JOURNAL OF VIROLOGY, Mar. 2003, p. 3495–3504 Vol. 77, No. 6
0022-538X/03/$08.00+0 DOI: 10.1128/JVI.77.6.3495–3504.2003


# Genetic Fate of Recombinant Adeno-Associated Virus Vector Genomes in Muscle

Bruce C. Schnepp, K. Reed Clark, Dori L. Klemanski, Christina A. Pacak, and Philip R. Johnson*

*Columbus Children's Research Institute, Columbus Children's Hospital, and Department of Pediatrics, College of Medicine and Public Health, The Ohio State University, Columbus, Ohio 43205*

Received 4 October 2002/Accepted 18 December 2002

**Recombinant adeno-associated virus (rAAV) vectors are promising human gene transfer vectors, because they mediate long-term gene expression in vivo. The vector DNA form responsible for sustained gene expression has not been clearly defined, but it has been presumed that the vector integrates to some degree and persists in this manner. Using two independent methods, we were unable to identify rAAV integrants in mouse muscle. In the first approach, we were unable to recover host cell-vector DNA junctions from a lambda phage library generated using transduced mouse muscle DNA that contained a high vector copy number. Following this result, we devised a PCR assay based on the principle that integrated rAAV vector sequences could be amplified using primers specific for mouse interspersed repetitive sequences (B1 elements). Using this assay, we analyzed transduced mouse muscle DNA isolated from 6 to 57 weeks after injection and did not detect amplification above background levels. Based on the demonstrated sensitivity of the assay, these results suggested that >99.5% of vector DNA was not integrated. Additional analyses using a novel DNA exonuclease showed that the majority of the rAAV vector DNA in muscle persisted over time as transcriptionally active monomeric and concatameric episomes.**

Recombinant adeno-associated virus (rAAV) vectors have been shown to mediate long-term gene transfer in vivo, particularly in organs composed primarily of nondividing cells (e.g., muscle, eye, liver, lung, and brain) (7, 16, 17, 28, 46, 59). Promising preclinical data from animal models have resulted in the initiation of human clinical trials for the treatment of cystic fibrosis and hemophilia B (23, 26, 52, 53). These advances notwithstanding, the basis for long-term gene expression remains undefined, and this uncertainty could hamper further development of these vectors in situations in which vector integration is a safety concern. Thus, as rAAV vectors are proposed for an increasing number of clinical applications, it becomes increasingly important to understand the relationship between vector and host DNA.

Wild-type (wt) AAV integrates primarily (~60%) in a site-specific fashion into human chromosome 19 in cultured cells (27, 29, 45). This site-specific integration appears to be mediated by virally encoded *rep* gene products through the recognition and binding of specific viral and cellular consensus sequences (18, 56). Importantly, almost nothing is known about wt AAV integration or persistence in vivo. Approximately 80% of adults have antibodies reactive with AAV-2 (5), but it is not known whether latent provirus persists (or in what form) within these individuals (39). To date, only a single study has examined wt AAV infection in primates (22). In this study, nine rhesus macaques were inoculated with wt AAV in the presence or absence of wt adenovirus. Using PCR, AAV DNA was found in several animals, while Southern blot analysis of DNA taken from sites of inoculation showed no positive signals. Site-specific integration (measured by PCR amplification of wt AAV-cellular DNA junctions) was detected in one animal (a very faint signal was seen in two others). Thus, it can be concluded that even when conditions are permissive for in vivo site-specific integration, this unusual biologic phenomenon is not a common event (22).

It is important that most rAAV-2 vectors lack the AAV rep gene and thus are not considered to have the potential to undergo targeted integration. Nonetheless, rAAV vectors have been shown to integrate in vitro in many cell types. However, the vast majority of these studies were performed with vectors encoding drug resistance markers that demonstrated integration of rAAV vector in drug-resistant clones (21, 30, 36, 38, 43, 50, 55). In studies in which rAAV vector persistence has been analyzed in the absence of selection in vitro, it appears that rAAV vectors integrated at a very low frequency. In fact, transformed cells in culture transduced with rAAV vectors (in the absence of selection) have shown a loss of transgene expression over time (2, 34). The loss of transgene expression corresponded to a loss of rAAV vector DNA, suggesting that rAAV vectors persisted predominately as episomes that declined in a dividing cell population. These observations stand in stark contrast with the stable gene expression found following retroviral vector transduction (2, 34).

