JOURNAL OF VIROLOGY, Dec. 1997, p. 9231–9247 Vol. 71, No. 12
0022-538X/97/$04.00+0

Cellular Recombination Pathways and Viral Terminal Repeat Hairpin Structures Are Sufficient for Adeno-Associated Virus Integration In Vivo and In Vitro

C. C. YANG,[1] X. XIAO,[1] X. ZHU,[1]† D. C. ANSARDI,[1]‡ N. D. EPSTEIN,[2] M. R. FREY,[3] A. G. MATERA,[3] AND R. J. SAMULSKI[1]*

Gene Therapy Center and Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7352[1]; National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20829[2]; and Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106[3]

Received 3 July 1997/Accepted 10 September 1997

The human parvovirus adeno-associated virus (AAV) is unique in its ability to target viral integration to a specific site on chromosome 19 (ch-19). Recombinant AAV (rAAV) vectors retain the ability to integrate but have apparently lost this ability to target. In this report, we characterize the terminal-repeat-mediated integration for wild-type (wt), rAAV, and in vitro systems to gain a better understanding of these differences. Cell lines latent for either wt or rAAV were characterized by a variety of techniques, including PCR, Southern hybridization, and fluorescence in situ hybridization analysis. More than 40 AAV-rAAV integration junctions were cloned, sequenced, and then subjected to comparison and analysis. In both immortalized and normal diploid human cells, wt AAV targeted integration to ch-19. Integrated provirus structures consisted of head-to-tail tandem arrays with the majority of the junction sequences involving the AAV inverted terminal repeats (ITRs). No complete viral ITRs were directly observed. In some examples, the AAV p5 promoter sequence was found to be fused at the virus-cell junction. Data from dot blot analysis of PCR products were consistent with the occurrence of inversions of genomic and/or viral DNA sequences at the wt integration site. Unlike wt provirus junctions, rAAV provirus junctions mapped to a subset of non-ch-19 sequences. Southern analysis supported the integration of proviruses from two independent cell lines at the same locus on ch-2. In addition, provirus terminal repeat sequences existed in both the flip and flop orientations, with microhomology evident at the junctions. In all cases with the exception of the ITRs, the vector integrated intact. rAAV junction sequence data were consistent with the occurrence of genomic rearrangement by deletion and/or rearrangement-translocation at the integration locus. Finally, junctions formed in an in vitro system between several AAV substrates and the ch-19 target site were isolated and characterized. Linear AAV substrates typically utilized the end of the virus DNA substrate as the point of integration, whereas products derived from AAV terminal repeat hairpin structures in the presence or absence of Rep protein resembled AAV–ch-19 junctions generated in vivo. These results describing wt AAV, rAAV, and in vitro integration junctions suggest that the viral integration event itself is mediated by terminal repeat hairpin structures via nonviral cellular recombination pathways, with specificity for ch-19 in vivo requiring additional viral components. These studies should have an important impact on the use of rAAV vectors in human gene therapy.

Adeno-associated virus (AAV) is a defective parvovirus that requires coinfection with a helper virus such as adenovirus (3) or herpes simplex virus (10) for efficient replication and production of progeny. In the absence of helper virus, AAV infection results in integration of the AAV genome into host genomic DNA. Subsequent superinfection of cells latently infected with AAV helper virus results in the rescue and replication of the AAV genome, thereby completing the virus life cycle (4). AAV integration appears to have no adverse effect on cell growth or morphology in vitro (18). Furthermore, latently infected cells established in cell culture appear to be stable and have maintained integrated viral DNA for >150 cell passages (5, 40). The efficient integration of AAV into the host cell genome to establish viral latency in the absence of helper virus possesses a feature that so far is unique among DNA viruses of eukaryotes. Analysis of genomic DNA isolated from latently infected cells indicated that AAV targets integration toward q13.4-ter of chromosome 19 (ch-19) (24–26, 40). Sequence analysis of several independent integration events showed that virus-host DNA junctions occurred within a 100-bp region (40). The chromosomal breakpoints isolated from wild-type (wt) latent cell lines differ from junctions established in a transient integration system, established by Berns and colleagues, using an Epstein-Barr virus (EBV) shuttle vector (15, 16) but are similar in that both require the same ch-19 locus. This targeted integration by wt AAV has been demonstrated in a variety of human cell lines, ranging from aneuploid 293, HeLa, Detroit 6, and IB3 cells to diploid WI-38, colon, and T cells (17, 20, 27, 33, 40).

The nonpathogenic nature of AAV, its site-specific targeting of integration in establishment of viral latency, and the ability of the virus to infect a wide range of cell lines and cell types suggest that AAV possesses many characteristics that make its

* Corresponding author. Mailing address: Gene Therapy Center, CB 7352, 7119 Thurston Bowles, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7352. Phone: (919) 962-3285. Fax: (919) 966-0907. E-mail: rjs@med.unc.edu.

† Present address: Department of Biological Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213.

‡ Present address: Department of Microbiology, University of Alabama—Birmingham, Birmingham, AL 35294.

recombinant derivatives attractive for use as gene transduction vectors in gene therapy. At this time, however, the structure of the AAV provirus in its latent state within ch-19, the characteristics of the wt virus that are retained by recombinant AAV (rAAV) vectors (in particular, the site-specific targeting of integration), and the actual DNA substrates used for viral integration are less clear. Recent studies using an EBV–ch-19 episomal integration assay have now established the minimal host *cis* sequence required for targeting (15, 16, 30, 31), thus providing essential information required to understand the targeted integration of wt AAV. In this work, we characterized (i) the provirus structure of wt AAV from various latent cell lines through PCR analysis and sequence determination of junction breakpoints; (ii) the integration of rAAV in its establishment of cellular latency through fluorescence in situ hybridization (FISH), Southern analysis, and bacterial cloning of provirus structures; and (iii) the in vitro integration of AAV DNA substrates into the ch-19 preintegration site. The results from the work described in this report indicate that head-to-tail structures are commonly observed for wt AAV proviruses with inversions of the genomic DNA and novel junctions involving the AAV p5 promoter region in the reverse orientation. Furthermore, we describe the first isolation and characterization of provirus clones derived from independent latent cell lines that contain virus-cell junction sequences. rAAV was observed to integrate at loci on chromosomes 1, 2, 7, and 17, with two independent isolates targeting the same locus of ch-2. All of these provirus structures exhibited microhomology to the chromosome target sequence and resulted in deletions of the inverted terminal repeat (ITR) sequences. Finally, the in vitro integration of various AAV DNA substrates into the ch-19 integration site was characterized both in the presence and absence of AAV Rep proteins. Analysis of the resulting AAV–ch-19 integration products that were cloned and sequenced revealed that while linear AAV substrates integrated into the target locus at the end of the DNA substrate, molecules that contained the AAV ITR hairpin structure generated junctions that were indistinguishable from wt AAV–ch-19 junctions characterized in vivo. The structures of junctions generated from wt, rAAV, and in vitro reactions in this study suggest that the AAV ITR hairpin is the required substrate for AAV integration and that host enzymes are sufficient for this integration event. While these results indicate that a common cellular pathway is utilized in the integration of rAAV and wt virus, the integration event itself does not appear to be a simple insertion of the viral genome into the host genome, and additional viral proteins (Rep) are required for targeting to ch-19. The understanding of the initial steps required for AAV and rAAV integration will have important implications for the use of this virus as a vector for human gene therapy.

MATERIALS AND METHODS

Cell lines and viruses. 293, HeLa, Detroit 6, and WI-38 cell lines were maintained on Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 100 μg of penicillin per ml, and 100 μg of streptomycin per ml in an environment maintained at 5% CO_2. AAV/OR1-OR2 and rAAV (DD-Neo) viral stocks were prepared by using the AAV plasmids pSMOR, pBS35 (38, 40), and pDD-Neo (47) (see below) according to methods previously described. Latently infected cell lines were established as previously described (40, 47). Subconfluent-monolayer tissue culture cells were infected with AAV/OR1-OR2 or DD-Neo viral stocks at a multiplicity of infection of ≥10. In the case of DD-Neo, the infected cells were placed under G418 selection 48 h postinfection. Individual drug-resistant colonies were isolated and expanded after 2 weeks of selection.

DNA cloning and sequencing. The AAV hybrid viral vectors pSMOR and pBS35 used to generate AAV/OR1-OR2 viral stocks have been described previously (40). The rAAV plasmid pDD-Neo used to generate recombinant viral vector stocks of DD-Neo was constructed by inserting the *Xho*I-*Sal*I neomycin resistance fragment from pMC1neo poly A (Stratagene) into the *Sal*I site of the rAAV plasmid pDD-2 (47). High-molecular-weight chromosomal DNA was extracted as previously described (35). PCR amplification products derived from cell lines latently infected with AAV were TA cloned into the plasmid pCR1000 according to the protocol provided by the supplier (Invitrogen). The cosmid pHC-RJ was kindly provided by N. D. Epstein.

Cloning of complete (or portions of) rAAV proviruses containing provirus junction sequences from genomic DNA was performed according to the following procedure. Ten to twenty micrograms of genomic DNA was digested to completion with *Bgl*II, *Hin*dIII, or *Ssp*I and recircularized with DNA ligase. DNA ligations were performed in a total volume of 1 ml in the presence of 12 Weiss units of T4 DNA ligase under conditions outlined by the supplier (New England BioLabs). A Bio-Rad Gene Pulser was used as described in the manufacturer's protocol for the transformation of electroporation-competent *Escherichia coli* (SURE strain; Stratagene) with the resulting recircularized DNA ligation products.

The integration target plasmid pRE2 used in the in vitro integration experiments was constructed by cloning the 2.7-kb *Bam*HI fragment containing the AAV–ch-19 integration site (40) into the *Bam*HI site of pGEM-3Z (Promega). "No-end" AAV and rAAV DNA substrates were prepared, as previously described (41), using the *Pvu*II fragment the AAV plasmid SSV9. Plasmid preparations and other routine DNA manipulations were performed by standard procedures (35). DNA was sequenced at the University of North Carolina Automated DNA Sequencing Facility on a model 373A DNA sequencer (Applied Biosystems) with a TaqDyedeoxy Terminator Cycle Sequencing Kit (Applied Biosystems).

PCR and DNA hybridization analysis. PCRs were carried out as described by the manufacturer (Perkin-Elmer Cetus) in a total reaction volume of 100 μl. The AAV primers H (5′-ACAAAGCTGTCAGAAATGCC-3′), T (5′-ATAAGTAG CATGGCGGGTTA-3′), and TR6 (5′-CCTCAGTGAGCGAGCGAGCG-3′) and the ch-19 primer C (5′-GCATAAGCCAGTAGAGCTCA-3′) were synthesized at the University of North Carolina Nucleic Acids Core Facility. PCRs proceeded for 30 cycles in a Perkin-Elmer GeneAmp 9600 thermal cycler programmed for 1 min at 94°C, 1 min at 58°C, and 1.5 min at 72°C. PCR products were analyzed on 1% agarose gels and then isolated from reaction mixtures by phenol extraction and ethanol precipitation by standard procedures (35).

Dot blot filters were prepared by spotting total genomic DNA (5 μg) or PCR products (4 μl of a 100-μl reaction volume) denatured in 0.4 N NaOH–0.6 M NaCl on Gene Screen Plus (New England Nuclear) nylon membranes soaked in the same solution. DNA panels containing high-molecular-weight chromosomal DNA were prepared by restriction enzyme digestion of total genomic DNA (10 to 20 μg) followed by separation on a 0.8% agarose gel and transferred to Gene Screen Plus nylon membranes as recommended by the manufacturer. The mouse-human somatic cell hybrid panels representing each individual human chromosome was obtained from Oncor (Gaithersburg, Md.). Random primed ^{32}P-labeled DNA probes were produced with a labeling kit and unlabeled probe DNA as described by the kit's supplier (Boehringer Mannheim, Indianapolis, Ind.). Southern hybridizations (42) were performed by the following hybridization and wash procedures. ^{32}P-labeled DNA probes were hybridized to DNA panels at 65°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–sodium dodecyl sulfate (SDS) overnight. Following hybridization, the membranes were washed for 30 min to 1 h in 2× SSC–1% SDS at 65°C and then for 30 min in 0.2× SSC–1% SDS at room temperature; then they were exposed to X-ray film with an intensifying screen.

