JOURNAL OF VIROLOGY, Oct. 1997, p. 7951–7959 Vol. 71, No. 10
0022-538X/97/$04.00+0

Adeno-Associated Virus Rep Proteins Target DNA Sequences to a Unique Locus in the Human Genome

RICHARD T. SUROSKY,[1]* MASASHI URABE,[2] SIMON G. GODWIN,[1] SUSAN A. MCQUISTON,[1] GARY J. KURTZMAN,[1] KEIYA OZAWA,[2] AND GEORGES NATSOULIS[1]†

Avigen, Inc., Alameda, California 94502,[1] and Department of Molecular Biology, Institute of Hematology, Jichi Medical School, Tochigi 329-04, Japan[2]

Received 24 March 1997/Accepted 23 June 1997

We have developed a system for site-specific DNA integration in human cells, mediated by the adeno-associated virus (AAV) Rep proteins. In its normal lysogenic cycle, AAV integrates at a site on human chromosome 19 termed AAVS1. We describe a rapid PCR assay for the detection of integration events at AAVS1 in whole populations of cells. Using this assay, we determined that the AAV Rep proteins, delivered in *cis* or *trans*, are required for integration at AAVS1. Only the large forms of the Rep protein, Rep78 and Rep68, promoted site-specific integration. The AAV inverted terminal repeats, present in *cis*, were not essential for integration at AAVS1, but in cells containing Rep, they increased the efficiency of integration. In the presence of the Rep proteins, the integration of a plasmid containing AAV inverted terminal repeats occurred at high frequency, such that clones containing the plasmid could be isolated without selection. In two of the five clones analyzed by fluorescence in situ hybridization, the plasmid DNA was integrated at AAVS1. In most of the clones, at least one copy of the entire plasmid was integrated in a tandem array. Detailed analysis of the integrated plasmid structure in one clone suggested a complex mechanism producing rearrangements of the flanking genomic DNA, similar to those observed with wild-type AAV.

A number of methods have been developed for integrating DNA at a specific location in the human genome. Site-specific integration can be achieved by homologous recombination using the endogenous cellular machinery. Unfortunately, this system is inefficient, and elaborate selection schemes are often required to identify cells containing an integrated sequence (1, 13). Heterologous site-directed recombination systems, such as those using the yeast FLP recombinase (41) or the bacterial Cre recombinase (48), function in mammalian cells and can be more efficient. The drawback of these systems is that the target site does not preexist in the human genome and must be introduced. A system combining the efficiency of site-directed recombination with an endogenous target sequence would have many applications for research and gene therapy in human cells.

Adeno-associated virus (AAV) is a human parvovirus with a single-stranded DNA genome of 4.7 kb (reviewed in references 24 and 38). The genome consists of two open reading frames encoding the Rep and Cap proteins, flanked at each end by 145-base inverted terminal repeats (ITRs). The ITRs are the only sequences required in *cis* for viral replication, packaging, and integration (18, 36, 45). In addition to the ability to form an extensive stem-loop structure, each ITR contains binding sites for the AAV Rep proteins, the Rep binding site (RBS) (43, 50) and a sequence recognized by the Rep protein nicking activity (the terminal resolution site [TRS]) (21). The RBS consists of tandem repeats of the tetramer GAGC (11, 35). The *cap* gene encodes three capsid proteins that form the viral particle (38). The *rep* gene encodes four overlapping proteins: the two large, Rep78 and Rep68, and the two small, Rep50 and Rep42, proteins.

* Corresponding author. Mailing address: Avigen, Inc., 1201 Harbor Bay Parkway, Suite 1000, Alameda, CA 94502. Phone: (510) 748-7174. Fax: (510) 748-7155. E-mail: surosky@avigen.com.

† Present address: Microcide Pharmaceuticals, Inc., Mountain View, CA 94043.

The Rep proteins are essential for viral replication, integration, and rescue of the provirus. In vitro studies of the Rep proteins have identified a sequence-specific binding activity (11, 55), a sequence-specific endonuclease activity (21), an ATP-dependent helicase (21), an ability to regulate transcription from AAV and other promoters (4, 29, 30, 57), and an ability to stimulate replication of AAV sequences (39, 54). These activities reside in the Rep78 and Rep68 proteins. The analysis of Rep binding by random oligonucleotide selection demonstrated that although fragments containing repeats of GAGC were preferentially bound, any fragment containing a single GAGC tetramer followed by several G residues would be bound by Rep at lower affinity (12).

AAV is capable of both latent and lytic infection. In the presence of a helper virus, such as adenovirus or herpes simplex virus, AAV initiates a productive viral infection. Prokaryotic clones of wild-type AAV virus are also infectious. When an AAV-containing plasmid is transfected in the presence of helper virus, AAV excises and the normal lytic cycle begins (32, 44). In the absence of helper virus, AAV integrates at a specific site on chromosome 19, termed AAVS1 (27, 28, 47). The provirus can excise and replicate upon subsequent introduction of a helper virus (10, 31, 36). The AAVS1 region was originally isolated as an 8.2-kb fragment (26), but recently Linden et al. (33) have defined a 33-base sequence as the minimum required for AAVS1 function. Interestingly, this sequence contains a RBS and a TRS very similar to those identified within the AAV ITRs, and mutations in either of these two elements in AAVS1 abolish site-specific integration.

The mechanism of AAV integration is not known, although putative intermediates and proviral structures have been analyzed. In vitro binding studies have demonstrated that the Rep proteins can mediate complex formation between the AAV ITR and a 109-bp sequence from AAVS1 containing the RBS and TRS (55). Proviral structures have been recovered from latently infected cell lines and from Epstein-Barr virus-based shuttle plasmids containing the AAVS1 region, and analysis

has revealed several consistencies. One junction in the genomic DNA was frequently in or adjacent to the RBS within AAVS1, while the other genomic DNA junction did not involve AAVS1 sequences (15). The genomic DNA flanking the insertion point was sometimes rearranged so as to produce deletions, duplications, or translocations (15, 26, 36). With respect to the virus, at least one of the junctions was within or adjacent to the ITR or one of the *rep* promoter regions (15, 36). Tandem copies of the virus were often observed (10, 31, 36). The complexity of these structures suggests a multistep mechanism for AAV integration.

A central role for the Rep proteins in AAV site-specific integration is suggested by experiments using recombinant virus. Examination of 22 latently infected cell lines by Kotin et al. (28) indicated that wild-type AAV integrated at AAVS1 in approximately 70% of the lines. In a bronchial epithelial cell line infected by AAV, 94% of the integration events were in chromosome 19 (22). In contrast, a recombinant AAV (rAAV) vector in which the *rep* and *cap* genes had been deleted was still capable of integration but did not integrate at AAVS1 (22). In addition, rAAV or plasmids carrying an AAV provirus in which the viral *cap* gene had been deleted, but retained Rep activity, were still capable of integration at AAVS1, demonstrating that the Cap proteins are dispensable for site-specific integration (49).

Based on these results, we speculated that it would be possible to achieve site-specific integration by using two plasmids: an integrating plasmid containing AAV ITRs and the second providing the AAV Rep proteins in *trans*. The advantage to this system is that no replicating virus is used and only the AAV ITRs, but no viral coding sequences, are integrated. In addition, because no virus is packaged, there may be no constraints on the size of the integrating sequence. We first describe a rapid and sensitive PCR-based assay for integration at AAVS1 in whole populations of cells. Using this assay, we detected site-specific integration with viral and plasmid constructs in the presence of Rep. Individual clones containing integrated AAV sequences were isolated without selection and characterized. Site-specific integration of the AAV sequences was demonstrated by using fluorescence in situ hybridization (FISH). Finally, the structure of one of the integrated sequences was determined, and a number of similarities with AAV proviral structures were observed.