A somewhat analogous situation appears to exist for plasmid DNA transfected into transformed cells in culture. There is a loss of plasmid DNA copies in cells over time due to the episomal nature of this DNA (14). However, under selective conditions, plasmid DNA has been shown to readily integrate in vitro, predominately in head-to-tail arrays (58). Conversely, several in vivo studies have demonstrated that plasmid DNA

* Corresponding author. Mailing address: Columbus Children's Research Institute, Children's Hospital, Room W590, 700 Children's Dr., Columbus, OH 43205. Phone: (614) 722-2735. Fax: (614) 722-3273. E-mail: johnsonp@pediatrics.ohio-state.edu.

persists only at very low levels following intramuscular injection in mouse tissue (31, 35).

Although the basis for rAAV genome persistence remains unclear, three possible events seem plausible: (i) nonhomologous recombination-integration into cellular DNA; (ii) maintenance of extrachromosomal forms; or (iii) a combination of integrated and episomal forms. Several studies have suggested, on the basis of hybridization patterns observed following Southern blot analysis of total cellular DNA from skeletal muscle, liver, and brain, that rAAV vector DNA might be integrated in vivo (7, 8, 15, 33, 59). In these studies, rAAV vector DNA was found associated with high-molecular-weight DNA, although no formal proof of integration was offered (i.e., isolation of vector-cellular junctions or proviruses). Shortly thereafter, based on fluorescent in situ hybridization data, rAAV vector sequences were reported to integrate in the liver (37); however, the interphase nature of the published chromosomal spreads made interpretation of data difficult. Also, rAAV integration was observed in a genetic selection model that identified rare insertional events in which a small population of hepatocytes transduced with a therapeutic rAAV vector were able to repopulate a defective liver (4). In another study, vector-cellular DNA junctions from mouse liver DNA were isolated by PCR (40). However, the frequency of these integrative events within the liver appeared to be very low, as evidenced by the complete loss of detectable rAAV vector in regenerating liver tissue following portal vein infusion (42).

In contrast, episomal forms of vector DNA have been easier to document and characterize in vivo. We previously observed a low-molecular-weight rAAV species that was characterized as a double-stranded episome (8). Additionally, other investigators using Southern blot hybridization have observed monomeric and dimeric rAAV vector forms in different tissues (1, 41, 42, 47, 48, 51). More recently, monomeric and multimeric circularized rAAV vector forms have been directly isolated from mouse skeletal muscle (13).

Considered together, these data do not allow us to clearly define the genetic fate of rAAV vector genomes in transduced tissues from animals. On the one hand, there appears to be evidence that rAAV can integrate, but it is likely a very rare event that remains unquantified. On the other hand, episomal forms of vector DNA appear to be commonly found in transduced tissues, although no attempt has been made to assign a relative frequency to episome formation.

Because of this uncertainty, we sought to quantitatively define the proportion of total rAAV vector genomes that persist in one form or the other in skeletal muscle tissue. To accomplish this goal, we used three distinct and complementary analytical approaches to examine the genetic fate of rAAV vector genomes in transduced muscle from mice. These analyses allowed us to conclude that rAAV vector genomes in muscle persist primarily (>99.5%) as large and small episomal concatemers. We were unable to demonstrate integration events in muscle even by using a sensitive PCR-based assay.

## MATERIALS AND METHODS

**Manipulation of nucleic acids.** Total DNA from cultured cells and mouse skeletal muscle was isolated using standard DNA isolation methods (44). Briefly, cells and tissues were lysed in 10 mM Tris (pH 8.0)–100 mM NaCl–25 mM EDTA–0.5% sodium dodecyl sulfate (SDS)–proteinase K (200 μg/ml) for 12 h at 56°C, followed by sequential phenol-chloroform extractions and ethanol precipitation. Genomic Southern blot analyses were performed using 3 μg of DNA fractionated on 0.8% agarose gels, which were subsequently dehydrated and used for in-gel hybridization (54). DNA hybridization conditions were 65°C for 16 h in buffer containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 1× Denhardt's reagent, and 200 μg of sonicated herring sperm DNA/ml. Rehydrated gels were washed twice at 60°C in 2× SSC–0.2% SDS for 30 min and then twice at 60°C in 0.2× SSC–0.2% SDS for 30 min. For Plasmid-Safe DNase treatments, DNA was incubated for 16 h at 37°C in 33 mM Tris (pH 7.8)–66 mM potassium acetate–10 mM magnesium acetate–0.5 mM dithiothreitol–1 mM ATP with 10 U of Plasmid-Safe ATP-Dependent DNase (Epicentre Technologies)/μg. The Columbus Children's Research Institute Sequencing Core Laboratory performed all DNA sequencing with an ABI 727 capillary electrophoresis sequencer.