In situ hybridization. Human metaphase spreads were prepared from cultured lymphoblasts for FISH by a standard methanol-acetic acid protocol. The cosmid clone pHC-RJ, containing the right-hand cellular junction preintegration region from the cell line DD16, was labeled by nick translation, incorporating biotin–16-dUTP (Boehringer Mannheim). Approximately 70 ng of labeled DNA was then ethanol precipitated along with 2 μg of human Cot 1 DNA (Gibco/BRL, Gaithersburg, Md.) and 9 μg of salmon sperm DNA for use on each metaphase slide. The hybridization solution consisted of 50% formamide–2× SSC–10% dextran sulfate at 37°C. After an overnight incubation at 37°C, washings were performed in 50% formamide–2× SSC at 42°C and then in 1× SSC at 60°C. Fluorescein isothiocyanate-conjugated avidin (Vector Laboratories, Burlingame, Calif.) was used for detection of hybridization signals. Counterstaining with 4′,6-diamidino-2-phenylindole (DAPI) generated the G/Q banding pattern. A cooled charge-coupled device camera (Photometrics, Tucson, Ariz.) was utilized in combination with a standard epifluorescence microscope (Zeiss Axioplan). The 16-bit source images were stored as normalized 8-bit Gray scale data files by using the software program CCD Image Capture (Yale University, New Haven, Conn.). High-plane parallel bandpass filters maintained the proper image registration. Image merging and pseudocoloring were accomplished with Gene Join (Yale University) on an Apple Macintosh computer. Finished color prints were produced with Adobe Photoshop 3.0.4 (Adobe Systems) and a Tektronix Phaser IISDX printer.

In vitro integration reactions. Cell extracts from untransfected 293 cells and 293 cells transfected with pHIV-Rep (2) were prepared as previously described (44) and dialyzed against 20 mM Tris-HCl–0.1 mM EDTA–50 mM NaCl–10% glycerol–1 mM dithiothreitol. Extracts were stored at −80°C prior to use. In vitro integration reactions were performed in 10 mM Tris-HCl–30 mM HEPES, pH 7.5, containing 7 mM $MgCl_2$, 0.1 mM dinucleoside triphosphates, 4 mM ATP, 40 mM creatine phosphate, 0.05 mM EDTA, 25 mM NaCl, 5% glycerol, 0.5 mM

dithiothreitol, and 1 μg of creatine phosphokinase. Typically, 0.5 to 1 μg of DNA substrate, 0.5 to 1 μg of DNA target, and approximately 200 μg of 293 cell extract protein were added to a reaction mixture to a total reaction volume of 30 μl. Reaction mixtures were incubated at 37°C for 4 to 6 h, after which they were incubated for 1 h at 37°C in the presence of 0.1 mg of proteinase K per ml. Following the proteinase K treatment, the resulting DNAs were subject to phenol extraction and ethanol precipitation. The DNA pellets were then resuspended in 25 μl of Tris-EDTA buffer, pH 7.0, containing RNase at 20 μg/ml. PCR amplification of in vitro integration products was performed, as outlined previously, with 1 μl of resuspended integration reaction mixture. TA cloning of PCR products into the vector pCR II was carried out, as described earlier in this report, by the supplier's protocol (Invitrogen). Positive clones were screened for the presence of AAV–ch-19 inserts by PCR with the previously described primers. Isolated clones were screened by digestion with *Eco*RI, which liberates the inserted DNA fragment from the cloning vector. Clones that contained inserts larger than 500 bp were selected for DNA sequence analysis.

RESULTS

Characterization of wt AAV–ch-19 integration in immortalized and normal diploid cells. Genomic DNA was isolated from several individual HeLa latent cell lines, a pool of WI-38 cells latently infected with AAV and the hybrid virus AAV/OR1-OR2, and the initial latently AAV-infected cell line Detroit 5. The resulting DNAs were subject to PCR amplification with several AAV-specific and a ch-19-specific primer combinations. The two AAV-specific primers utilized (H and T) are unique to the Rep and Cap regions, respectively, and read out toward the AAV ITRs at either end of the AAV genome, while the ch-19-specific primer (C) used was selected to a 100-bp region within the ch-19 preintegration locus previously identified (25, 40). The resulting PCR amplification products were subjected to dot blot analysis, and individual junctions derived from AAV–ch-19 integration events were cloned and sequenced.

The dot blot analysis used to identify PCR products from cell lines latently infected with AAV is depicted in Fig. 1A. A probe for ch-19 junction-specific sequences was used to analyze products of PCRs with the C primer alone as well as with T-C and H-C primer combinations, whereas a probe specific to the AAV ITR was used to analyze products of PCR with the T primer alone, the H primer alone, and the H-T primer combination. No detectable PCR product that hybridized with either probe was produced with any primer combination in uninfected HeLa cells. All cell lines shown, with the exception of C12, exhibited PCR products produced with the T-C primer combination that hybridized with the ch-19 junction-specific probe (Fig. 1A). Furthermore, with the exception of G12, PCR products generated with the H-C primer combination hybridized to this same probe. To generate both head and tail junctions with the same cellular sequence, an inversion and duplication of the genomic and/or viral DNA had to have taken place. No detectable PCR product hybridized to this probe in reactions containing only the single ch-19-specific primer C, suggesting that this inversion-duplication was directly related to wt AAV integration. With the AAV ITR-specific probe and the H-T primer combination, all cell lines exhibited ITR-positive PCR products. Only one cell line, G11, demonstrated evidence for an AAV tail-to-tail product, while no head-to-head structures were detected. The positions of the wt AAV breakpoints in relation to the ch-19 sequence are shown in Fig. 1B. The summary of these results supports a common integration structure for wt AAV which primarily involves head-to-tail tandems with cellular junctions containing rearrangements of genomic and/or viral DNA. These results are consistent with the original characterization of wt AAV integration in the Detroit 6 cells as previously described (23) and with a model for targeted AAV integration proposed by Berns and coworkers (30, 31). In addition to the ITR, the AAV p5 promoter (see below) was the only other common viral sequence we observed being used in the generation of junctions.

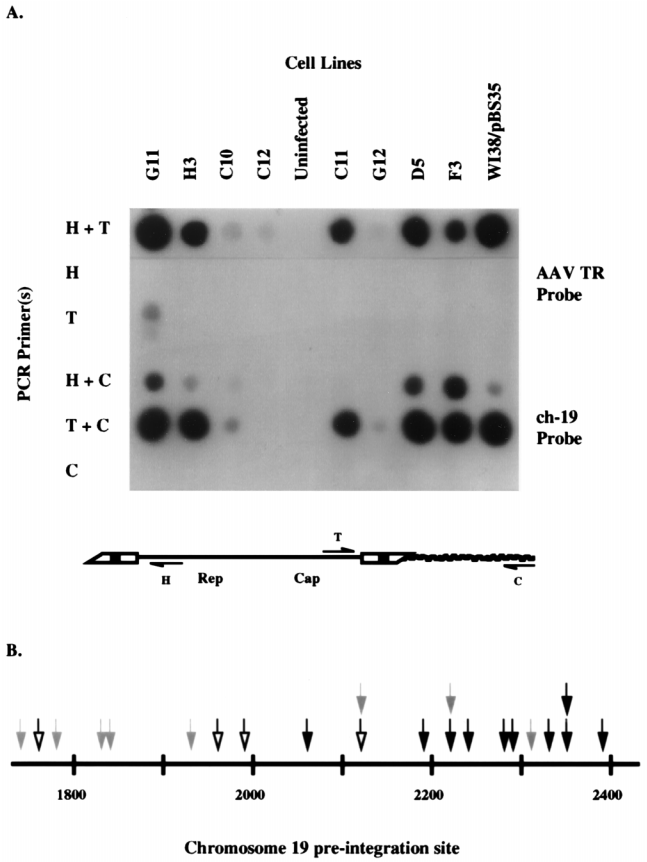


FIG. 1. (A) Dot blot analysis of PCR amplification products generated from AAV-cell junctions. Genomic DNAs isolated from various cell lines that were latently infected with AAV are listed horizontally. PCR amplification was performed with various combinations of the AAV primers H, which hybridizes to AAV sequences at positions 400 to 381, and T, which hybridizes to positions 4498 to 4514, and/or the ch-19 cellular primer C, which hybridizes to the AAV preintegration site present on the cloned *Bam*HI fragment found on ch-19. The various probes used are described in the text. (B) Locations (in nt) of chromosomal breakpoints of cloned AAV–ch-19 junctions derived from WI-38 cell lines (black arrows), A′–ch-19 junctions (open arrows), and P5–ch-19 junctions (gray arrows). See the text for a more detailed description.

To extend the analysis of wt integration in normal diploid cells, we focused on junctions generated in WI-38 cells. A pooled infection was used to generate PCR junction products. The PCR products produced with the various primer combinations indicated in Fig. 1A were cloned into *E. coli*, and individual isolates were selected by colony hybridization to a ch-19-specific probe. Ten unique positive clones were then sequenced. The junction breakpoint locations within the AAV ITR and the ch-19 preintegration site are listed in Table 1. Breakpoints in the ch-19 integration site were observed within an approximately 270-bp region mapping to nucleotides (nt) 2062 to 2346 of the *Bam*HI subclone containing the AAV ch-19 integration site described previously (40). The locations of these ch-19 breakpoints listed in Table 1 are shown in the map depicted in Fig. 1B. The cloned junctions contained breakpoints scattered throughout the putative hairpin structure of the AAV ITR (Fig. 2A and 2D). Several of these junctions contained sufficient portions of the ITR to allow determination of the original orientation (flip or flop) of the hairpin. Six of these clones contained junctions with simple

TABLE 1. AAV–ch-19 junctions from WI-38 latent cell lines[a]

Clone	Breakpoint (nt)	
	AAV	pRE2
TA-W1[b]	4607	2346
TA-W4	4576[c]	2062[c]
TA-W8[d]	4587	2191
TA-W12	4595	2277
TA-W13[e]	4586	2235
TA-WW6	4599	2328
TA-WW14	4592	2388
TA-WW16	4606	2345
TA-WW18[f]	4591	2290
TA-WW20	4465	2215

[a] Junctions were cloned as described in Materials and Methods. AAV and pRE2 breakpoint locations are based on nucleotide map positions as described in reference 40.
[b] Contains a 4-bp insertion (CACT) between the AAV and pRE2 breakpoints.
[c] Previously published in reference 40.
[d] Contains an insertion of AAV nt 265 to 158 between the AAV and pRE2 breakpoints.
[e] Contains an insertion of AAV nt 266 to 155 between the AAV and pRE2 breakpoints.
[f] A 5-bp insertion (TGGTC) between AAV and pRE2 breakpoints matches sequence in B′ of the AAV ITR. Inserted DNA sequences are listed reading from the AAV ITR out toward the cellular junction sequence.

crossovers between viral (AAV ITR) and cellular (ch-19) sequences. The breakpoint locations within the AAV ITRs of these clones are depicted in Fig. 2A.

The four remaining clones contained DNA sequences between the breakpoint of the AAV terminal repeat and the ch-19 target sequence that could not be accounted for by a simple crossover between viral terminal repeat and cellular DNA sequences and are noted in Table 1. The lengths of these sequences inserted between contiguous viral and cellular sequences ranged from 4 to 115 bp. These inserted sequences at the virus-cell junction matched DNA sequences present in the AAV genome. Two of these clones, TA-W1 and TA-WW18, contained short stretches ($\leq$5 bp) of DNA that matched sequence further downstream within a complete AAV ITR. The presence of these sequences can also be accounted for by a simple deletion within the AAV ITR. The other two clones, TA-W8 and TA-W13, contained significantly larger (>100 bp) insertions at the virus-cell junctions. These insertions matched sequence immediately upstream of the P5 promoter in the AAV genome and were in an inverted orientation relative to the AAV ITR. One of them is a simple 113-nt insertion of nt 266 to 154 from AAV. The other contains a 118-nt insertion of nt 265 to 158 from AAV, with a rearrangement of DNA sequence between nt 217 to 211.

No complete AAV ITRs were observed in any of the virus-host junctions derived from WI-38 latent cell lines (Table 1) that were cloned through PCR amplification with AAV primers upstream of the start of the AAV ITR. Although it may be argued that the deletions in the AAV ITRs observed in these junctions are an artifact of the PCR amplification itself, it has been previously shown that it is possible to prepare complete and replication-competent AAV ITRs by PCR amplification (47). Furthermore, the junctions listed in Table 1 are similar to AAV-cell junctions characterized previously in aneuploid cell lines (23, 30, 40) in that incomplete ITRs and, in some cases, short stretches of nucleotides inserted at the virus-host junction are observed. Moreover, AAV ITR-cell junctions cloned from PCR amplification products (see below) derived from DD-Neo latent cell lines were identical to junctions that were cloned directly into *E. coli*. This result serves as a control and indicates that the PCR reactions did not cause significant alterations in the junctions that were characterized in this study.