MATERIALS AND METHODS

The nucleotide coordinates used to describe the AAV clones refer to the AAV sequence determined by Srivastava et al. (51).

Cell lines, vectors, and plasmids. The 293 cell line was derived from human embryonic kidney cells transformed with sheared human adenovirus type 5 DNA (American Type Culture Collection, Rockville, Md.). 293/WTAAV, containing AAV integrated at AAVS1, was constructed by infecting 293 cells with wild-type AAV. Random clones were isolated and tested for site-specific integration in a PCR assay (see below). *Escherichia coli* DH5α was used as a host for all bacterial transformations (17). The structure and encapsidation of rAAV, in which the viral coding sequence has been replaced by a cytomegalovirus (CMV)-*lacZ* fusion gene, and infection of virus into 293 cells have been described elsewhere (23).

pAAV, identical to psub201 (46), is a plasmid carrying a replication-competent AAV provirus. Two mutant derivatives of pAAV were constructed. pAAVM1 contains a deletion in the *cap* region between the *Apa*I sites at positions 2943 and 4040. pAAVM2 was generated by digesting pAAV with *Bam*HI, end filling with Klenow enzyme, and ligating with T4 ligase, resulting in the insertion of four extra bases. This frameshift mutation affects all four forms of Rep.

Plasmid pRepCap contains the entire AAV coding region without the ITRs cloned into the *Not*I site of pBSII (Stratagene, La Jolla, Calif.). The insert in this plasmid is identical to that of pAAVAd (45). pRep contains the AAV coding region of pAAVM1 without the ITRs cloned into a pBSII derivative. In pRSVRepCap, the p5 promoter has been replaced with the Rous sarcoma virus (RSV) promoter. Specifically, a fragment spanning positions +310 to +4484 of AAV was inserted downstream of the RSV promoter in place of the chloramphenicol acetyltransferase gene in the expression plasmid pORSVCAT (Stratagene). p2ITRlacZ is a derivative of psub201 (46) in which the *Xba*I fragment carrying *rep* and *cap* was replaced with an expression cassette fusing the CMV promoter to *lacZ*. The CMV promoter consists of a 654-bp *Spe*I-*Sac*II fragment from −584 to +70 of the CMV immediate-early gene (7). The *lacZ* gene was obtained as a *Sma*I-*Dra*I fragment from pCMVβ (Clontech, Palo Alto, Calif.). p1ITRlacZ is a derivative of p2ITRlacZ in which a 250-bp ITR-containing fragment was deleted. This plasmid retains a single ITR 5′ of the CMV promoter. The CMV-*rep* gene was constructed by fusing the *Spe*I-*Sac*II fragment containing the CMV promoter (see above) with an AAV fragment from +310 to +2943 containing the *rep* gene and the 5′ end of the *cap* gene. The *PGK-rep* gene was constructed by fusing a 534-bp *Avr*II-*Pst*I fragment containing the *PGK* promoter to an AAV fragment from positions +310 to +4484. A portion of the *cap* gene was then deleted by removing the sequence between the *Apa*I sites at positions 2943 and 4040. The dihydrofolate reductase (DHFR) fragment used in the transfection studies was a 370-bp *Afl*III-*Bam*HI fragment. The *lacZ* fragment used in the transfection studies was a 1,113-bp *Cla*I-*Sac*I fragment.

Plasmids that express individual Rep proteins were constructed by PCR-based mutagenesis (20). All four Rep proteins are translated from the same open reading frame. Therefore, to express only one of the large Rep proteins (Rep78 or Rep68), the ATG at position 993, corresponding to the first methionine of the small Rep proteins (Rep52 and Rep40), was changed to GGA. Moreover, to express only Rep78 or Rep52, derived from the unspliced mRNA, the splice donor site was destroyed by changing AGG to CGC at position 1905. Rep68 and Rep42, derived from the spliced mRNA, were expressed from a construct containing a deletion of the entire intron sequence from positions 1907 to 2227. The mutagenized *rep* genes were expressed by using the CMV promoter (see above) or the native AAV promoters (p5 for Rep78 or Rep68 and p19 for Rep52 or Rep40).

Probes. The AAVS1 probe was the 3.5-kb *Eco*RI-*Kpn*I fragment from AAVS1 (26). The *lacZ* probe was the 4.5-kb *Not*I fragment from p2ITRlacZ containing the CMV-*lacZ* cassette. The plasmid backbone probe was the 1.9-kb *Afl*III-*Eco*RI fragment from pBR322 (6). The *rep* probe was the 4.1-kb *Xba*I-*Xba*I fragment from psub201; it corresponds to bp 310 to 4484 of the AAV genome.

Primers. The AAVS1 primers were 79 (5′-ACTTTGAGCTCTACTGGCT TC-3′), 80 (5′-GGAGGATCCGCTCAGAGG-3′), and 107 (5′-CGGGGAGG ATCCGCTCAGAGGTACA-3′). Primers 81 (5′-AGGAACCCCTAGTGATG GAGT-3′) and 100 (5′-CGGCCTCAGTGAGCGAGCGAGC-3′) were derived from the d and a regions of the AAV ITR, respectively (51). The plasmid backbone primer was 5′-CGCCAGGGTTTTCCCAGTCACGAC-3′. The DHFR primer was 5′-GGGGGTACCCAAATCGTGCATGCCGTCTAT-3′, and the *lacZ* primer was 5′-CCCGACCTCTCTATCGTGCG-3′. The two primers used in the inverted PCR to recover the preintegration site were 5′-GGGTGAGGAAAGT CAAGAT-3′ and 5′-GCCCCATATGGTTTCAATG-3′.

DNA and cell manipulations. Plasmid and mammalian DNA purification, DNA sequencing, bacterial transformation, and Southern blot analysis were performed as described previously (3). Hirt supernatants were prepared as described elsewhere (19). Cells were assayed for β-galactosidase activity by trypsinizing and replating at very low (<5%) density. Cells were allowed to reattach for 12 h prior to staining. Detection of β-galactosidase by 5-bromo-4-chloro-3-indolyl-β-D-galactoside staining and *o*-nitrophenyl-β-D-galactopyranoside assays were performed as described previously (3). Cells were transfected by the $CaPO_4$ method (56) and infected with wild-type AAV or rAAV as described elsewhere (23). In experiments in which both plasmid and viral DNAs were introduced, the cells were first transfected with the plasmid. After 8 h, the medium was changed and the virus was added.

DNA fragments used for transfection were isolated after digestion of 1 μg of plasmid DNA with the appropriate restriction enzyme and fractionation on an agarose gel. The DNA fragment was cut out of the gel and extracted by using Geneclean (Bio101, Vista, Calif.). Single-cell clones were isolated by flow cytometry on a FACS Vantage (Becton Dickinson, San Jose, Calif.). Inverse PCR was performed as described previously (40) after digestion of the genomic DNA with the enzyme *Pst*I and ligation under dilute conditions. The PCR products were cloned by using a TA cloning kit (Invitrogen, San Diego, Calif.). FISH was performed as described previously (52). Probes were labeled by the random priming method (14).

