nienhuis, and R. J. Samulski.** 1992. Regulated high level expression of a human gamma-globin gene introduced into erythroid cells by an adeno-associated virus vector. Proc. Natl. Acad. Sci. USA **89:**7257–7261.
45. **Xiao, X., J. Li, and R. J. Samulski.** 1996. Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. J. Virol. **70:**8098–8108.
46. **Yalkinoglu, A. O., H. Zentgraf, and U. Hubscher.** 1991. Origin of adeno-associated virus DNA replication is a target of carcinogen-inducible DNA amplification. J. Virol. **65:**3175–3184.