To characterize junctions with breakpoints in the A′ region of the AAV ITR, a PCR amplification of virus-cell junctions that did not require polymerization through the entire AAV ITR was designed. Junctions from the HeLa latent cell lines G11, H3, and C11 were PCR amplified from genomic DNA by using the AAV primer TR6 and the ch-19 primer C. The AAV primer TR6 hybridizes to the A region of the AAV ITR and matches sequence in the A′ region (Fig. 2). Four independent junctions (three from H3 and one from C11) that possessed ITR breakpoints in the A′ region (nt 4655 to 4675) were characterized. These junctions and their respective breakpoint locations within the AAV ITR and the ch-19 target region are listed in Table 2. AAV ITR breakpoints were located between nt 4655 and 4673, whereas ch-19 breakpoints were located between nt 1763 and 2119. The physical structure of the clone TA-H14 is depicted in Fig. 2B.

In addition to the expected A′-cell junctions, a second group of novel junctions was produced by PCR amplification with the TR6-C primer pair. These junctions possessed AAV breakpoints that fell within the AAV P5 promoter region upstream of the start of the Rep open reading frame (nt 146 to 320) joined to ch-19 sequence, whose breakpoints fell between nt 1731 and 2294. The physical structure of the clone TA-H21 is also depicted in Fig. 2B. Nine junctions of this type (six from G11 and three from H3) were characterized and are listed in Table 2. Two of these junctions contained 11- to 12-bp insertions of DNA sequence at the virus-cell junction. These two groups of clones are consistent with virus-cell junctions possessing complete AAV ITRs with breakpoints within the A′ region or provirus structures involving a head-to-tail tandem structure with a virus-cell breakpoint in the P5 region of the AAV genome. However, because there is no information on sequence upstream of the TR6 primer, neither the integrity of the ITR structure 5′ to the PCR primer nor the exact structure of the provirus upstream of the A′ region can be commented on based on these results.

Besides the ITR- and P5-specific ch-19 junctions, the only other PCR product we obtained with any frequency involved viral head-to-tail junctions. Two unique clones containing head-to-tail junctions, obtained by using a combination of H and T primers, are shown in Fig. 2C. A physical map of the two unique junction sequences is depicted in Fig. 2C. Both of these junctions contain significant deletions within the AAV terminal repeat. The clone TA-ht5 crosses over from the A region of the tail terminal repeat into the D region of the head terminal repeat, while the clone TA-ht11 crosses over from the C region of the tail terminal repeat in the flop orientation into AAV sequence upstream of the P5 promoter at the head. While Southern blot analysis data supported the existence of fragments consistent with intact ITRs in these head-to-tail arrangements, none of the PCR products we isolated contained a complete terminal repeat. It is not clear, however, whether deletions and/or rearrangements contained in these head-to-tail clones were an artifact (47) or were produced during the integration process. The locations of all the unique AAV breakpoints within the putative terminal repeat hairpin of the AAV junction clones listed in Tables 1 and 2 are depicted in Fig. 2D.

Analysis of rAAV provirus structures. Numerous attempts to isolate an intact wt AAV provirus structure resulted in failure. While various clones could be isolated, they either were not stable in *E. coli* or rearranged as we propagated the bacteria as previously described (6) (data not shown). To characterize an intact AAV provirus structure, we focused on a

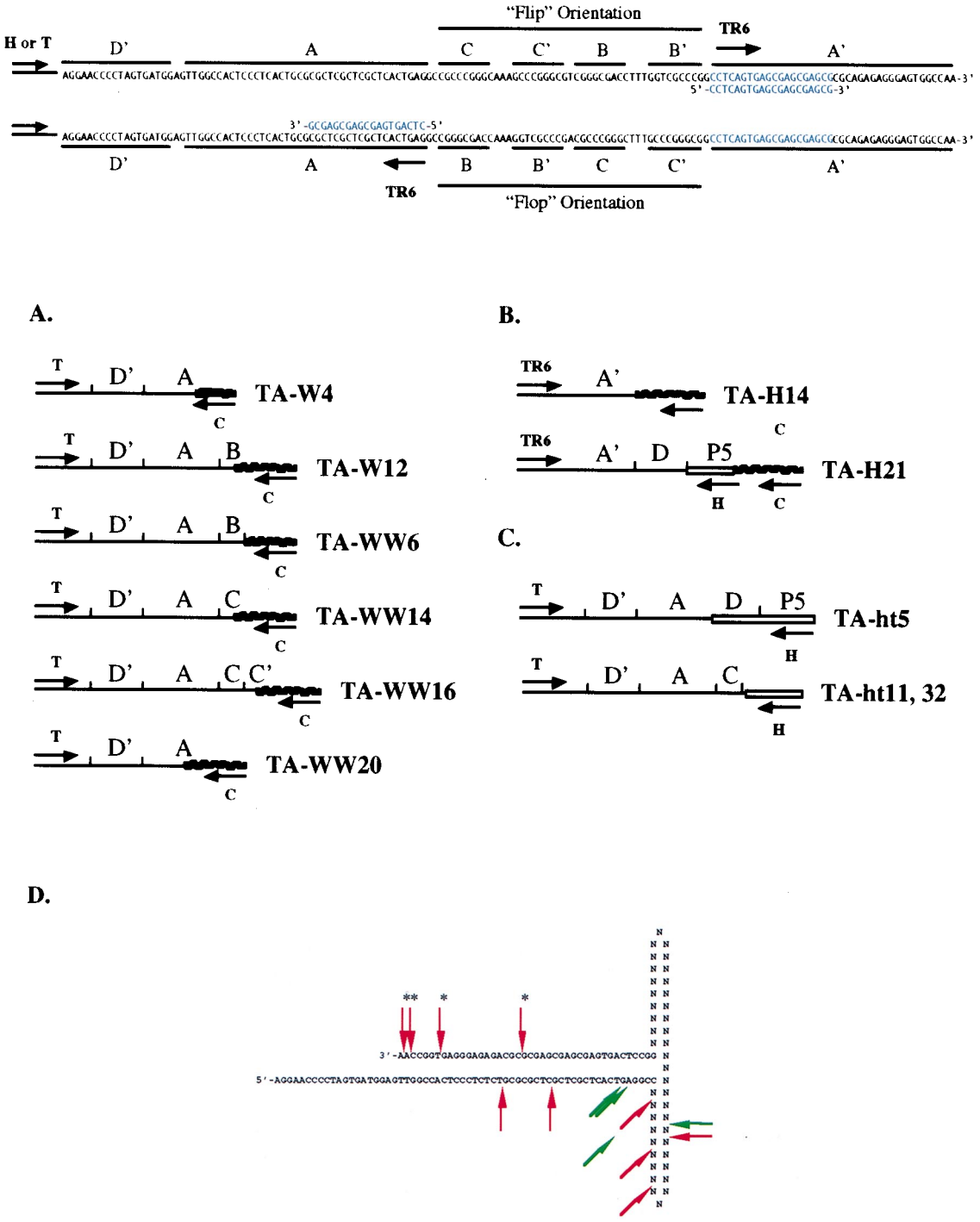


FIG. 2. (A) Structural maps of unrearranged AAV–ch-19 junctions derived by PCR amplification of genomic DNA from latently infected WI-38 cell lines. Nucleotide maps of the AAV ITR in its two possible orientations are shown at the top. (B) Structural maps of representative AAV–ch-19 junctions derived by PCR amplification of latently infected HeLa cells with AAV primer TR6 and ch-19 primer C. (C) Structural maps of head-to-tail structures derived by PCR amplification with AAV primers H and T of genomic DNA from HeLa cells latently infected with AAV as previously described. The AAV terminal repeat regions are indicated as previously described. (D) Physical location of the AAV junction breakpoints within the putative AAV terminal repeat hairpin structure. Asterisks indicate junctions derived by PCR amplification with the AAV primer TR6. Green arrows indicate junctions that contained insertions of nucleotides between ITR and cellular sequences. The nucleotides in the B-B′ and C-C′ regions are listed as N, since the purpose of this map is to indicate physically the portion of the ITR remaining at the junction, regardless of the ITR orientation (flip or flop).

rAAV vector containing neomycin and ampicillin resistance genes and a bacterial origin of replication. The construct, DD-Neo, is shown in Fig. 3. Recombinant stocks of DD-Neo were generated from the unique vector plasmid pDD-Neo as previously described (47). The strategy for retrieving intact rAAV provirus, also outlined in Fig. 3, involved direct cloning of the vector from genomic DNA isolated from established latent cell lines (G418 resistant) into *E. coli* by using the bacterial origin

TABLE 2. AAV junctions from HeLa cells amplified with primers TR6 and C[a]

Clone	Breakpoint (nt)	
	AAV	pRE2
TA-H14	4673	1991
TA-H15	4674	1957
TA-H20	4668	2119
TA-C25	4655	1763
TA-G2[b]	269	1827
TA-G5	264	1930
TA-G6	294	1783
TA-G8	270	1744
TA-G9	281	2099
TA-G12[c]	262	2120
TA-H16	160	1838
TA-H21	288	2222
TA-H23	278	2307

[a] Junctions were cloned as described in Materials and Methods. AAV and pRE2 breakpoint locations are based on nucleotide map positions as described in reference 40.
[b] Contains an 11-bp insertion (TTAAATACCCA) between the AAV and ch-19 breakpoints.
[c] Contains a 12-bp insertion (TGGGTATTTAAG) between the AAV and ch-19 breakpoints.

of replication and an ampicillin selection marker contained on the vector. Detroit 6 cells were used for infection with recombinant viral stocks of DD-Neo in the presence of the selective antibiotic geneticin G418. Individual drug-resistant colonies were selected, and cell lines that had undergone single or few vector integration events, as confirmed by Southern analysis, were chosen for provirus cloning. All of these clones were negative for rescue when assayed by wt AAV-adenovirus coinfection.

Cloning of the rAAV provirus. Genomic DNA from six DD-Neo latent cell lines established in Detroit 6 human cells was isolated and digested with *Bgl*II, a restriction enzyme that does not cut the DD-Neo vector. Southern hybridization analysis of *Bgl*II-digested DNA from six latent cell lines with a probe specific for the DD-Neo vector indicated the presence of unique *Bgl*II fragments larger than the 4-kb vector itself in five of the six cell lines (data not shown). One cell line, DD2, exhibited two DNA fragments that hybridized to the vector probe that were larger than pDD-Neo. Digestion of genomic DNA with *Hin*dIII or *Ssp*I, each of which cuts the vector once, revealed the presence of two DNA fragments that contained vector sequences, whereas the cell line DD2 showed four DNA fragments that hybridized to the vector probe (data not shown). These data supported single or relatively simple integration patterns, which were further characterized by direct genomic cloning of the provirus.