PCR based assay for site-specific integration. For each assay, 5×10^5 cells were either infected with AAV or transfected with 1 μg of plasmid DNA. After 2 days in culture, total genomic DNA was extracted from the pool of cells. An aliquot of DNA corresponding to 10^4 human genomes was used in a PCR. The PCR mixture also contained 1× *Taq* buffer (10 mM Tris Cl [pH 8.3], 1.5 mM $MgCl_2$, 50 mM KCl), 2 μM each primer, 200 μM deoxynucleoside triphosphate, and 1 U of *Taq* polymerase. The cycling conditions were 94°C for 3 min, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 3 min. The PCR mixture was transferred to a Zeta Probe membrane (Bio-Rad, Hercules, Calif.) by using a dot blot apparatus and hybridized to an AAVS1 probe according to the manufacturer's protocol. The probe was a random-primed ^{32}P-labeled PCR fragment (14) generated in a reaction with 0.1 μg of human genomic DNA as template and primers 79 and 80. Amplification conditions were as described above. The cycling conditions were 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, repeated 35 times.

Immunoblotting analysis. Two days after transfection, approximately 10^6 cells were trypsinized and pelleted. Cells were lysed in 50 mM Tris (pH 7.5)–1 mM EDTA–0.15 M NaCl–1% sodium deoxycholate–0.1% sodium dodecyl sulfate–1% Nonidet P-40–1 mM phenylmethylsulfonyl fluoride. Protein concentrations were determined by the bicinchoninic acid protein assay (Pierce, Rockford, Ill.), and 1 to 10 μg of protein was analyzed. Proteins were fractionated on a 10% polyacrylamide gel (3). The protein was transferred to a nitrocellulose membrane by using an electroblotting apparatus (Hoefer, San Francisco, Calif.), and the membrane was blocked with 5% nonfat milk (3). The membrane was incubated with a mouse monoclonal antibody directed against AAV Rep proteins (American Research Products, Belmont, Mass.) followed by anti-mouse immunoglobulin-linked horseradish peroxidase antibody (Amersham, Arlington Heights, Ill.). Chemiluminescence detection was performed with the ECL system (Amersham).

Recovery of integrated plasmid DNA. Ten micrograms of genomic DNA was digested with the restriction enzyme *Hin*dIII and fractionated on an agarose gel. All DNA fragments in the 10- to 25-kb size range were cut out of the gel and extracted by the phenol method (34). The DNA was ligated and used to transform bacteria, selecting for carbenicillin resistance. Comparison of restriction maps of the recovered plasmid and of the genomic region containing the integrated plasmid verified that no rearrangements had occurred during recovery.

RESULTS

Site-specific integration of AAV sequences is Rep dependent. A rapid assay for integration at AAVS1 was developed for identifying constructs capable of site-specific integration (Fig. 1A). The template for this PCR-based assay was genomic DNA extracted from an unselected pool of cells that had been infected or transfected with AAV-derived constructs. One primer from the AAVS1 region and a second primer from the integrating DNA were used for amplification. AAV integration occurs in a region spanning at least 100 bp in AAVS1, and so we expected the PCR products to be heterogeneous in size (15, 47). Therefore, all specific amplification products, irrespective of size, were detected by dot blot hybridization using an AAVS1-derived probe situated immediately adjacent to the AAVS1-specific primer.

Analysis of 293 cells infected by wild-type AAV at different multiplicities of infection (MOIs) was used to verify the utility of this assay (Fig. 1B). Primers from different portions of the ITR (81 and 100) in combination with AAVS1-specific primers (80 and 107) produced positive signals. No signal was detected from either infected or noninfected cells with only the AAVS1 primer (primer 80) (data not shown). A faint signal was detected with a single ITR-specific primer (primer 81) only in infected cells. This signal may have resulted from two integration events on either side of the AAVS1 region that is used as a probe. The limits of detection of this assay were determined by a mixing experiment using uninfected 293 cells and 293/WTAAV cells containing an AAV integration at AAVS1 which produced a positive signal in the PCR assay. These cells were mixed in various proportions, but the total number of genomes assayed was kept constant at 10^4. A positive signal could readily be detected with as few as 0.1% (10 in 10^4) 293/WTAAV genomes (Fig. 1C).

An rAAV vector in which all viral sequences except the ITRs had been replaced by a CMV-*lacZ* fusion gene was examined for the ability to integrate at AAVS1. No evidence for site-specific integration was observed when this vector was infected at an MOI of 10^5 (Fig. 1B). However, in cells transfected with the *rep*-expressing plasmid pRep or pRepCap and subsequently infected with the rAAV vector at an MOI of 10^5, site-specific integration was readily detected. This finding indicated that the *rep* gene delivered in *trans* could direct the integration of the rAAV vector to AAVS1.

AAV sequences contained in plasmids were then analyzed for site-specific integration. First, the AAV provirus pAAV was transfected and produced a positive signal (Fig. 1B). Two derivatives of pAAV each containing a mutation in one of the

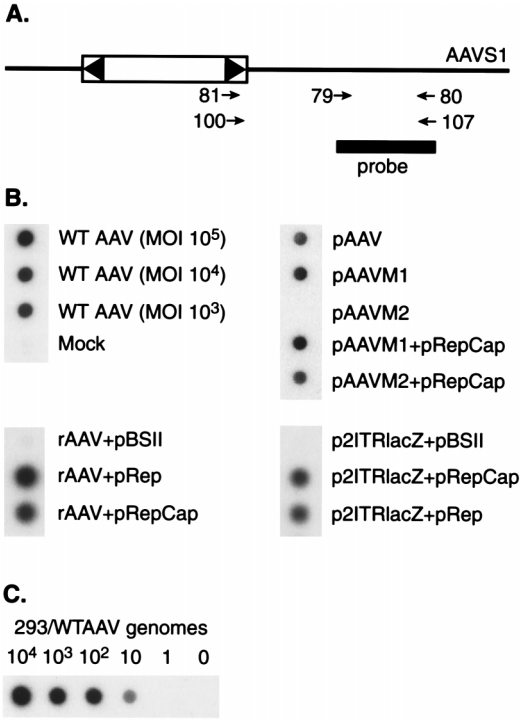

FIG. 1. PCR assay for the detection of integration at AAVS1. (A) Diagram showing the components of the assay. AAV or any ITR-bound sequence is integrated at AAVS1. The open box indicates the AAV coding region or other marker gene. The boxed arrowheads represent the AAV ITRs. The primers used in the PCR analysis are indicated by the arrows with the associated numbers. Primers 81 and 100 are derived from the AAV ITR. Primers 79, 80, and 107 are derived from AAVS1. The probe, generated by using primers 79 and 80, spans a 297-bp region in AAVS1. (B) Site-specific integration of various DNA molecules. 293 cells were infected and/or transfected with the viruses or plasmids indicated. The structures of the transfected DNAs are described in Materials and Methods. Three different MOIs of wild-type AAV were tested. The rAAV was infected at an MOI of 10^5. All MOIs are expressed as the number of single-stranded genomes per cell, and all viral preparations were assayed by quantitative dot blot hybridization. One microgram of plasmid DNA was used in the transfections. Genomic DNA was prepared, and 10^4 genomes were used as a template in a PCR with an AAV-derived and AAVS1-derived primer (80-81 or 100-107). The products were immobilized on a filter membrane and hybridized with the AAVS1 probe. WT, wild type. (C) Mixing experiments to determine the sensitivity of the PCR assay. Various ratios of 293 cells and 293/WTAAV cells, in which AAV is integrated at AAVS1, were mixed and DNA was prepared from a total of 10^4 cells. The number of 293/WTAAV cells in each sample is indicated.