**Cells and viruses.** HeLa, 293, and C2C12 cells were purchased from the American Type Culture Collection (Rockville, Md.). HeLa, 293, and C2C12 cells were maintained in Dulbecco's minimal essential medium supplemented with 10% fetal bovine serum and penicillin-streptomycin. rAAV producer cell lines were maintained in the above-described medium supplemented with 700 μg of G418/ml. Production of rAAV vectors encoding the mouse β-glucuronidase gene or the chicken ovalbumin gene were generated from producer cell line GUS-g2 or ova-6, respectively (9). Both vectors contain the human cytomegalovirus (HCMV) promoter/enhancer for high-level transgene expression. rAAV was generated by growing $10^{10}$ producer cells for 5 days in a Corning Cell Cube and then infecting the cells with adenovirus type 5 at a multiplicity of infection (MOI) of 50. After 72 h, the cells were concentrated by low-speed centrifugation (1,000 × *g*) and resuspended at a density of $10^7$ cells/ml in 20 mM Tris (pH 8.0)–1 mM $MgCl_2$–150 mM NaCl. Cells were lysed by treatment with 0.5% deoxycholate and Benzonase (35 units/ml). Vector purification was achieved using heparin affinity column chromatography as described previously (6). Using the PE Applied Biosystems Prism 7700 TaqMan sequence detector system, DNase-resistant particle (DRP) titers on both vectors were determined as previously described (6). The HCMV promoter-specific primer-probe sequences used were as follows: forward primer, 5′ TGAAATCCCCGTGAGTCAA 3′ (300 nM final concentration); reverse primer, 5′ CATGGTGATGCGGTTTTGG 3′ (200 nM final concentration); and HCMV probe, 5′ FAM-CCGCTATCCACGCCCATTGATG-TAMRA 3′ (200 nM final concentration). The presence of wt-like AAV was assessed by a double-blind passage assay using 1% of the rAAV stock as previously described (6). Based on assay sensitivity (1 IU), each vector preparation contained <1 IU of wt-like AAV per $10^{11}$ rAAV particles (DRP). The presence of replication-competent adenovirus in the rAAV preparations was assayed by passing 1% of the rAAV vector stocks onto 293 cells and scoring for adenovirus cytopathic effect after 7 days of incubation (6). The vector stocks did not possess detectable replication-competent adenovirus (<1 PFU per $10^{11}$ rAAV particles). Vector stocks contained <1 endotoxin unit per 50-μl dose (Limulus Ameobecyte Assay; Cape Cod Inc.). Adenovirus type 5 was purified from infected 293 cells by CsCl step and isopycnic gradient centrifugations according to standard protocols (19).

rAAV/eGFP-zeo stable cell lines were generated by transducing C2C12 cells with an rAAV vector encoding an enhanced green fluorescent protein (eGFP)-zeocin fusion gene at an MOI of $10^4$ DRP per cell. After 48 h, the cells were plated into 100-mm-diameter dishes at a density of $1 \times 10^5$ cells/dish and selected in the above-described medium supplemented with 800 μg of zeocin/ml. The stable cell line pool was passed 10 times (~40 cell doublings) in selective medium to ensure that extrachromosomal rAAV vector forms were lost via dilution.

**Animals.** Adult male BALB/c mice were used for all studies described in this paper, and experiments were carried out in accordance with local institutional and National Institutes of Health guidelines. The animals were anesthetized with Avertin (240 mg/kg of body weight, given intraperitoneally by injection intramuscularly into the hind limb). A skin incision (0.5 cm in length) was made over the distal femur, and using a 28-gauge needle, $2.5 \times 10^{11}$ DRP of rAAV/gus or $1.0 \times 10^{11}$ DRP of rAAV/ova (50-μl volume) was injected along the longitudinal axis of the quadriceps muscle. Control animals received PBS in an identical manner. The entire quadriceps muscle group was removed and rapidly frozen in liquid nitrogen prior to DNA extraction.