*Bgl*II-digested genomic DNA was used for direct cloning of plasmids containing vector-cell DNA junctions. The digested DNA was recircularized by T4 DNA ligase, and the ligation mixture was used to transform *E. coli*. Ampicillin-resistant transformants were grown in liquid medium, and the resulting plasmid clones were isolated and characterized. Six plasmid clones from five independent cell lines, chosen on the basis of size (>4 kb) and the presence of restriction fragments characteristic of the DD-Neo vector, were characterized. Two independent clones, DD2-1 and DD2-2, were isolated from the cell line DD2. Restriction enzyme mapping of the provirus clones revealed that the vector DD-Neo had integrated into the host genome intact in all cases, with breakpoints located in or near the terminal repeats (Fig. 4A). One clone, derived from the latent cell line DD18, possessed a breakpoint in the vector about 200 bp downstream from the ampicillin resistance gene. DNA sequence analysis of the provirus clones indicated that none of the vector integrants contained a complete viral terminal repeat at the virus-cell junction (Fig. 4A). The breakpoints within the vector ITR at the crossover between vector and host for both the left (5′ end, or head) and right (3′ end, or tail) junctions are shown in Fig. 4A. As with the cloned wt AAV junctions, several of these DD-Neo junctions contained sufficient portions of the terminal repeats to indicate the vector orientation of the terminal repeat (e.g., flip for the right junction of DD16 and flop for the right junction of DD5). The locations of these cloned vector junction breakpoints within

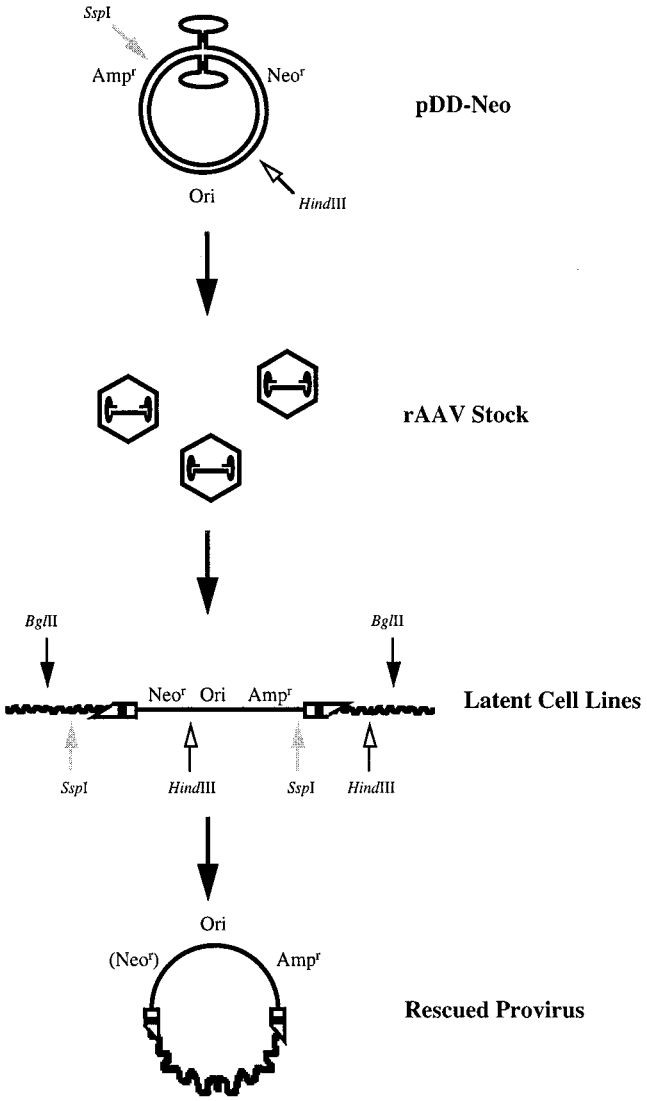
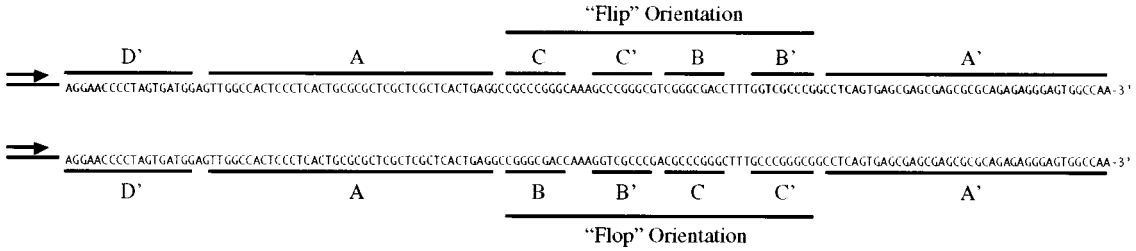
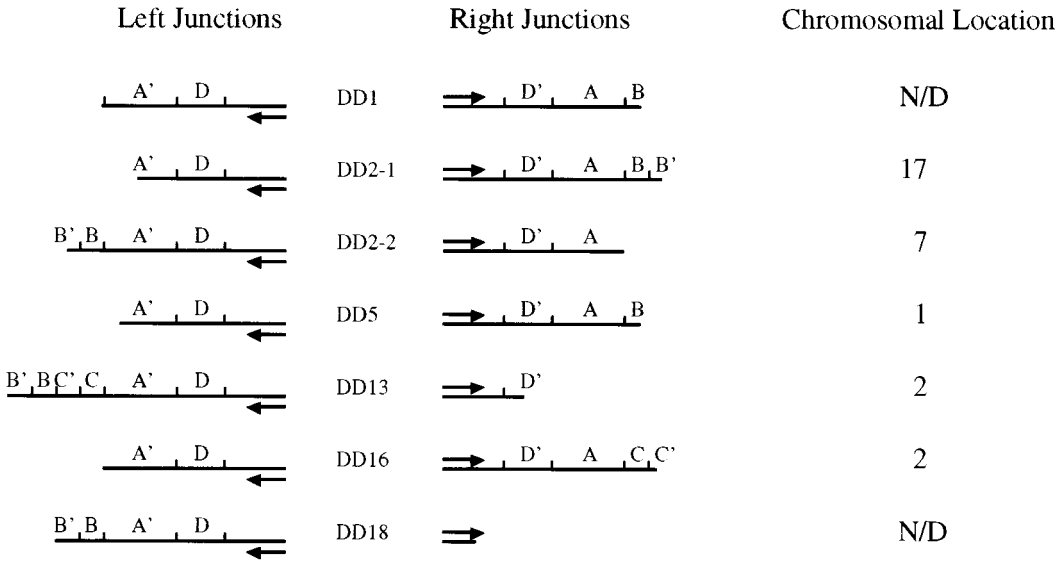


FIG. 3. Strategy for the construction, establishment, and isolation of rAAV provirus DNA. pDD-Neo was used to generate recombinant virus stocks, which in turn were used to establish DD-Neo latent cell lines. Genomic DNAs isolated from individual drug-resistant colonies were digested with restriction enzymes that either cut outside the vector genome or cut once in the vector, recircularized with DNA ligase, and propagated in *E. coli*. The resulting rescued proviruses containing vector-cell junctions were sequenced and further characterized by Southern analysis and by FISH to human chromosomes.



A.



B.

FIG. 4. (A) DD-Neo rAAV terminal repeat sequences present at left- and right-hand vector-cell junctions of individual latent cell lines. Proviruses whose chromosomal locations have been mapped by hybridization to rodent-human somatic cell hybrid panels (see text for description) are indicated on the right. A nucleotide map of the AAV ITR is shown at the top in its two possible orientations (flip or flop). N/D, not determined. (B) Physical locations of the DD-Neo junction breakpoints within the putative AAV terminal repeat hairpin. The nucleotides in the B-B′ and C-C′ regions are listed as N, since the purpose of this map is to indicate physically the portion of the ITR remaining at the junction, regardless of the ITR orientation (flip or flop).

the putative AAV terminal repeat hairpin are depicted in Fig. 4B.

Attempts to clone a complete provirus from the latent cell line DD1 by the techniques previously described proved unsuccessful. Southern analysis of the genomic DNA from the latent cell line DD1 indicated that the provirus was contained on a *Bgl*II fragment in excess of 30 kb (data not shown). To characterize the junctions for this provirus, genomic DNA was digested with either *Hin*dIII or *Ssp*I and was recircularized as described for *Bgl*II genomic DNA, and the resulting pool was used to transform *E. coli*. By using one of these enzymes that uniquely cuts the vector, the left and right junctions of the

integrated provirus may be cloned out separately (Fig. 3). Ampicillin-resistant clones containing provirus junction sequences were isolated from both ligation pools and characterized. DNA sequence analysis revealed that a single type of clone was derived from each pool. Breakpoints in the ITR for the junctions from DD1 are also shown in Fig. 4A. As with the complete proviruses cloned from the other five latent cell lines, neither junction contained a complete AAV terminal repeat. However, the right junction sequence for DD1, derived from the cloning of *Hin*dIII digestion fragments, contains a 12-bp stretch of sequence, inserted at the vector-cell junction, matching sequence present in the C′ region of the AAV ITR. As with the AAV–ch-19 junctions described earlier in this report, this result can be accounted for by either an insertion of this fragment at the vector-cell junction or a deletion of the B′ and C sequences within the AAV ITR.

As with the cloning of PCR amplification products, the cloning and propagation of rAAV junctions may have altered the vector-cell junction sequences, given the potential for instability of AAV terminal repeat sequences in *E. coli* (8, 28, 37, 39). To address this question, vector-cell junction sequences derived from the provirus clones described in this work were compared to junctions cloned through PCR amplification of genomic DNA. Primers specific to cellular DNA sequences derived from junction clones were used with primers specific to the head and tail regions of the DD-Neo vectors. Agarose gel mobility and DNA sequence analysis revealed that the PCR products of junctions amplified from genomic DNA and the respective provirus clones were identical in both size and sequence (data not shown). Furthermore, the single-copy provirus in DD16 was directly cloned in independent experiments, and the vector-cell junctions from both experiments were identical in sequence. While it may be argued that both techniques may be subject to potential deletions, it is unlikely that two independent protocols would produce identical artifacts. These results together indicate that the junctions cloned in this work, either directly or from PCR amplification products, are authentic and not experimental artifacts.

Analysis of the first 500 bp of cellular DNA sequence adjoining the vector-cell crossover of these cloned proviruses indicated that none of the sequences present at either the left or right junctions in any of the seven clones shared identity (data not shown). Furthermore, when compared to the ch-19 preintegration site for AAV, none of the sequence obtained from these vector-cell junctions matched the sequence of the *Bam*HI subclone (40) containing a 2.7-kb fragment of the AAV preintegration site or the 4.1-kb *Eco*RI fragment AAVS1 (24), supporting non-ch-19-targeted integration (see Fig. 7 and data not shown).

Characterization of vector integration loci and chromosome mapping. Using the DNA sequence obtained from the provirus-cell junctions of the plasmid clone pDD16, PCR primers were designed to produce DNA probes for the left and right junctions of this provirus. Random primed ^{32}P-labeled fragments of the left and right junctions were used to probe *Bgl*II-digested total genomic DNA from several latent cell lines. Figure 5A depicts a Southern blot of genomic DNA from several cell lines latently infected with DD-Neo and probed with a ^{32}P-labeled DNA fragment specific to the right junction of the provirus in the latent cell line DD16. In all cell lines, the probe hybridized to a DNA fragment corresponding to a molecular weight of approximately 4 kb. In addition, the probe hybridized to a DNA fragment corresponding to a molecular weight of approximately 8.5 kb in the cell line DD16 and to a 10-kb *Bgl*II fragment present in the cell line DD13. The 8.5-kb and the 10-kb *Bgl*II fragments from the latent cell lines DD16 and DD13, respectively, comigrated with the provirus clones derived from those two cell lines (data not shown). The left-junction probe from DD16 hybridized to a 7-kb *Bgl*II fragment from all of the latent cell lines, as well as to an 8.5-kb fragment in the cell line DD16 and a 10-kb fragment in DD13 (Fig. 5B). These results support integration by two independent rAAV vectors into a cellular DNA sequence. Furthermore, Southern analysis of the preintegration site indicates that the left and right junctions of DD16 are not contiguous, since the left and right junction probes from DD16 hybridize to different genomic DNA fragments (Fig. 5A and B). These results suggest that rearrangement of the genomic DNA occurs during or after vector insertion, similar to the wt (data not shown). This observation is of extreme interest since an independent study has also mapped the integration of rAAV vectors into ch-2, suggesting that a preferred integration is being observed with these vectors (20).

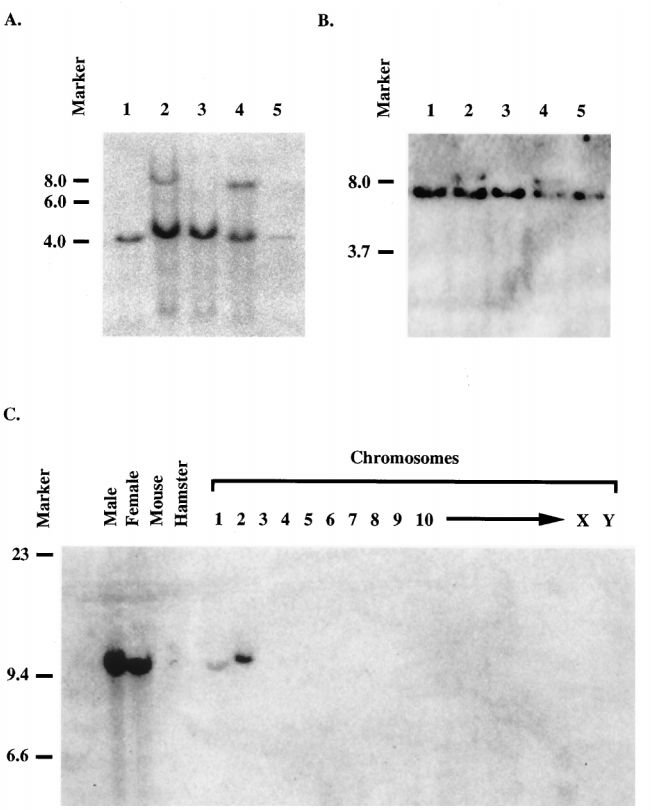


FIG. 5. (A) Southern blot of genomic DNA isolated from individual cell lines latently infected with DD-Neo, as described in the text, and probed with cellular sequences from the right-hand junction derived from the cell line DD-16. Lane assignments are indicated horizontally, and molecular size markers (in kilobases) are noted vertically. Lanes: 1, DD5; 2, DD13; 3, Detroit 6; 4, DD16; 5, DD18. (B) Southern blot of genomic DNA isolated from individual latent cell lines as described in the text and probed with cellular sequences from the left-hand junction derived from the cell line DD-16. Lane assignments are the same as for panel A. (C) Southern blot of a somatic cell hybrid panel (ONCOR, Inc.) probed with cellular junction sequences derived from the cell line DD-16. Marker locations and lane assignments are indicated horizontally.