AAV open reading frames were tested. A 1-kb deletion in the *cap* gene did not prevent site-specific integration of the resulting provirus (pAAVM1) (Fig. 1B). In contrast, a frameshift mutation in the *rep* gene, which affected all four Rep proteins (pAAVM2), abolished site-specific integration. Site-specific integration of pAAVM2 could be restored by cotransfection with plasmid pRepCap. These experiments indicated that the Rep but not the Cap proteins are required for site-specific integration. The ability of the Rep proteins to direct plasmid integration at AAVS1 was also demonstrated by transfection experiments with two plasmids: one containing a CMV-*lacZ* fusion gene flanked by AAV ITRs (p2ITRlacZ); and either pRep or pRepCap. Site-specific integration of plasmid p2ITRlacZ was observed when cells were transfected with pRep or pRSVRepCap, but upon transfection with the negative control plasmid pBSII, integration at AAVS1 was not observed (Fig. 1B).

Rep increases the number of cells showing long-term expression. The PCR assay demonstrated site-specific integration but

TABLE 1. Effect of the Rep protein on long-term expression from an ITR-containing plasmid

Plasmids	% β-Galactosidase-positive cells[a] at indicated day after transfection							
	2	5	8	13	20	27	35	41
p2ITRlacZ + pBSII	95.1	42.0	10.6	1.3	0.7	1.9	0.5	1.4
p2ITRlacZ + pRSVRepCap	95.6	46.8	11.2	8.2	5.0	5.9	6.6	5.0

[a] An aliquot of 293 cells from the transfected pool was removed and stained for β-galactosidase activity. For each sample, 300 cells were examined.

TABLE 2. Effects of different *rep* constructs on long-term expression and site-specific integration

rep plasmid	% β-Galactosidase-positive cells[a]	PCR assay
pBSII	0.5	−
pRep (wild type)	4.7	+
pCMVRep	3.3	+
pRSVRepCap	6.2	+
pPGKRep	5.3	+

[a] Plasmid p2ITRlacZ was cotransfected with the plasmids indicated, and cells were assayed for β-galactosidase expression 3 weeks after transfection. For each sample, 300 cells were examined. Each value is an average of two independent experiments.

provided no information about expression of the integrated DNA. Therefore β-galactosidase expression was examined after cotransfection of p2ITRlacZ with pRSVRepCap or the control plasmid pBSII. Mixing experiments using β-galactosidase-positive and -negative cells showed no changes in ratios after prolonged culturing, demonstrating that β-galactosidase expression has no effect on growth rate (data not shown). After 2 days, 95% of the cells were β-galactosidase positive in both pools, but this number declined rapidly (Table 1). After 13 days, in the pool of cells not receiving *rep*, 2% or less of the cells were β-galactosidase positive. In contrast, in the cells transfected with pRSVRepCap, 5 to 10% were β-galactosidase positive, and this level was maintained for at least 41 days. This increase most likely resulted from more efficient integration of the *lacZ* gene in the *rep*-transfected population, yielding a higher percentage of stable β-galactosidase-expressing cells. Alternatively, the Rep proteins may somehow have stabilized the p2ITRlacZ episomal DNA.

The effects of different Rep protein levels on long-term β-galactosidase expression were evaluated by using plasmids containing the *rep* gene in which the wild-type p5 promoter had been replaced by various heterologous promoters. First, the levels of Rep proteins produced by these fusion genes were determined 2 days after transfection by Western blot analysis (Fig. 2). The wild-type gene on plasmid pRepCap produced primarily the smaller, Rep52 and Rep40, proteins (lane 2), with lesser amounts of the larger, Rep78 and Rep68, proteins, a profile similar to that obtained from the wild-type virus (lane 6). In contrast, all of the *rep* genes with heterologous promoters (CMV, RSV, and *PGK*) produced the large Rep proteins in greater abundance. pCMVRep (lane 5) yielded the highest amount of large Rep proteins (this lane contains one-fifth the total protein of the other lanes), followed by pRSVRepCap (lane 4) and pPGKRep (lane 3). In a time course analysis of Rep protein expression from all constructs, production peaked at 2 days after infection or transfection and then slowly declined (data not shown).

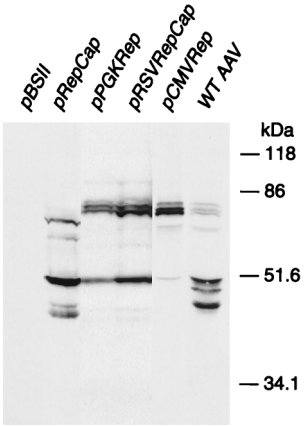

FIG. 2. Western blot analysis of Rep expression from different constructs. Wild-type (WT) AAV (MOI of 10^5) was infected or *rep*-containing plasmids (1 μg) were transfected into 293 cells, and after 2 days cell lysates were prepared. Each lane contains 10 μg of protein except for the pCMVRep sample, which contains 2 μg of protein. Cells transfected with pBSII were used as a negative control. Positions of size markers are indicated at the right.

In transfection experiments with p2ITRlacZ and plasmids containing these *rep* constructs, little difference in long-term β-galactosidase expression was observed (Table 2). All *rep* constructs yield between 3.3 and 6.2% β-galactosidase-positive cells. A positive signal from the PCR assay was observed when p2ITRlacZ was transfected with each of the *rep* constructs (Table 2). In agreement with these results were those of an experiment in which the level of the p2ITRlacZ plasmid was fixed (1 μg) and the amount of pRSVRepCap was varied (0.1, 1.0, and 10.0 μg): no difference in long-term β-galactosidase expression was observed (data not shown). The results suggest that the levels of Rep proteins produced in these experiments do not dramatically affect long-term expression from an ITR-containing plasmid.

The contribution of the individual Rep proteins to the site-specific integration reaction was examined by using constructs engineered to express only one of the Rep proteins. Western blot analysis verified that each construct expresses a single Rep protein (Fig. 3A). Each of these plasmids was transfected with plasmid p2ITRlacZ, and site-specific integration was analyzed by using the PCR assay (Fig. 3B). Both of the large Rep proteins could mediate integration at AAVS1, whether expressed from the CMV or from the p5 promoter. In contrast, no evidence of site-specific integration was observed with the small Rep proteins.

***cis*-acting sequences and site-specific integration.** The importance of the ITR sequences for site-specific integration and long-term expression was examined. Plasmids with one or two ITRs behaved identically in the PCR and long-term expression assays. We constructed a derivative of plasmid p2ITRlacZ which was identical to the parental plasmid except that one of the two ITRs was deleted (p1ITRlacZ). After cotransfection with pRSVRepCap, each plasmid produced a signal in the PCR assay and similar percentages of blue cells in long-term β-galactosidase assays (Fig. 4 and Table 3). Surprisingly, a control plasmid containing no ITR sequences (p0ITRlacZ) yielded a signal in the PCR assay after cotransfection with pRSVRepCap when primers from the plasmid backbone and AAVS1 were used (Fig. 4). In addition, gel-purified fragments of 370 bp from the DHFR gene or 1,113 bp from the *lacZ* gene also produced a signal when the DHFR-AAVS1 or *lacZ*-AAVS1 primer set was used, respectively. The signal was observed only when the fragments were transfected with a