**Phage library construction and screening.** Total cellular DNA was isolated from mouse skeletal muscle at 12 weeks postinoculation with rAAV/ova ($10^{11}$ DRP). TaqMan PCR revealed that this muscle DNA sample possessed 36 vector copies/nucleus. Using a QIAEX II extraction kit (Qiagen, Inc.) according to the manufacturer's specifications, 4 μg of DNA was digested with *Eco*RI and *Xho*I (*Xho*I cuts once within the rAAV vector sequence) and separated on a 0.8% agarose gel. DNA ranging in size from 3 to 12 kb was extracted from the gel. The recovered fragments were ligated, using a Uni-ZAP XR vector system (Strat-

agene), into lambda phage vector arms (predigested with *Xho*I and *Eco*RI) and packaged (using a Uni-ZAP XR Gigapack cloning kit according to the manufacturer's instructions) into recombinant phage. An analysis of 50 rescued phagemids revealed that the phage library was 98% recombinant, with an average insert size of 4 kb. Using standard methods (44), a total of $1.5 \times 10^6$ phages were screened with an HCMV-specific radiolabeled probe.

**B1 PCR.** Total cellular DNA (20 to 40 ng) isolated from mouse skeletal muscle or C2C12 cells was used in all B1 PCRs; this amount of DNA gave the greatest level and size distribution of B1 amplicon amplification. Higher sample DNA amounts appeared to rapidly exhaust the primer and nucleotide pools, resulting in PCR products with a smaller size distribution. The DNA was first digested with *Pac*I (to reduce sample viscosity) and then amplified in 20 mM Tris-HCl (pH 8.75)–10 mM KCl–10 mM $(NH_4)_2SO_4$–2 mM $MgCl_2$–400 μM each deoxynucleoside triphosphate–5 U of Herculase Hotstart DNA polymerase (Stratagene)–2 μg of B1 primer in a 50-μl reaction mixture. The B1-repeat primer sequence used was 5′ AGTTCCAGGACAGCGAGGGCTAYACAGA 3′. The reaction conditions were as follows: denaturation at 95°C for 2 min, followed by 10 cycles of 95°C for 30 s, 64°C for 30 s, and 72°C for 7 min 30 s, with an additional 20 cycles of 95°C for 30 s, 64°C for 30 s, and 72°C for 10 min 30 s. A portion (2.5 μl) of the B1 PCR mixture was then used as the template for real-time PCR (TaqMan) to quantitate vector sequence-specific amplification in the B1 PCR. The TaqMan PCR conditions used were as follows: 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min in a 25-μl reaction volume. For sample quantification, a plasmid DNA standard curve was generated in triplicate using of $10^1$ to $10^6$ plasmid copies; experimental values were obtained by extrapolation (using system software) from the plasmid DNA standard curve (coefficient of linearity, ≥0.995).

To determine the initial rAAV vector copy number in the muscle samples, 25 ng of total muscle DNA was used in a standard TaqMan PCR with the HCMV primer-probe combination. The resulting vector copy number was divided by 4,167 nuclei (assuming 6 pg per nuclei) to obtain the number of vector copies per nucleus. The ovalbumin-specific TaqMan PCR primers and probe were as follows: forward, 5′ CTGACCATCCATTCCTCTTCTGT 3′ (300 nM final concentration); reverse, 5′ GAAACACATCTGCCAAAGAAGAGA 3′ (300 nM final concentration); and probe, 5′ FAM-TCAAGCACATCGCAACCAACGCC-TAMRA-3′ (200 nM final concentration). To analyze the integrity of genomic DNA, the following mouse *gus*-specific primers were used: forward, 5′ ACCCT GCGGTTGTGATGTG 3′ (50 nM final concentration); reverse, 5′ AATATG CGG CGGGTTTCAG 3′ (300 nM final concentration); and probe, 5′ FAM-T GTGGCCAATGAGCCTTCCTCTGC-TAMRA 3′ (200 nM final concentration).