Rodent-human somatic cell hybrid panels digested with *Bam*HI, each containing the total DNA from an individual human chromosome, were used to determine the chromosomal locations of the left and right junctions of the provirus cloned from the latent cell line DD16. Both probes hybridized to the panel corresponding to ch-2. Neither fragment hybrid-

ized to hamster or mouse DNA. The left-junction probe hybridized to a 10-kb *Bam*HI fragment and exhibited some weaker cohybridization to a similarly sized fragment in ch-1 (Fig. 5B). The right-junction probe hybridized to a 7-kb *Bam*HI fragment, with no cohybridization to any other panel (data not shown). A probe specific for the left junction from another clone of the cell line DD2 exhibited hybridization to a 1-kb *Bam*HI fragment present on ch-17, as well as sequences present in the hamster (data not shown). The locations of the junction breakpoints within the vector ITR of the individual proviruses are shown in Fig. 4A, and their locations within the putative ITR hairpin are shown in Fig. 4B. The cellular sequences at the vector-cell junctions of the provirus clones from the cell lines DD5, DD16, DD2-2, and DD2-1 were used as probes, and they have been mapped to chromosomes 1, 2, 7, and 17, respectively, by use of the same rodent-human somatic hybrid panels (Fig. 4A). There is a remote possibility, however, that these sequences have been translocated in Detroit 6, given their aneuploid nature or as a result of the integration event itself. Although this possibility cannot be ruled out at this time, FISH analysis of the latent cell lines supported the results obtained with the somatic cell hybrid panel (data not shown).

In situ chromosome and DNA sequence analysis of the ch-2 provirus preintegration site. The cosmid pHC-RJ, which contains the right-junction preintegration region for the provirus contained in the cell line DD16, was isolated and used as a probe for in situ chromosome analysis to further map the location of this preintegration region in Detroit 6 cells. The hybridization signals localized exclusively to ch-2, as shown in Fig. 6A. DAPI (G/Q) banding and FLpter analysis (64 to 70%) of approximately 20 metaphase spreads showed that the pHC-RJ signals were located within the q24-q31 subregion. Several individual examples from independent metaphase spreads are shown in Fig. 6B. The minor differences in size of the individual chromosomes can be attributed to the different states of condensation during metaphase for chromosomes coming from different spreads. The localization of the hybridization signals is diagrammed by the ideogram in Fig. 6C. The cosmid pHC-RJ was sequenced with primers derived from cellular sequences from the plasmid pDD16. The left-junction probe for DD16 did not hybridize to this cosmid, consistent with the results we observed for Southern analysis (Fig. 5A and 5B), and this is indicative of rearrangement (deletion and/or translocation) of genomic DNA at the provirus integration site.

The preintegration sequence surrounding the right (3′, or tail) vector-cell junction for DD16 is shown in Fig. 7, along with the preintegration sequences for the left junction of DD2-1, located on ch-17, and the right junction of DD2-2, located on ch-7. Comparison of terminal repeat and cellular preintegration sequences reveal 2 or 3 bp (depicted in boldface in Fig. 7) of microhomology that cannot be distinguished in viral and cellular DNA. rAAV provirus structures were similar to wt integrants in that both flip and flop orientations of the ITRs were observed, microhomology between the viral and cellular breakpoints existed, and vector integration resulted in rearrangement of cellular DNA. In addition, we saw an accumulation of similar breakpoints between rAAV and wt ITR sequences (compare Fig. 2C with Fig. 4B). Unlike the wt, rAAV vectors did not appear to integrate as tandems, and although we observed a common integration site on ch-2, targeting to ch-19 was completely lost.

In vitro integration. The in vitro integration of several AAV and rAAV DNA substrates into the AAV ch-19 target site located on a circular plasmid (pRE2) was performed under conditions described earlier in Materials and Methods. Four different AAV DNA substrates were examined for their ability to integrate in vitro into the ch-19 AAV integration sequence contained on the plasmid pRE2: no-end AAV DNA, described previously (34, 41); the linear *Xba*I fragment derived from AAV plasmid SSV9; the SSV9-derived linear *Sma*I fragment; and the SSV9-derived linear *Pvu*II fragment. These AAV substrates are depicted in Fig. 8A.

In vitro integration of no-end AAV substrates. The DNA sequences of 24 individual clones derived from the PCR amplification of products of in vitro integration reactions between no-end DNA and pRE2 were obtained. Two primer combinations were used in the PCR amplification of the products from the in vitro integration reactions: primer pair H-C and primer pair T-C. These clones and the locations of the virus-cell breakpoints are listed in Table 3. Integration reactions carried out with extracts of 293 cells that had been transfected with pHIV-Rep are noted as such. Integration junctions were obtained both in the presence and absence of Rep in the cell extract.

Nineteen independent junction clones were isolated from the PCR amplification of products of in vitro integration reactions between no-end AAV DNA and pRE2 with primers C and H. Of these junction clones, 11 contained breakpoints that were located in the AAV terminal repeat region (nt 4 to 184). The remaining junction clones contained breaks within the P5 promoter region of AAV (nt 192 to 352). None of these virus-cell junctions contained a complete, intact AAV terminal repeat. A map of the physical locations of the AAV breakpoints of the 11 in vitro junction clones that were located in the AAV ITR is depicted in Fig. 8B. Breakpoints in the pRE2 target DNA for 15 of these clones mapped between nt 1903 and 2377. The pRE2 (ch-19) breakpoints from these junctions are depicted in Fig. 8C. The four remaining clones contained rearrangements of the substrate and/or target DNA sequences. Insertions of these sequences were observed in these clones between either an AAV breakpoint in the P5 region or the AAV terminal repeat and a pRE2 breakpoint mapping between nt 2070 and 2395. The rearranged DNA sequences are noted in Table 3. Two of these clones, 34.4.5 and 15.5.9, contained insertions of AAV sequences from the capsid coding regions. The other clones contained more complex rearrangements, in which insertions of both AAV- and pRE2-associated sequences occurred in both direct and inverted orientations. All integration junctions that contained insertions or inversions of AAV and/or target sequences at the junction breakpoints were generated with 293 cell extracts in the absence of Rep protein.

In vitro integration junctions from five individual clones derived by PCR amplification with the T-C primer combination were also characterized. All of the junctions cloned with the T primer contained AAV breakpoints within the AAV terminal repeat region (nt 4489 to 4675) and were located within a 10-bp region surrounding the Rep nicking site (nt 4566). Breakpoints in pRE2 of these clones were located between nt 2042 and 2202. One of these junction clones contained an insertion of AAV sequences from nt 2358 to 2639, between the AAV ITR and pRE2 breakpoints. These clones are also summarized in Table 3, and the physical map locations of the AAV ITR and pRE2 breakpoints are also shown in Fig. 8B and C.

In vitro integration of linear AAV derivatives into pRE2. In vitro integration junctions derived from linear, double-stranded AAV DNA substrates that contain portions of the complete AAV terminal repeats and pRE2 were PCR amplified with primers H and C and cloned in a manner similar to junctions derived from the in vitro integration of no-end DNA. These linear AAV substrates were derived from the AAV plasmid

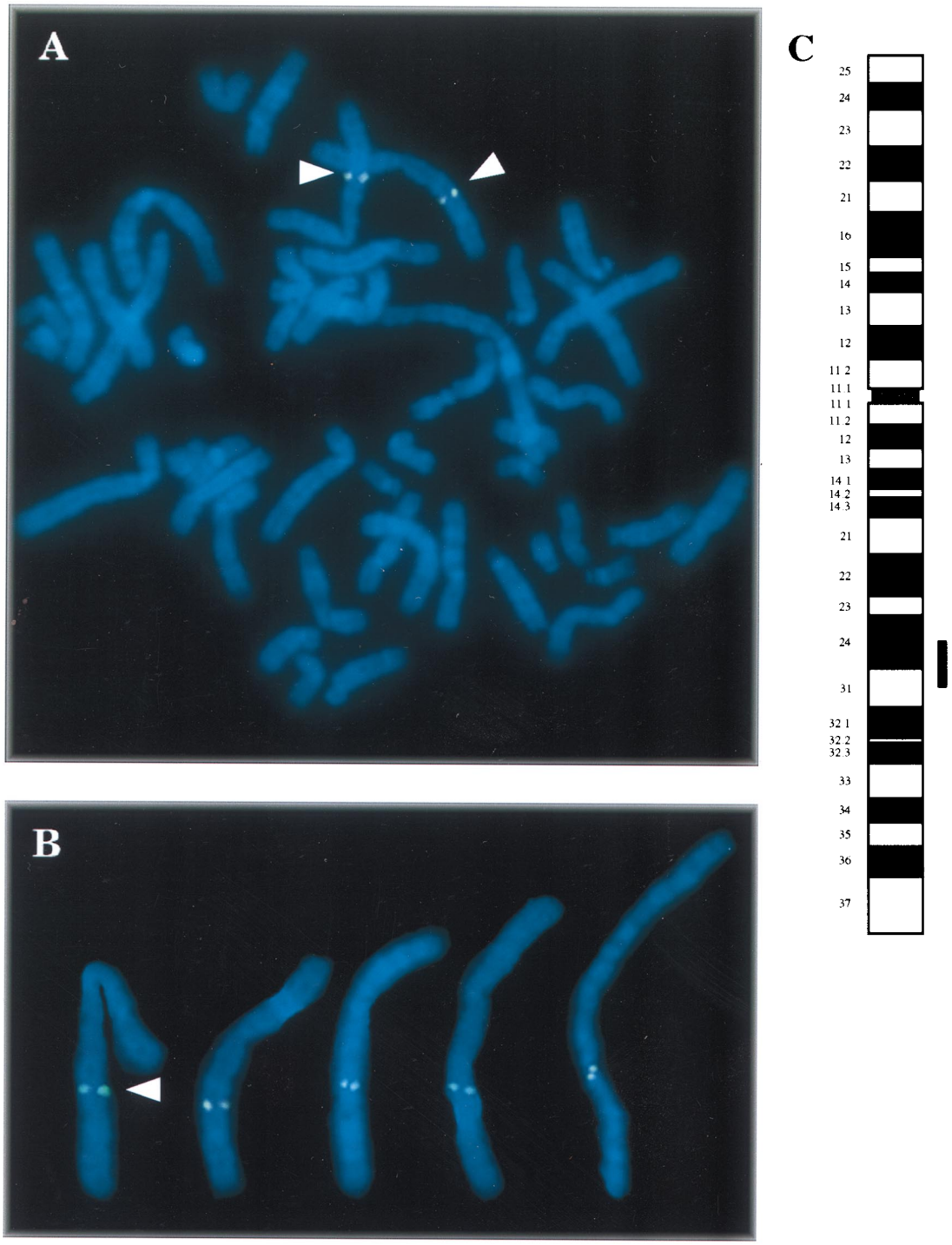


FIG. 6. (A) FISH of metaphase human chromosome spreads. (B) Fluorescein isothiocyanate-conjugated avidin was used to detect hybridization signals by use of cellular junction sequences derived from the cell line DD16, while a counterstain with DAPI generated the G/Q banding pattern. Individual ch-2 from several metaphase spreads indicating hybridization to the long arm in the interval q24-q31. (C) Ideogram of ch-2 indicating locations of hybridization signals determined by using the previously described probe.

SSV9. The *Pvu*II fragment contains the entire linear AAV genome with two complete intact AAV terminal repeats, while the *Sma*I fragment contains the AAV genome and half of each AAV terminal repeat. The *Xba*I fragment contains the AAV genome without AAV terminal repeat sequences. Integration junctions between these substrates and pRE2 were isolated by PCR amplification and TA cloning as previously described for no-end DNA. The junctions that were characterized in these experiments are listed in Table 4.