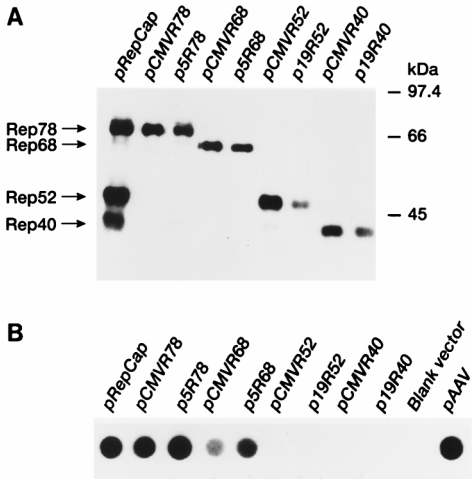

FIG. 3. (A) Western blot analysis of the individual Rep-expressing plasmids. The CMV promoter-driven (pCMVR78, pCMVR68, pCMVR52, and pCMVR40), p5 promoter-driven (p5R78 and p5R68), and p19 promoter-driven (p19R52 and p19R40) constructs were transfected into 293 cells, and after 2 days lysates were prepared. Plasmid pRepCap, which expresses all four Rep proteins, was used as a control. Five micrograms of protein prepared from cells transfected with the p5- or p19-driven constructs or 1 μg of protein prepared from cells transfected with the CMV-driven constructs was loaded in each lane. (B) Detection of integration at AAVS1. Plasmid p2ITRlacZ and the *rep* expression plasmids were cotransfected into 293 cells. Genomic DNA prepared from the cells was analyzed as described for Fig. 1B. Cells cotransfected with pAAV and p2ITRlacZ were used as a positive control. Cells cotransfected with a plasmid lacking a *rep* expression cassette and p2ITRlacZ were used as a negative control.

rep-containing plasmid. In the absence of Rep proteins, a positive signal was never observed. Mixing experiments demonstrated that integration of non-ITR-containing sequences at AAVS1 was a rare event, occurring 10- to 100-fold less frequently than with ITR-containing sequences (data not shown). Differences were observed between plasmids with and without ITRs in long-term β-galactosidase assays. A 10-fold increase in β-galactosidase-positive cells was obtained after cotransfection of pRepCap and p2ITRlacZ or p1ITRlacZ compared to pRepCap and p0ITRlacZ (Table 3). Thus, the ITR was not essential for site-specific integration but did increase the percentage of cells showing long-term expression when transfected with a *rep*-expressing plasmid.

Analysis of clones expressing β-galactosidase. Cells expressing β-galactosidase were analyzed in greater detail after the isolation of single-cell-derived clones. The clones were isolated without selection from two independent pools of cells 4 days after cotransfection with plasmids pRSVRepCap and p2ITRlacZ, using flow cytometry to isolate single cells. After

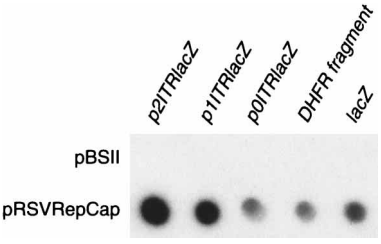

FIG. 4. *cis*-acting sequences required for site-specific integration. Plasmids or DNA fragments were transfected with either control plasmid pBSII or pRSVRepCap. The PCR assay was carried out as described for Fig. 1B. One primer used in the assay was 107 from AAVS1, and the second was specific for the integrating DNA (described in Materials and Methods).

TABLE 3. Long-term expression from plasmids with zero, one, and two AAV ITRs

lacZ plasmid	% β-Galactosidase-positive cells[a]	
	pBSII	pRSVRepCap
p0ITRlacZ	1.0	0.3
p1ITRlacZ	0.4	7.4
p2ITRlacZ	0.7	6.2

[a] The *lacZ*-containing plasmids were cotransfected with either the pBSII control plasmid or pRSVRepCap. Cells were assayed for β-galactosidase activity 3 weeks after transfection. For each sample, 300 cells were examined. Each value is an average of two independent experiments.

expansion, an aliquot from each clone was examined for β-galactosidase expression. Four of seven and three of 22 clones obtained from the two pools were positive for β-galactosidase expression. Note that the 57% positive cells from the first pool was unusually high compared to the percentage typically obtained (Table 1). A pool of cells transfected with p2ITRlacZ and pBSII yielded no β-galactosidase-positive clones from a total of 22 tested. Among the positive clones, the expression level of β-galactosidase was stable even after 2 months of growth, as determined by *o*-nitrophenyl-β-D-galactopyranoside assays (data not shown).

DNA hybridization was used to analyze the seven β-galactosidase-positive clones. Genomic DNA from each clone was digested with restriction enzyme *Hin*dIII, an enzyme that has no recognition site in AAVS1 or plasmid p2ITRlacZ. The DNA was hybridized with probes from the AAVS1 region, the *rep* gene, the CMV-*lacZ* cassette, and the plasmid backbone from p2ITRlacZ. The seven clones fell into two categories (Table 4). Six of the clones contained no *rep* sequences but hybridized to both the CMV-*lacZ* and plasmid backbone probes, suggesting that the provirus did not excise from the plasmid prior to integration. This was confirmed by an additional digest with *Spe*I, which cuts once within plasmid p2ITRlacZ, followed by hybridization with the CMV-*lacZ* probe. All but one of the clones produced a band identical in size to plasmid p2ITRlacZ, indicating that at least one entire copy of the plasmid had been integrated (data not shown). The six clones also showed at least one band in addition to the band from the wild-type AAVS1 region, suggesting that rearrangements of this locus had occurred (Fig. 5). In the seventh clone, hybridization analysis suggested that a concatemer of p2ITRlacZ and pRSVRepCap had integrated at a site other than AAVS1. The *rep*, *lacZ*, and bacterial DNA probes all hybridized to a single band of the same

TABLE 4. DNA hybridization analysis and rescue analysis of the β-galactosidase-positive clones

Clone	Hybridization[a] with probe for:				Rescue[b]
	AAVS1	*lacZ*	pBR322	*rep*	
1	+	+	+	−	+
2	−	+	+	+	+
3	+	+	+	−	+
4	+	+	+	−	+
5	+	+	+	−	−
6	+	+	+	−	+
7	+	+	+	−	+

[a] For the AAVS1 probe, + indicates the presence of at least one band in addition to the band present in 293 cells. For the *lacZ*, pBR322, and *rep* probes, + indicates the presence of at least one hybridizing band.

[b] + indicates the presence of a 4.7-kb sequence in the Hirt supernatant that hybridized with the *lacZ* probe.

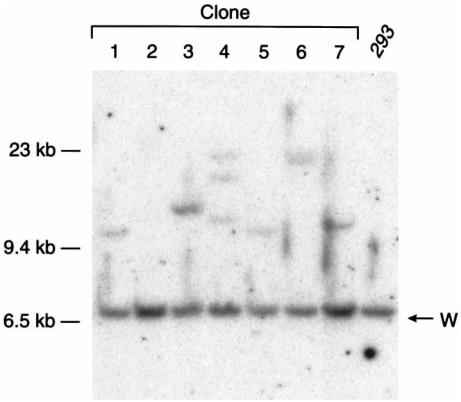

FIG. 5. DNA hybridization analysis of β-galactosidase-positive clones, using an AAVS1 probe. Genomic DNA was prepared from the seven β-galactosidase-positive clones; 10 μg was digested with *Hin*dIII, fractionated, and hybridized with an AAVS1 probe. 293, untransfected 293 cells. The wild-type fragment (W) is more intense than the other hybridizing bands because there are four copies of chromosome 19 in 293 cells.

molecular weight (data not shown). Only the wild-type AAVS1 fragment was observed in this clone (Fig. 5). Five β-galactosidase-negative clones from the p2ITRlacZ-pRSVRepCap transfection were analyzed in a similar manner. One of the clones contained plasmid backbone sequences, but no *lacZ* or *rep* sequences or rearrangements of AAVS1 were detected (data not shown). None of the 12 clones produced a positive signal in the PCR assay.