**Cloning and expression analysis of rAAV in vivo vector forms.** One microgram of total muscle DNA, isolated from an rAAV/ova-injected mouse, was digested with *Eco*RI (an enzyme which does not cut within the rAAV vector) and treated with Plasmid-Safe DNase as described above. This DNA was used as a template in a PCR mixture containing 20 mM Tris (pH 8.4), 50 mM potassium chloride, 1.5 mM magnesium chloride, 1 mM deoxynucleoside triphosphates, and 2.5 U of Platinum *Taq* DNA polymerase (Gibco BRL) in the following reaction conditions: 94°C for 2 min for 1 cycle and 94°C for 30 s, 55°C for 30 s, and 72°C for 5 min for 30 cycles. The forward primer used was 5′ AACAGCTGCAGATCAA GCCAGAGAGCTCAT 3′, and the reverse primer was 5′ TTGATCTGCAGC TGTTTGAAAGTTGATAGG 3′. The PCR product was cloned into a TOPO-TA cloning vector (Invitrogen). Positive colonies were identified using a radioactive HCMV probe by a standard colony hybridization method (44). rAAV vector fragments were released by *Pst*I digestion and self-ligated using T4 DNA ligase (New England Biolabs). Using Superfect Transfection reagent (Qiagen), the ligation reaction mixture was directly transfected into $3 \times 10^5$ HeLa cells. A cell lysate was generated after 48 h, electrophoresed on a SDS–10% polyacrylamide electrophoresis gel, and subsequently transferred to a polyvinylidene difluoride membrane (Amersham). The membrane was blocked for 10 min at 25°C using TBST (10 mM Tris [pH 8.0], 150 mM NaCl, 0.1% Tween 20) plus 5% nonfat dry milk. The primary antibody was a monoclonal anti-chicken egg albumin (Accurate Chemical) and was hybridized in blocking solution at 1:1,000 for 1 h. The blot was washed three times for 10 min in TBST and incubated for 30 min in TBST plus 2.5% dry milk with a horseradish peroxidase-conjugated anti-mouse secondary antibody (Vector Laboratories) at a 1:10,000 dilution. The blot was washed as above and developed using an ECL Plus detection system (Amersham).

## RESULTS

**Direct cloning of host cell-vector junctions directly from rAAV-transduced muscle.** We previously reported the association of rAAV vector DNA with high-molecular-weight DNA on Southern blots of transduced mouse brain and skeletal muscle tissue (7, 8). From these data and those of others (15, 33, 59) we inferred that integrated forms of rAAV vector DNA might be present in vivo. To further characterize the form of vector DNA, we attempted to isolate vector-cellular junctions that are predicted to form following rAAV integration into the host genomic DNA. Total cellular DNA from transduced mouse muscle (see Materials and Methods) was digested with *Eco*RI (does not cut within the rAAV/ova genome) and *Xho*I (cuts once within the rAAV/ova genome) and then size selected to enrich for restriction fragments containing host cell-AAV junctions. Quantitative PCR revealed that rAAV vector sequences were present in the genomic DNA at 36 vector copies per nucleus. A lambda phage library was generated and screened with an HCMV-specific probe (see Materials and Methods). Upon screening $1.5 \times 10^6$ recombinant phages, we failed to isolate any clones containing viral-cellular junctions, despite achieving twofold genomic coverage ($P = 0.999$) (10).

**PCR-based assay for rAAV integration.** To complement and extend the direct cloning approach described above, we developed a PCR-based assay that was designed to detect integrated rAAV vector DNA in mouse tissue. This approach was predicated on the random amplification of mouse genomic DNA that is flanked by Alu-like mouse interspersed repetitive sequences (known as B1 elements) by using a B1-specific PCR primer. Mouse B1 repetitive elements are approximately 135 bp long and are present at 80,000 to 100,000 copies per haploid genome (25). Thus, B1 elements comprise about 1% of the total mouse genome and exist on average every 15 kb. Analysis of B1 PCR products using mouse genomic DNA as the template and a single B1 primer revealed a smear ranging in size from >17 kb down to 200 bp (data not shown). Thus, the size distribution of our optimized B1 PCR reflected the expected size distribution based on B1 repeat frequency.