No junction clones were obtained from the linear *Pvu*II fragment. Fifteen junctions between pRE2 and the *Sma*I and *Xba*I linear AAV substrates were cloned and characterized. Six

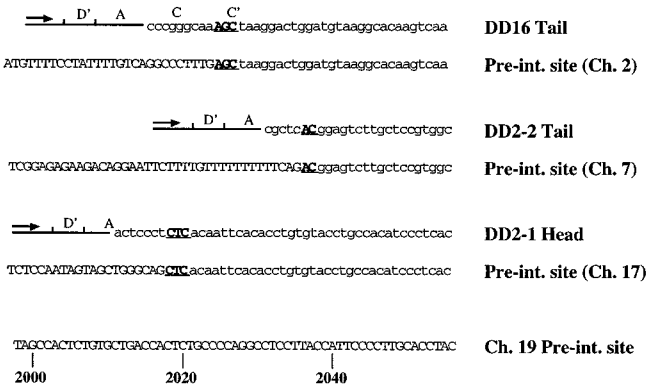


FIG. 7. DNA sequences of the vector-cell junctions for three independent provirus clones and their respective chromosomal preintegration loci. The ch-19 preintegration (Pre-int.) locus is shown at the bottom of the figure. DNA sequences upstream of the integration site are in capital letters, while DNA sequences that cannot be differentiated from vector or host DNA sequences are in boldface and underlined. The preintegration region of ch-19 displayed is from nt 1998 to 2056 of the *Bam*HI subclone containing the AAV target site described previously (40), where the majority of AAV–ch-19 junctions were observed in that study.

of the seven junctions obtained from the in vitro integration of the *Sma*I linear AAV substrate into pRE2 were joined at the *Sma*I site terminus, whereas five of the eight junctions obtained from the in vitro integration of the *Xba*I linear AAV substrates into pRE2 were joined at the *Xba*I site terminus. The remaining substrate-target junctions that were cloned contained deletions of DNA sequences from the end of the AAV substrate from the *Sma*I site (nt 46) or the *Xba*I site (nt 184). Breakpoints for these junction clones in the pRE2 target sequence were located between nt 1867 and 2367.

DISCUSSION

In this report, we have characterized the sequence structure of integrated wt AAV proviruses located in ch-19 of diploid and aneuploid cells and the structure, sequence, and location of the provirus of rAAV from latent tissue culture cells and described what we believe to be the first analysis of in vitro integration of various AAV substrates into a plasmid containing the human ch-19 target site. From these results, the following conclusions can be made. Establishment of wt AAV latency primarily involves head-to-tail viral structures, with the ITRs and p5 promoter sequences being involved in the crossover (Fig. 9A). Virus integration typically results in duplication and inversion of virus and/or host DNA sequences at the target site. Breakpoints in the chromosome and ITR are clustered (Fig. 9B) and, although not sequence specific, show microhomology between ITR and host DNA sequences at the crossover point. Characterization of direct clones of single-copy rAAV proviruses indicates that the vector genome integrates intact, with no rearrangement of vector sequences except for the ITRs which form the crossover points (Fig. 9A). The structure of the host genome is rearranged to some extent, and the vector has lost specificity for ch-19. As with wt, rAAV also demonstrates breakpoints in the ITR that are clustered (Fig. 9B). Analysis of in vitro junctions between various AAV substrates and the AAV ch-19 target site indicates that the terminal-repeat hairpin is the likely substrate required for integration in the establishment of viral latency. Substrate-target DNA sequence rearrangements sometimes seen in vitro are consistent with the structures seen in wt AAV integration. The observed junctions seen in all these systems indicate that wt, recombinant, and in vitro AAV integration uses a hairpin ITR for an integration substrate (Fig. 9B) and generates cellular junctions that are similar (exhibiting microhomology) and suggests that the integration event itself is mediated for the most part by cellular factors.

AAV-rAAV integration and its requirements. To date, the cloning of a complete wt AAV provirus has proven to be elusive. Recently, Berns and coworkers have proposed a model describing the site-specific integration of AAV (30, 31). This model requires the Rep binding site (RBS) and the Rep nicking site (terminal resolution site) and involves multiple strand switching during novel DNA strand synthesis within a Rep-mediated complex formed between AAV and the AAV target site. The multiple-strand switches proposed in this model could account for the various provirus structures we observed. Furthermore, additional cellular machinery required for DNA replication and repair is necessary for completion of the integration process. The results obtained in this study characterizing wt AAV, rAAV, and in vitro junctions indicate that this cellular machinery and the AAV ITR are sufficient for AAV or rAAV integration. At the same time, however, these results do not infer or imply any less of a role of the viral Rep protein in the targeted integration observed for wt AAV and how it may affect the formation of some of the various provirus structures observed during the establishment of viral latency.

If the hairpin structure is the correct substrate for integration and intact ITRs are not retained during or after integration, then the presence of head-to-tail tandems in provirus structures would strongly support the occurrence of a replication step prior to integration, as illustrated in Fig. 9A, step I. This tandem substrate could then participate in integration, forming the virus-cell breakpoints we observed (Fig. 9A, steps II and III), and retain sufficient sequences of the ITR for the rescue process (39) through the head-to-tail tandem. A similar model was proposed in 1993 to explain how rescue and replication could ensure with provirus structures that contained rearranged ITR junctions (36). While this model is attractive, a primary concern with analysis of latent cell lines for structural information on the mechanism of AAV integration is the possibility of subsequent events occurring before characterization can take place (2 to 4 weeks to derive clonal cell lines). This concern is echoed in the literature with the observation that high-passage-number, latently wt-AAV-infected cell lines produce free-monomer AAV genomes, suggesting that a dynamic provirus state exists even after extended latency (11, 32). It is very possible that the complex viral structures we have characterized with wt AAV in this study are a result of vector amplification and rearrangement after integration (Fig. 9A, step IV). Transient expression of Rep would be sufficient to initiate a rescue step, which in turn could generate nicked substrates that would initiate replication in situ, generating substrates reminiscent of the onion skin model proposed for simian virus 40 in the early 1980s. The results support the existence of a complex provirus structure in latently infected cell lines, which is consistent with previous observations (5, 11, 27, 32, 46); this complex structure has been reproduced in an EBV episome containing the AAV ch-19 target sequence, in which the provirus frequently possesses a head-to-tail structure (15).

An effort to characterize AAV integration without the concern of rearrangement during the extended time period required to generate clonal cell lines has supported the use of the transient integration system established by Berns and colleagues. Several classes of AAV–ch-19 junctions were observed in the study of wt AAV integration into an EBV shuttle

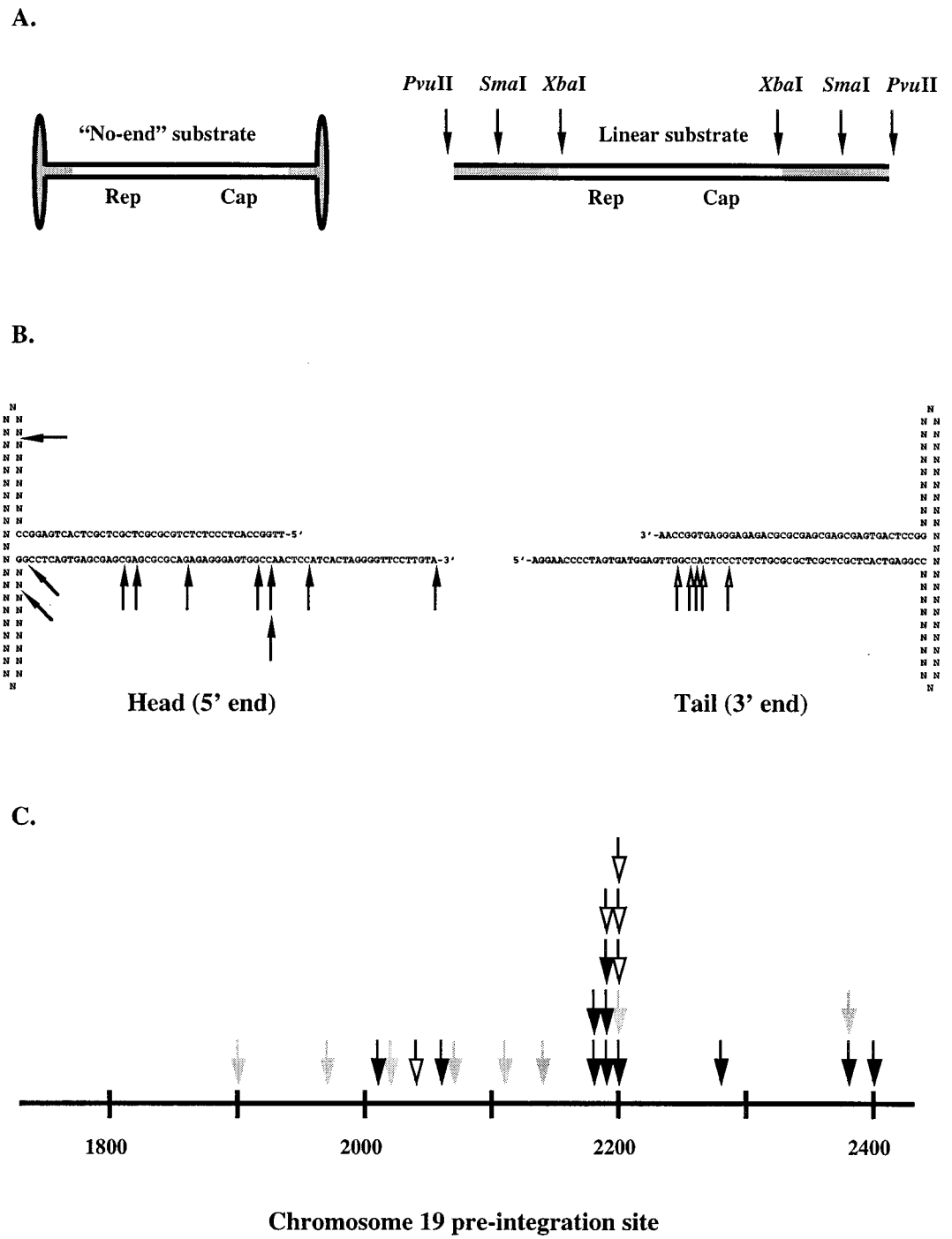


FIG. 8. (A) AAV DNA substrates for in vitro integration reactions derived from the AAV plasmid SSV9. The AAV ITR sequences are shaded. No-end DNA has been described previously (40). Locations of the *Xba*I, *Sma*I, and *Pvu*II restriction sites are as indicated. (B) Breakpoint locations within the AAV ITR for in vitro integration junctions generated between pRE2 and no-end DNA. ITR breakpoints derived by PCR amplification of the head (5′ end; black arrows) and tail (3′ end; open arrows) junctions are as indicated. (C) pRE2 (ch-19) breakpoint locations from in vitro integration reactions. Black and open arrows indicate where ch-19 sequences are joined to head and tail ITR sequences, respectively. The gray arrows indicate ch-19 breakpoints joined to AAV sequences within the P5 promoter region.

vector containing the ch-19 AAV target site (15, 16). In addition to the junctions between viral and ch-19 sequences formed within the AAV ITR, virus-host junctions were observed at or near the AAV P5 promoter, as well as between AAV sequences and the EBV episome or unknown DNA sequences (15). In this report, we characterized several AAV–ch-19 junctions cloned from both diploid and aneuploid cell lines. In addition to the previously described junctions formed between the AAV ITR and the target site, we cloned several junctions that involved virus-cell crossovers within the AAV P5 promoter, junctions which parallel those observed previously in the EBV shuttle vector integration system. These results sug-

TABLE 3. In vitro integration junctions from no-end DNA and pRE2

Clone	± Rep[a]	Breakpoint (nt)[b]	
		AAV	pRE2
Derived from PCR amplification with primers H and C			
14.4.30	−	285	2200
14.4.38	−	270	2108
14.6.15	+	111	2285
14.6.21	+	120	2377
14.6.22	+	280	2135
14.6.40	+	120	2174
14.5.3.Z	−	111	2282
14.5.6.Z	−	140	2193
14.5.9.Z[c]	−	115	2395
14.6.2.Z	+	35	2194
14.6.5.Z	+	255	1903
14.6.7.Z	+	248	2021
14.6.15.Z	+	195	1972
14.6.17.Z	+	88	2056
14.6.19.Z	+	73	2200
34.4.5[d]	−	352	2070
34.4.6	−	71	2013
34.4.11[e]	−	245	2384
34.4.12[f]	−	153	2283
Derived from PCR amplification with primers T and C			
15.5.7	−	4566	2042
15.5.8	−	4568	2199
15.5.9[g]	−	4567	2202
15.5.11	−	4564	2201
15.5.12	−	4572	2192

[a] +, Rep protein present during integration; −, Rep protein absent during integration.
[b] Nucleotide locations are based on map locations as described in reference 40.
[c] Contains an insertion of pRE2 nt 105 to 433 joined to pRE2 nt 523 to 466 between the indicated breakpoints.
[d] Contains an insertion of AAV nt 3635 to 3027 between the indicated breakpoints.
[e] Contains an insertion of pRE2 nt 387 to 666 joined to AAV nt 4275 to 4424 and pRE2 nt 995 to 1180 between the indicated breakpoints.
[f] Contains an insertion of pRE2 nt 4542 to 4421 between the indicated breakpoints.
[g] Contains an insertion of AAV nt 2358 to 2639 between the indicated breakpoints.

gest that the integration of wt AAV in the EBV shuttle vector retains many of the characteristics that have been described for the integration of wt AAV at ch-19. Since no other prominent viral sequences involved with cellular junctions were observed, it suggests that the RBS located in the P5 promoter is most likely responsible for the involvement of this DNA at the crossover points. It should also be noted that although the virus-cell junctions appear similar in both systems, 9 of over 40 AAV clones from the EBV shuttle vector contained the complete AAV genome, and only 2 of these exhibited the ability to rescue. This is in contrast to latently wt-AAV-infected cell lines, which exhibit significantly higher frequencies of viral rescue. This suggests that there may be as-yet-unidentified differences between the transient integration and stable latent cell lines.