The ability of the *lacZ* sequences to be rescued was examined in the β-galactosidase-positive clones. The clones were transfected with a *rep*-containing plasmid, pRep, and subsequently infected with adenovirus. Hirt supernatants prepared from the clones were fractionated on a gel and hybridized with a *lacZ* probe. Six of the seven clones produced a 4.7-kb sequence that hybridized with the probe, the size expected for a sequence containing the CMV-*lacZ* gene flanked by ITRs (Table 4). Larger sequences, which presumably represent multimers of the 4.7-kb unit-length fragment, were also detected in the preparations from the six clones.

The location of the integrated p2ITRlacZ plasmid was determined by FISH in five of the β-galactosidase-positive clones with a rearrangement at AAVS1 (Fig. 6). Metaphase chromosome spreads were hybridized with probes derived from plasmid p2ITRlacZ, AAVS1, and the DNA ligase gene (also located at 19q13). In two of the clones, the plasmid was localized to 19q13 at or near the AAVS1 locus. In one of the clones, the plasmid was located on the other arm of chromosome 19; in two clones, the plasmid was found on another unidentified chromosome.

Structure of the integrated *lacZ* sequences. The structure of the integrated DNA was determined in one of the clones in which the *lacZ* gene was localized to AAVS1 by FISH analysis. DNA hybridization suggested that the entire plasmid, including the bacterial replication origin and β-lactamase selectable marker, was present on a single *Hin*dIII fragment. Therefore, genomic DNA was digested with *Hin*dIII, diluted, ligated, and then used to transform *E. coli*. One plasmid recovered was analyzed and found to contain approximately 1.7 copies of p2ITRlacZ (Fig. 7). At one of the junctions, the *lacZ* and AAVS1 sequences were juxtaposed. In AAVS1, the crossover was at a position 2 bp from the RBS. With respect to the plasmid, the junction was more than 1 kb from the nearest RBS in a region bearing no homology to the region of AAVS1 involved in the crossover.

At the other junction, it appeared that DNA from three different regions was present (Fig. 7). Unique genomic DNA was joined to a sequence with homology to the repetitive *Alu* element which was juxtaposed to plasmid sequence. The junction in the plasmid occurred adjacent to an RBS present in the ITR. Joined to the plasmid was 20 bp of sequence with homology to the internal portion of an *Alu* element (8). This sequence was joined to unique DNA which did not originate from the sequenced portion of AAVS1 and had no homology to known sequences. Using inverted PCR, we cloned this DNA and the flanking sequences from the preintegration site in the parental 293 cell line. A sequence with homology to an *Alu* repeat was not found adjacent to this DNA in 293 cells. This finding suggested that the junction resulted from two crossovers, one between this unknown genomic DNA and an *Alu* element and one between the *Alu* sequence and the plasmid. Alignment of the sequences involved in the recombination revealed small regions of homology adjacent to the junctions. Interestingly, one small region of homology was the sequence GAGC, the tetramer that forms the RBS.

DISCUSSION

In this study, we achieved site-specific integration of plasmid DNA into the human genome by using components of AAV. We demonstrated that integration of a plasmid at AAVS1 requires the presence of the AAV *rep* gene which may be provided either in *cis* or *trans*. In the presence of Rep proteins, long-term expression of a *lacZ* gene on an ITR-containing plasmid was increased 5- to 10-fold, to a level which allowed direct isolation of β-galactosidase-positive clones in the absence of selective pressure. The analysis of several clones suggested that the entire plasmid had been integrated in a tandem arrangement. In two of the five clones, the integration of the *lacZ* plasmid was at or near AAVS1 on chromosome 19. The structure of the integrated plasmid and flanking sequence was determined in one of the clones and indicated a complex mechanism of insertion producing genomic rearrangements.

We believe that the increase in long-term β-galactosidase expression in the presence of the *rep* gene reflects an increase in levels of integration stimulated by Rep. The integrated *lacZ* gene could then be passed on to all progeny with fidelity. A recent study has shown that a recombinant AAV vector, transduced into monkey bronchial epithelial cells, can be main-

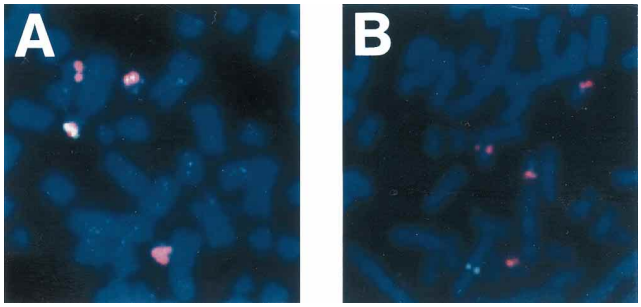

FIG. 6. FISH analysis of β-galactosidase-positive clones. Metaphase chromosome spreads prepared from two clones were analyzed by FISH. The green fluorescence indicates the p2ITRlacZ probe, and the red fluorescence indicates the AAVS1 probe. These clones as well as the untransfected 293 cell line contain four copies of chromosome 19. In panel A, the p2ITRlacZ sequence colocalizes with the AAVS1 locus on one homolog; in panel B, the p2ITRlacZ sequence is present on a different chromosome.