Regardless, one significant question we addressed in this study is the potential difference in the establishment of viral latency in aneuploid (D6 and HeLa) versus normal diploid (WI-38) cells. Because of the finite lifetime of these cells (50 ± 10 doublings), a mixed population of infected cells was characterized. However, the results (shown in Fig. 1A and 2D) are consistent with data from studies with HeLa cells and the latently AAV-infected cell line Detroit 5, which strongly support the observation that there is no significant difference between wt AAV infection of normal diploid cells and that of aneuploid cells.

Cloning of the rAAV provirus and its structural characterization. We have also cloned complete rAAV proviruses from several individually established latent cell lines. Even though our screen was set up to characterize low-copy-number integrants, in all cases the provirus remained intact, without any of the complex structures observed with the wt virus; e.g., head-to-tail tandem arrays were not present in these cell lines, nor was there rearrangement of the rAAV genome. All of the observed vector-cell junctions were in or near the AAV ITR. Furthermore, none of the cloned rAAV proviruses was located at the ch-19 target site characterized for the wt virus, and none of the junction sequences shared significant homology with ch-19. It must be noted that the latently rAAV-infected cell lines were established in the absence of Rep protein, which has previously been suggested to be required for targeted integration of wt AAV at ch-19 (7, 12, 45). Given the absence of Rep in the establishment of these cell lines, targeted integration of rAAV to ch-19 might not be expected. Although none of these proviruses was located in ch-19, Southern analysis of the parent cell lines indicated that two proviruses were located in the same chromosomal locus on ch-2. This is of interest since an independent observation of vector integration into ch-2 has been reported (20). It is not clear whether this represents a recombination hot spot in aneuploid cell lines or discriminating potential of the vector in the absence of Rep.

The cellular junction sequences obtained from cloned pro-

TABLE 4. In vitro integration junctions from linear AAV substrates and pRE2

Clone (endonuclease)[a]	Breakpoint (nt)	
	AAV	pRE2
AAV (*Sma*I)		
34.10.4	49	2362
34.10.6	49	2333
34.10.9	49	2101
34.10.10	49	2324
39.22.2	49	2109
39.22.3[b]	271	2111
39.22.5	49	2063
AAV (*Xba*I)		
34.13.2	185	2152
34.13.4	185	2230
34.13.8	185	2093
34.13.9	185	2230
39.28.2[c]	185	2257
39.28.6	205	2140
39.28.7	192	1869
39.28.8	236	1867

[a] Junction clones were derived by PCR amplification with primers H and C.
[b] Contains an insertion of AAV nt 460 to 526 between the indicated breakpoints.
[c] Contains an insertion of pRE2 plasmid backbone nt 3879 to 3818 between the indicated breakpoints.

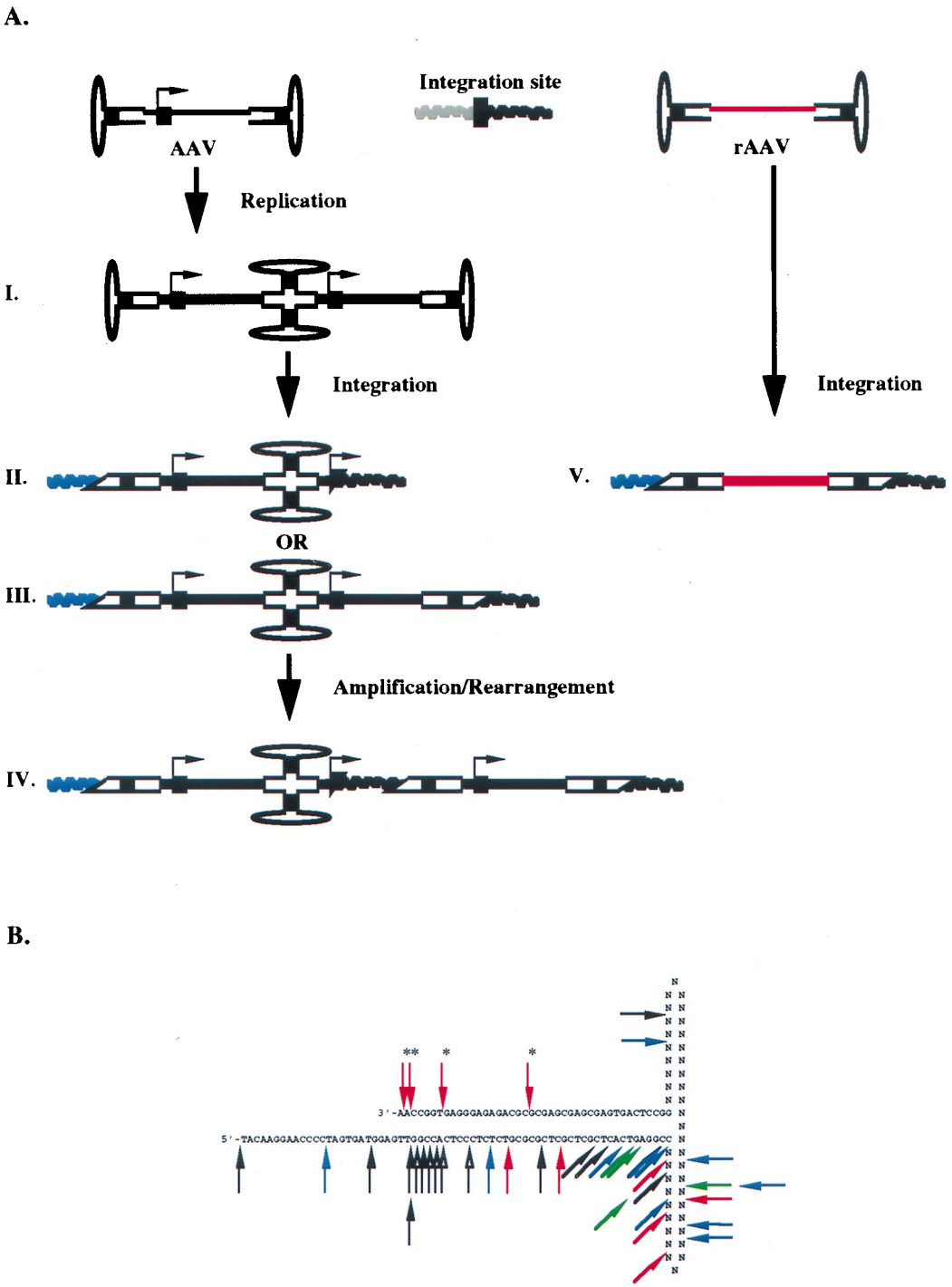


FIG. 9. (A) Scheme outlining the integration of AAV-rAAV. Rep binding sites are indicated by the black boxes on the AAV genome (in the A stem of the ITR or in the P5 promoter, as indicated by the arrow) and at the chromosomal integration site. The pathway on the left describes the integration of wt AAV, in which head-to-tail multimers of AAV may be generated in replication prior to integration (I) or in an amplification-rearrangement step following integration (IV). The integration event may include deletion, duplication, and/or inversion of host DNA sequences, which can occur at either of these two steps. Potential provirus integration structures are also depicted, which include junctions in the P5 promoter (II) or the AAV ITR (III). The pathway on the right outlines the integration of rAAV, in which a single copy of the unrearranged vector is inserted in the genome (V). Rearrangement of the host genome occurs during the integration event. This must also be a pathway for wt AAV, based on results described in reference 15. See the text for further details. This figure is not meant to be exhaustive and may not be complete in detail. (B) Physical map of the locations of junction breakpoints within the putative AAV ITR for in vitro integration junctions (in black), along with junction breakpoints from wt AAV (in red, WI-38 junctions without rearrangements; in green, junctions with insertions at breakpoints) and rAAV (DD-Neo; in blue). Asterisks indicate junctions derived by PCR amplification with the AAV primer TR6. The nucleotides in the B-B′ and C-C′ regions are listed as N, since the purpose of this map is to indicate physically the portion of the ITR remaining at the junction, regardless of the ITR orientation (flip or flop). See the text for a more detailed description.

viruses described in this work were compared to known DNA sequences currently in the GenBank database. BLAST analysis (1) revealed DNA sequence matches for the junctions from three clones: the left junction from DD2-1, the left junction from DD5, and the right junction from DD2-2. All of these sequence matches correspond to cDNA clones obtained from human brain cDNA sequence-tagged sites (21), fetal liver cDNA expressed sequence tags (19), and human ch-7 sequence-tagged sites (9). From these sequence matches, the preintegration sites for the left junction of the provirus in cell line DD2-1, located in ch-17, and the right junction of the provirus in DD2-2, located in ch-7, could be derived (Fig. 8). The locations of these proviruses support the lack of ch-19 targeting but may favor the existence of transcriptionally active regions of the genome. This would parallel what has been suggested for the integration of retroviruses, which have been hypothesized to target integration sites that are transcriptionally active (for a review, see reference 13) and nearby CpG islands (22). These are characteristics that have been noted also for the ch-19 target site for AAV (24). In addition, it is intriguing to note that both wt and rAAV have been determined to integrate in ch-17 (this work and reference 43), although a detailed analysis of the exact provirus locations has not been carried out. Given that the wt virus does not exclusively target the ch-19 site in its establishment of viral latency, this result suggests that ch-17 may be among the secondary target sites for both wt and rAAV located in the human genome.

The vector-provirus junctions characterized in this report were similar to provirus junctions at ch-19 observed with wt AAV (23, 29, 40), in that the junctions contained incomplete ITRs, with vector-cell breakpoints being scattered throughout the AAV ITR and patchy homology between cellular and vector sequences at the junction (Fig. 4 and 9B). These results support the suggestion that both wt AAV and rAAV vectors integrate along a pathway involving nonhomologous recombination, as has been previously proposed (24). In addition, examples of wt AAV inserts that were cloned from the EBV shuttle vector at the ch-19 site (15) possess the same provirus structure as these vectors (i.e., an intact genome, with breakpoints within the AAV ITRs at both ends), suggesting that the wt virus can also integrate intact in a fashion, as is observed for rAAV in this study. The structures of these clones indicate that while one end of the provirus is joined to the ch-19 target sequence, the other end of the provirus is joined to DNA sequence unrelated to the ch-19 preintegration site. These observations are also similar to those of the vector proviruses characterized in this study, in which the left and right junctions of the rAAV provirus are derived from unrelated locations within the host (Fig. 5A and B). This is indicative that there may be significant rearrangement of the host genome occurring during the establishment of latency in rAAV. There is little evidence, however, that significant genomic rearrangement does not occur with wt AAV. It has been reported that rAAV proviruses generated in the absence of selection and in the presence of Rep protein are arranged in head-to-tail tandems, much like those structures seen for wt AAV (14). This suggests that rAAV may establish latency in a manner similar to wt AAV in the presence of Rep protein. These results together are consistent with the integration of both wt AAV and rAAV involving some of the same pathways and host factors during the establishment of viral latency. Although rAAV apparently does not target ch-19 in the absence of Rep, the observed characteristics of the vector-cell integration junctions (breakpoints in the ITRs and microhomology at the crossover) and provirus structure of rAAV have also been observed in its wt counterparts, regardless of the presence or absence of Rep. These results further support the importance of cellular proteins in the integration of AAV and rAAV.