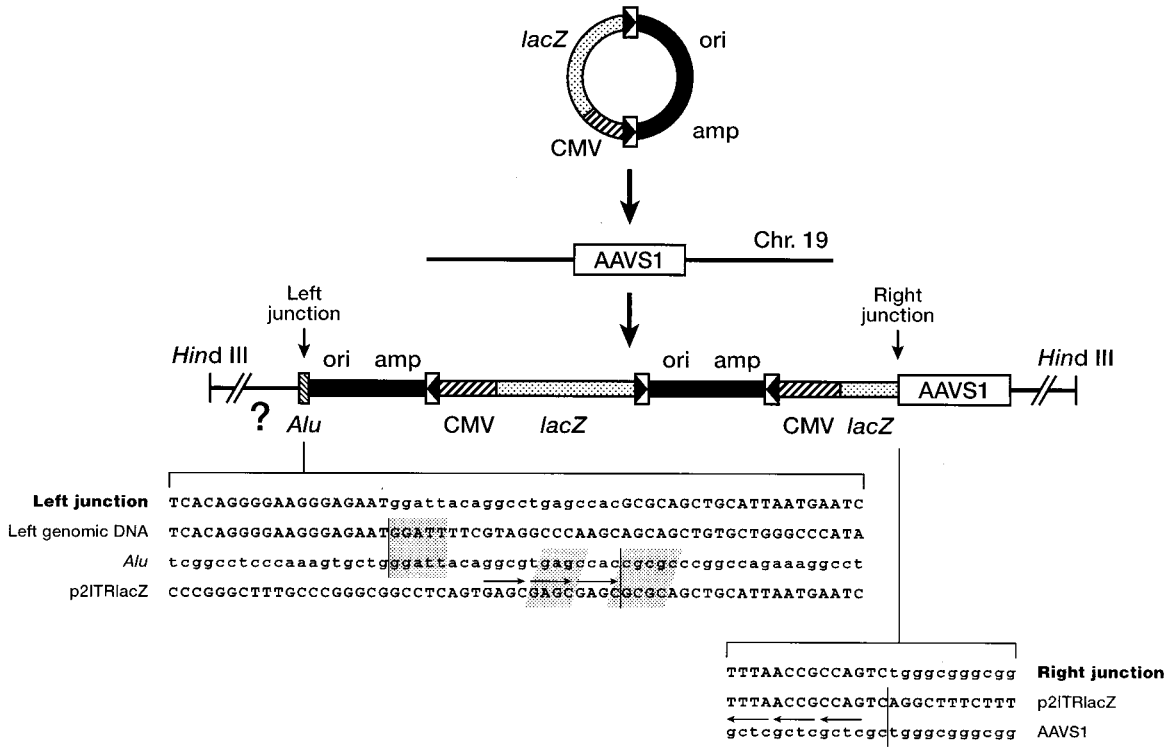

FIG. 7. Structure of the region containing plasmid p2ITRlacZ integrated at AAVS1. At the top is shown the structure of plasmid p2ITRlacZ and the AAVS1 region. Below is shown the structure of the integrated DNA in the AAVS1 region. The various portions of the plasmid are indicated. The open box is the sequenced portion of AAVS1. The boxed arrowheads represent the AAV ITRs. At the bottom the sequence of the junctions are aligned with the sequence of the parental molecules. The vertical lines indicate the apparent crossover points (in some cases the exact point is ambiguous). The GAGC repeats present at the RBS of the ITR or AAVS1 are overlined by arrows. The short regions of homology present adjacent to the crossover point in the parental molecules are indicated by the shaded boxes. Chr., chromosome.

tained for at least 3 months in episomal form (2). This appears not to be the case in the clones that we examined, as FISH analysis demonstrated that the plasmid DNA sequences were associated with the metaphase chromosomal DNA.

Although the sample size is small, the frequency of site-specific integration that we observe in this system is somewhat lower than that determined for wild-type AAV and other AAV constructs. In a study of bronchial epithelial cells, in 94% of the cells containing an integrated wild-type AAV, the virus was mapped to chromosome 19 by FISH (22). In another survey by Kotin et al. (28), approximately 70% of the latently infected cell lines contained AAV at AAVS1. In a recent study of AAV proviral plasmids, 75% were integrated at AAVS1 (49). However, in the latter two studies, rearrangements at AAVS1 were used as a measure of site-specific integration, and our results demonstrate that this may produce an overestimate (see below). The virus used in the first two studies may more efficiently integrate at AAVS1 than the ITR-containing plasmid because of either the presence of free ends or the presence of ITRs at the ends. It is also possible that some internal AAV sequences, which are entirely absent in plasmid p2ITRlacZ, help to promote integration at AAVS1. This possibility is supported by the observation that the AAV p5 promoter sequence (which contains an RBS) is frequently observed at the recombinant junctions and may play some role in the integration reaction (15).

The fact that the two clones that contain an integration event at AAVS1 did not produce a positive signal in the PCR assay was not unexpected and points out a shortcoming of the assay. The limitation of the size of a sequence that can be amplified places many constraints on the structure of the integrated DNA that will produce a positive signal. The site of the integration in AAVS1, as well as the distance and position of the AAV ITR with respect to the junction, will determine if a PCR product is made. Frequently the viral junction is not at an ITR but at some internal sequence (15). Due to these constraints, many cells that actually contain a site-specific integration event will appear negative in the assay. The use of long-distance PCR techniques or additional primers might partially overcome this problem.

Although the Rep proteins are required for site-specific integration and increased long-term expression, it is not clear from our results whether only site-specific integration is increased or whether Rep also contributes to random integration as well. Based on the ability of Rep to bind highly degenerate RBS, we suspect the Rep may also promote random integration or integration at sites with only slight resemblance to the wild-type RBS (12). All activities required for site-specific integration appear to reside in the large Rep proteins. Using constructs that express each one of the four Rep proteins individually, we readily detected integration at AAVS1 in the presence of Rep78 or Rep68, but no site-specific integration was detected with the smaller Rep proteins. Similar results were obtained when other Rep activities, such as in vitro replication (39), DNA binding (11, 42, 55), and endonuclease activity (21), were examined and found to reside in the large Rep proteins.

One unexpected result was the detection of site-specific integration in the PCR assay using a substrate without any portion of an ITR sequence. We believe that two factors contrib-

uted to this result: the ability of the Rep proteins to bind DNA sequences with little resemblance to an RBS and the high degree of sensitivity of the PCR assay. In vitro binding studies have demonstrated that low-affinity Rep binding will occur with any sequence containing a single GAGC followed by a run of G's (12). Such a pattern can be found in each of the non-ITR-containing sequences that we tested. These fragments could have been bound by Rep, albeit inefficiently, and then integrated at AAVS1. A *lacZ* fragment that contained no apparent Rep binding site did not produce a positive signal in the assay (16). Since the PCR assay could detect integration events in 0.1% (10 positives in 10^4 cells) of the population, even rare events could yield a positive signal in this assay. Other assays used to detect integration at AAVS1, such as DNA hybridization or FISH analysis, would not be expected to detect such rare events.

The β-galactosidase-positive clones displayed an inconsistency between the results of the DNA hybridization and FISH analyses. In six clones examined by DNA hybridization, a rearrangement at AAVS1 was observed, suggesting that the *lacZ* plasmid had integrated at that location. However, when five of these clones were analyzed by FISH, only two showed integration at or near AAVS1. One possible explanation is that integration of the plasmid at AAVS1 may be an unstable intermediate in the integration process. Initially one or both ends of the plasmid may be joined to AAVS1 sequences, but perhaps as a result of strand switching, which is thought to be important in the integration reaction (24, 33), the plasmid may ultimately integrate at another locus. However, the formation of this intermediate may result in rearrangements at AAVS1. Alternatively, the rearrangements may be independent of any integration event and may result solely from the activity of the Rep proteins. We have demonstrated that infection of wild-type AAV or transfection of a plasmid containing a CMV-*rep* fusion gene can produce rearrangements at AAVS1 in the absence of integration (16).

In most of the β-galactosidase-positive clones, the entire plasmid had integrated and did so as tandem repeats. This was verified in the one clone that was analyzed in detail and suggested by DNA hybridization analysis of the other clones. Thus, the AAV provirus portion of the plasmid was not excised prior to integration, although the plasmid represents a substrate from which AAV can be rescued (32, 44). This is in contrast to a previous study of AAV-containing plasmids in which no plasmid backbone was detected in five of the six clones with integrated AAV sequence (49). The reason for this difference is unknown. Two groups observed 2 to 20 tandem repeats of the AAV genome after transduction of an AAV vector or AAV, and like the integrants that we examined, these proviruses could be rescued upon addition of the appropriate helper viruses (10, 36). Head-to-tail arrangements of integrated AAV were also observed in the episomal Epstein-Barr virus model system (15). The mechanism that generates these tandem repeats is unknown and may result from replication or recombination of the integrating DNA prior to or during integration. A rolling circle mechanism has been proposed for the viral integrants, and for our plasmid integrants it seems especially plausible since the plasmid template is already circular (15, 31, 36). It is clear that the mechanism that generates these structures is independent of normal AAV replication which generates head-to-head or tail-to-tail repeats (5, 18, 53).

The plasmid-genome junctions in the clone that we analyzed show many structural similarities to those generated by wild-type AAV. With respect to the viral junctions, it was frequently observed that one junction was in or adjacent to the ITR while the second was at some AAV internal sequence (15, 25, 31). With respect to the chromosome, one junction was frequently at or near the RBS in AAVS1. Interestingly, in studies using the Epstein-Barr virus shuttle system, when one chromosomal junction was in AAVS1, the second was never in AAVS1 but in some vector sequence or a sequence unrelated to the vector (15). Finally, a single junction which resulted from two recombination events has also been associated with AAV integration (26). One recombination event between two sequences within AAVS1 resulted in a rearrangement of this region, and a second event generated the junction between AAV and AAVS1. The structures of all of these junctions suggests a mechanism in which the formation of one viral and one chromosomal junction is tightly constrained while the second occurs randomly.

All of the experiments described in this report were carried out in 293 cells, which constitutively express the adenovirus E1a protein. The E1a protein activates *rep* gene expression (9), so that the results obtained in this cell line may not be typical of other human cells. Although our studies indicate that the efficiency of the integration reaction is not influenced by levels of Rep protein, we are presently examining this reaction in K562 and CEM-A cells. Preliminary experiments indicate that Rep protein also increases the efficiency of integration of ITR-containing sequences in these cell lines, and clones that appear to contain integration events at AAVS1 are now being analyzed (37).

This study demonstrates the feasibility of using components of AAV to mediate the nonviral, site-specific integration of DNA into the human genome. Although several interesting aspects of the mechanism were revealed, many questions about this system remain. The utility of this system as a research tool and as a new approach to gene therapy will depend on our ability to increase the efficiency of the integration reaction and minimize unwanted genomic rearrangements.

ACKNOWLEDGMENTS

We thank K. I. Berns for the AAVS1 clone and W.-L. Kuo for assistance with the FISH analysis. We also thank P. Colosi for helpful discussions and J. Boeke, L. Cuoto, and G. Podsakoff for critical reading of the manuscript.

This work was supported in part by NCDDG grant U01-A135382-01.

REFERENCES

1. **Adair, G. A., R. S. Nairn, J. H. Wilson, M. M. Seidman, K. A. Brotherman, C. MacKinnon, and J. B. Scheerer.** 1989. Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells. Proc. Natl. Acad. Sci. USA **86:**4574–4578.
2. **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.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1995. Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
4. **Beaton, A., P. Palumbo, and K. I. Berns.** 1989. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in *trans* by the Rep protein. J. Virol. **63:**4450–4454.
5. **Berns, K. I.** 1990. Parvovirus replication. Microbiol. Rev. **54:**316–329.
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. **Boshart, M., F. Weber, G. John, K. Dorsch-Hasler, B. Fleckenstein, and W. Shafner.** 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell **41:**521–530.
8. **Britten, R. J., W. F. Baron, D. B. Stout, and E. H. Davidson.** 1988. Sources and evolution of human *Alu* repeated sequences. Proc. Natl. Acad. Sci. USA **85:**4770–4774.
9. **Chang, L.-S., Y. Shi, and T. Shenk.** 1989. Adeno-associated virus P5 promoter contains an adenovirus E1A inducible element and a binding site for

the major late transcription factor. J. Virol. **63:**3479–3488.
10. **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.
11. **Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urelay, B. Safer, and R. M. Kotin.** 1994. Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in *Escherichia coli*. J. Virol. **68:**797–804.
12. **Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin.** 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J. Virol. **69:**7334–7338.
13. **Doetschman, T., N. Maeda, and O. Smithies.** 1988. Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA **85:** 8583–8587.
14. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
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. **Godwin, S., and R. T. Surosky.** Unpublished data.
17. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
18. **Hauswirth, W. W., and K. I. Berns.** 1979. Adeno-associated virus DNA replication: nonunit-length molecules. Virology **93:**57–68.
19. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26:**365–369.
20. **Hölscher, C., M. Hörer, J. A. Kleinschmidt, H. Zentgraf, and R. Heilbronn.** 1994. Cell lines inducibly expressing the adeno-associated virus (AAV) *rep* gene: requirements for productive replication of *rep*-negative AAV mutants. J. Virol. **68:**7169–7177.
21. **Im, D. S., and N. Muzyczka.** 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell **61:**447–457.
22. **Kearns, W. G., S. A. Afione, S. B. Fulmer, M. G. 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.
23. **Kessler, P. D., G. M. Podsakoff, X. Chen, S. A. McQuiston, P. C. Colosi, L. A. Matelis, G. J. Kurtzman, and B. J. Byrne.** 1996. Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. Proc. Natl. Acad. Sci. USA **93:**14082–14087.
24. **Kotin, R. M.** 1994. Prospects for the use of adeno-associated virus as a vector for human gene therapy. Hum. Gene Ther. **5:**793–801.
25. **Kotin, R. M., and K. I. Berns.** 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. Virology **170:**460–467.
26. **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.
27. **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.
28. **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.
29. **Labow, M. A., L. Graf, Jr., and K. I. Berns.** 1987. Adeno-associated virus gene expression inhibits cellular transformation by heterologous genes. Mol. Cell. Biol. **7:**1320–1325.
30. **Labow, M. A., P. L. Hermonat, and K. I. Berns.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J. Virol. **60:** 251–258.
31. **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.
32. **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.
33. **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.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka.** 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J. Virol. **68:**4988–4997.
36. **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.
37. **McQuiston, S. A., and R. T. Surosky.** Unpublished data.
38. **Muzyczka, N.** 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. Curr. Top. Microbiol. Immunol. **158:**97–129.
39. **Ni, T. H., X. Zhou, D. M. McCarty, I. Zolotukhin, and N. Muzyczka.** 1994. In vitro replication of adeno-associated virus DNA. J. Virol. **68:**1128–1138.
40. **Ochman, H., M. M. Medhora, D. Garza, and D. L. Hartl.** 1990. Amplification of flanking sequences by inverse PCR, p. 219–227. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. Academic Press, Inc., San Diego, Calif.
41. **O'Gorman, S., D. T. Fox, and G. M. Wahl.** 1991. Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science **251:**1351–1355.
42. **Owens, R. A., J. P. Trempe, N. Chejanovsky, and B. J. Carter.** 1991. Adeno-associated virus Rep proteins produced in insect and mammalian expression systems: wild-type and dominant-negative mutant proteins bind to the viral replication origin. Virology **184:**14–22.
43. **Owens, R. A., M. D. Weitzman, S. R. Kyostio, and B. J. Carter.** 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J. Virol. **67:**997–1005.
44. **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.
45. **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.
46. **Samulski, R. J., L. S. Chang, and T. Shenk.** 1987. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. J. Virol. **61:**3096–3101.
47. **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.
48. **Sauer, B., and N. Henderson.** 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc. Natl. Acad. Sci. USA **85:**5166–5170.
49. **Shelling, A. N., and M. G. Smith.** 1994. Targeted integration of transfected and infected adeno-associated virus vectors containing the neomycin resistance gene. Gene Ther. **1:**165–169.
50. **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.
51. **Srivastava, A., E. W. Lusby, and K. I. Berns.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. **45:**555–564.
52. **Stokke, T., C. Collins, W. L. Kuo, D. Kowbel, F. Shadravan, M. Tanner, A. Kallioniemi, O. P. Kallioniemi, D. Pinkel, L. Deaven, and J. W. Gray.** 1995. A physical map of chromosome 20 established using fluorescence in situ hybridization and digital image analysis. Genomics **26:**134–137.
53. **Straus, S., E., E. D. Sebring, and J. A. Rose.** 1976. Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis. Proc. Natl. Acad. Sci. USA **73:**742–746.
54. **Tratschin, J.-D., I. L. Miller, and B. J. Carter.** 1984. Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function. J. Virol. **51:**611–619.
55. **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.
56. **Wigler, M., A. Pellicer, S. Silverstein, and R. Axel.** 1978. Biochemical transfer of single copy eukaryotic genes using total cellular DNA as donor. Cell **14:**725–731.
57. **Wonderling, R., and R. Owens.** 1996. The Rep68 protein of adeno-associated virus type 2 stimulates expression of the platelet-derived growth factor B c-*cis* proto-oncogene. J. Virol. **70:**4783–4786.