In vitro integration of AAV substrates into the ch-19 target site. Finally, several AAV–ch-19 integration junctions were generated in vitro, using several AAV substrates and the ch-19 target site contained on a plasmid in both the presence and absence of Rep, and were subjected to comparison and analysis. Junctions between ch-19 and AAV substrates containing the AAV ITR (no-end DNA) yielded integration products that resembled AAV–ch-19 junctions formed in vivo both at ch-19 (40) and on an EBV shuttle vector (15). Linear AAV substrates, however, typically exhibited end-to-end joining between target and substrate, utilizing the end of the AAV substrate formed by the restriction enzyme digestion employed to generate the substrates in question. Attempts to obtain junctions from the linear *Pvu*II AAV substrates were unsuccessful. We have no working explanation for this result. Although the cloning procedure used to characterize in vitro junctions (PCR amplification with specific primers) is only capable of selecting for ch-19 target integration events, these results suggest that the AAV ITR hairpin is the authentic structure required for integration of AAV or rAAV. The apparent lack of a requirement for Rep for ch-19 integration may be due in part to the nature of these experiments, in which the ch-19 target is introduced in excess, and the previously mentioned constraint of the cloning procedure selecting only for these integration events. These results are in no way meant to suggest anything with regard to the significance of the role of Rep in targeted integration in vivo, nor do they preclude the potential for integration at other locations in this in vitro system. What may be noteworthy, however, is that the junctions generated on the 3′ no-end substrate generated the majority of breakpoints at the AAV nicking site. This may suggest that all ITRs are not equal. We also observed a bias for the 3′ end of the wt forming the integration junction with the ch-19-specific primer we used. These results may imply that there is polarity to the integration step and that the 3′ end of the viral genome [maybe due to the poly(A) site or RNA transcription] is preferred. In addition, the ability to generate breakpoints at the AAV nicking site in the absence of Rep supports our previously published observation of a host enzyme capable of recognizing and resolving the ITR hairpin structure (47).

The work described in this report indicates that the AAV ITR and as-yet-unknown cellular factors are the only requirements for AAV and rAAV integration. Any model proposed for AAV-rAAV integration needs to address this issue. A potential scheme for AAV-rAAV integration is presented in Fig. 9A. The scheme presented in Fig. 9A outlines several of the steps that may occur during the establishment of AAV-rAAV latency. With wt AAV, there must somehow be targeting to ch-19 along with the generation of head-to-tail tandem arrays during the integration process. This may involve an initial replication step of the virus (Fig. 9A, step II) followed by integration or amplification-rearrangement (Fig. 9A, step IV), with inversions of viral and cellular sequences potentially occurring at either one of these potential steps. The details of these processes likely include steps such as Rep tethering of the substrates through the RBS with strand switching during DNA replication (30, 31, 45). Our results can be accounted for by the model for site-specific AAV integration proposed recently (30, 31). On the other hand, rAAV integration in the absence of Rep appears to follow a different pathway. The resulting proviruses are intact, with breakpoints located within the AAV ITRs (Fig. 9A, step V). As with wt AAV, there appear to be rearrangements of the host genome during the

integration process. Based on results described previously, the wt virus also must follow this pathway of integration, since this provirus structure has been observed (15). This does not indicate, however, that the pathway for the integration of rAAV in the absence of Rep is a major pathway for the integration of its wt counterpart. Our integration studies with AAV plasmids and Rep supplied in *trans* support the occurrence of targeted integration and suggest that more than Rep DNA binding activity is required for this reaction (45a).

Conclusions. The results presented here describe the characterization of the provirus structures of wt and rAAV and the characterization of the structure of junctions derived from wt and rAAV and in vitro integration events. The integration and structure of the wt virus at the ch-19 target site in both aneuploid (HeLa and D6) and primary (WI-38) tissue culture cells appear to be similar. We also have examined the provirus structure of recombinant vectors by direct cloning of what we believe to be the first proviruses isolated from latent cell lines. Although more than one vector had integrated into the same locus (ch-2), this site was not on ch-19, and no other provirus clone had integrated into the wt target site. Although there is no indication of recombinant vectors targeting ch-19 as with the wt, these results indicate that vectors can target the same integration locus and that one of the integration loci characterized has the same chromosomal location as has been seen for the wt virus (ch-17). Finally, in vitro integration of various AAV- and rAAV-related substrates into a plasmid containing the ch-19 target site was performed in 293 cell extracts. These integration products, generated in the presence and absence of Rep protein, suggested that substrates containing the AAV terminal repeat hairpin structure are the authentic integration substrates in vivo.

The locations of the various junction breakpoints within the AAV-rAAV ITR characterized in this report are summarized in Fig. 9B. These results indicate that the junction breakpoints from the different systems examined in this report (wt, recombinant, and in vitro) are similar. Given this similarity in junctions from these different systems, it can be concluded that the AAV ITR hairpin and cellular recombination pathways are the only factors required for the establishment of AAV-rAAV latency. It has previously been suggested that AAV integration into the genome is facilitated by the nonhomologous recombination pathways of the host (23, 24). The in vitro integration products obtained in both the presence and the absence of viral Rep protein and the rAAV provirus junctions characterized support this hypothesis, as no significant difference between in vitro integration junctions produced in the presence or absence of Rep protein can be observed. What cellular factors are involved in the integration of AAV and rAAV vectors, however, cannot be ascertained at this time or from these studies. These studies have provided critical information with regard to determining the steps involved in wt and rAAV integration and should help in the evaluation of the potential of this vector system for human gene therapy.

ACKNOWLEDGMENTS

C.C.Y. and X.X. contributed equally to the work described in this report.

This work was supported by NIH grants HL51818 and DK51880.

We thank Terry Van Dyke and Doug McCarty for critically reading the manuscript.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Antoni, B. A., A. B. Rabson, I. L. Miller, J. P. Trempe, N. Chejanovsky, and B. J. Carter.** 1991. Adeno-associated virus Rep protein inhibits human immunodeficiency virus type 1 production in human cells. J. Virol. **65:**396–404.
3. **Atchison, R. W., B. C. Casto, and W. M. Hammond.** 1965. Adenovirus-associated defective virus particles. Science **149:**754–756.

3a. **Ballard, S. G., and D. C. Ward.** 1993. Fluorescence in situ hybridization using digital imaging microscopy. J. Histochem. Cytochem. **41:**1755–1759.

4. **Berns, K. I.** 1984. Adeno-associated virus, p. 563–592. *In* H. S. Ginsberg (ed.), The adenoviruses. Plenum Press, New York, N.Y.
5. **Berns, K. I., A. Cheung, J. Ostrove, and M. Lewis.** 1982. Adeno-associated virus latent infection, p. 249. *In* B. W. J. Mahy, A. C. Minson, and G. K. Darby (ed.), Virus persistence. Cambridge University Press, Cambridge, United Kingdom.
6. **Berns, K. I., J. Kort, K. H. Fife, E. W. Grogan, and I. Spear.** 1975. Study of the fine structure of adeno-associated virus DNA with bacterial restriction endonucleases. J. Virol. **16:**712–719.
7. **Berns, K. I., and R. M. Linden.** 1995. The cryptic life style of adeno-associated virus. Bioessays **17:**237–245. (Review.)
8. **Bohenzky, R. A., R. B. LeFebvre, and K. I. Berns.** 1988. Sequence and symmetry requirements within the internal palindromic sequences of the adeno-associated virus terminal repeat. Virology **166:**316–327.
9. **Bouffard, G. G., L. M. Iyer, J. R. Idol, V. V. Braden, A. F. Cunningham, L. A. Weintraub, R. M. Mohr-Tidwell, D. C. Peluso, R. S. Fulton, M. P. Leckie, and E. D. Green.** 1997. A collection of 1814 human chromosome 7-specific STSs. Genome Res. **7:**59–64.
10. **Buller, R. M. L., J. E. Janik, E. D. Sebring, and J. A. Rose.** 1981. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. J. Virol. **40:**241–247.
11. **Cheung, A. K. M., 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.
12. **Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. M. 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.
13. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
14. **Duan, D., K. J. Fisher, J. F. Burda, and J. F. Engelhardt.** 1997. Structural and functional heterogeneity of integrated recombinant AAV genomes. Virus Res. **48:**41–56.
15. **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.
16. **Giraud, C., E. Winocour, and K. I. Berns.** 1994. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. Proc. Natl. Acad. Sci. USA **91:**10039–10043.
17. **Goodman, S., X. Xiao, R. E. Donahue, A. Moulton, J. Miller, C. Walsh, N. S. Young, R. J. Samulski, and A. W. Nienhuis.** 1994. Recombinant adeno-associated virus-mediated gene transfer into hematopoietic progenitor cells. Blood **84:**1492–1500. (Erratum, **85:**862, 1995.)
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. **Hillier, L., N. Clark, T. Dubuque, K. Elliston, M. Hawkins, M. Holman, M. Hultman, T. Kucaba, M. Le, G. Lennon, M. Marra, J. Parsons, L. Rifkin, T. Rohlfing, M. Soares, F. Tan, E. Trevaskis, R. Waterston, A. Williamson, P. Wohldmann, and R. Wilson.** 1995. The WashU-Merck EST Project. Unpublished.
20. **Kearns, W. G., S. A. Afione, S. B. Fulmer, M. C. Pang, D. Erikson, M. Egan, M. J. Landrum, T. R. Flotte, and G. R. Cutting.** 1996. Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. Gene Ther. **3:**748–755.
21. **Khan, M. S., D. J. Hryb, G. A. Hashim, N. A. Romas, and W. Rosner.** 1990. Delineation and synthesis of the membrane receptor-binding domain of sex hormone-binding globulin. J. Biol. Chem. **265:**18362–18365.
22. **Kitamura, Y., Y. M. Lee, and J. M. Coffin.** 1992. Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. Proc. Natl. Acad. Sci. USA **89:**5532–5536.
23. **Kotin, R. M., and K. I. Berns.** 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. Virology **170:**460–467.
24. **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.
25. **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.
26. **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.
27. **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.
28. **Laughlin, C. A., J. D. Tratschin, H. Coon, and B. J. Carter.** 1983. Cloning of

infectious adeno-associated virus genomes in bacterial plasmids. Gene **23:** 65–73.
29. **Leonard, C. J., and K. I. Berns.** 1994. Adeno-associated virus type 2: a latent life cycle. Prog. Nucleic Acid Res. Mol. Biol. **48:**29–52. (Review.)
30. **Linden, R. M., P. Ward, C. Giraud, E. Winocour, and K. I. Berns.** 1996. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA **93:**11288–11294.
31. **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.
32. **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.
33. **Muro-Cacho, C. A., R. J. Samulski, and D. Kaplan.** 1992. Gene transfer in human lymphocytes using a vector based on adeno-associated virus. J. Immunother. **11:**231–237.
34. **Nahreini, P., S. H. Larsen, and A. Srivastava.** 1992. Cloning and integration of DNA fragments in human cells via the inverted terminal repeats of the adeno-associated virus 2 genome. Gene **119:**265–272.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
36. **Samulski, R. J.** 1993. Adeno-associated virus: integration at a specific chromosomal locus. Curr. Opin. Genet. Dev. **3:**74–80. (Review.)
37. **Samulski, R. J., K. I. Berns, M. Tan, and N. Muzyczka.** 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:**2077–2081.
38. **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.
39. **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.
40. **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, **11:**1228, 1992.)
41. **Snyder, R. O., R. J. Samulski, and N. Muzyczka.** 1990. In vitro resolution of covalently joined AAV chromosome ends. Cell **60:**105–113.
42. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
43. **Walz, C., and J. R. Schlehofer.** 1992. Modification of some biological properties of HeLa cells containing adeno-associated virus DNA integrated into chromosome 17. J. Virol. **66:**2990–3002.
44. **Ward, P., and K. I. Berns.** 1991. In vitro rescue of an integrated hybrid adeno-associated virus/simian virus 40 genome. J. Mol. Biol. **218:**791–804.
45. **Weitzman, M. D., S. R. Kyostio, R. M. Kotin, and R. A. Owens.** 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:**5808–5812.
45a. **Xiao, W., and R. J. Samulski.** Submitted for publication.
46. **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.
47. **Xiao, X., W. Xiao, J. Li, and R. J. Samulski.** 1997. A novel 165-base-pair terminal repeat sequence is the sole *cis* requirement for the adeno-associated virus life cycle. J. Virol. **71:**941–948.