In this assay, rAAV vector DNA integrated between two B1 elements is amplified and the amplified products (containing rAAV integrants) are then quantified using TaqMan PCR. The assay readout is expressed as fold amplification of the rAAV vector DNA (which occurs in the B1 PCR) over the carryover input rAAV vector sequences in the subsequent TaqMan PCR. Hence, net sequence amplification (over input) of the rAAV target sequence following the B1 PCR is diagnostic for rAAV integration. Importantly, episomal rAAV vector DNA is not amplified using the B1 PCR strategy and no net amplification of episomal rAAV vector DNA is predicted to occur.

For example, consider the situations illustrated in Fig. 1. In the case in which there are only nonintegrated copies of rAAV vector DNA (Fig. 1, left lane), no amplification occurs (episomes not flanked by B1 sequences) and there is no fold amplification over input. In the case in which there are only integrated forms (Fig. 1, center lane), the input signal (100 integrants) is amplified 1,000-fold in the B1 PCR. In the last case, both episomal and integrated forms are present (number in episomal form > number in integrated form). The same degree of amplification occurs, but the fold amplification is decreased because the episomal forms are detected in the TaqMan PCR quantification. However, this level of amplification still allows integration events to be detected.

To experimentally validate these concepts, we first asked whether exogenous plasmid sequences (episomes) would sig-

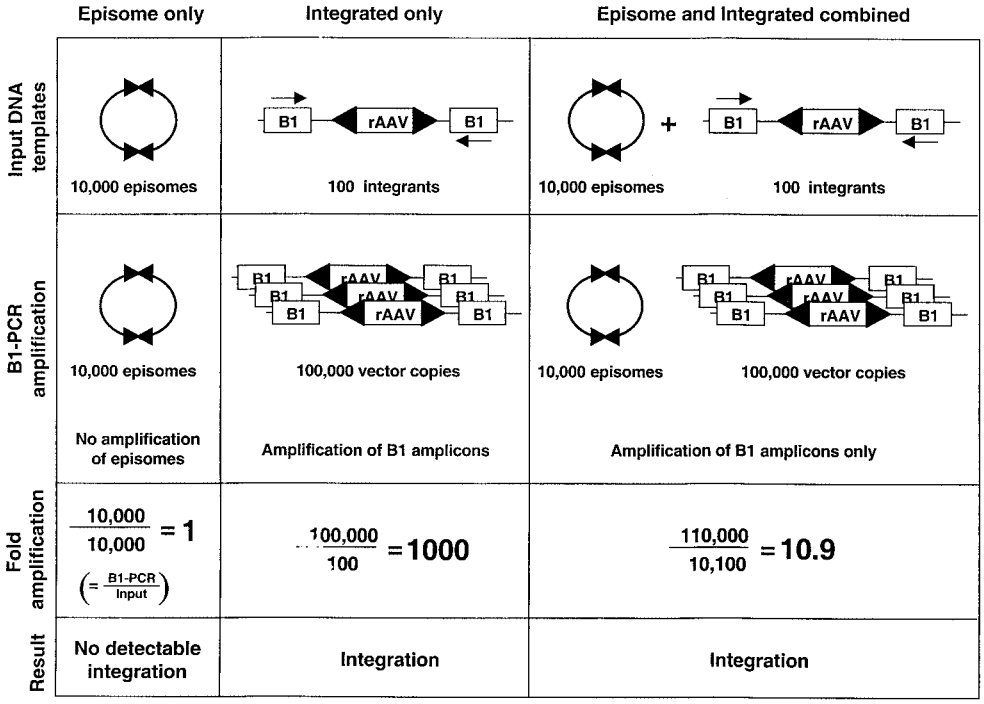


FIG. 1. Schematic of the B1 PCR strategy. Shown in this schematic are three potential forms of rAAV vector genomes in the host cell nucleus. As shown in the left panel, episomal vector forms do not amplify in the B1 PCR because B1 sequence elements are not present. As shown in the middle panel, integrated rAAV vector DNA are expected to amplify in the B1 PCR. Finally, as shown in the right panel, when both forms are present only the integrated proviral vector DNA amplifies in the B1 PCR. As demonstrated here, nonamplified episomal vector DNA carries over into the subsequent TaqMan PCR and is detected as input signal.

nificantly amplify in this assay (Table 1). Differing copy numbers ($1.5 \times 10^0$ to $1.5 \times 10^4$) of circular plasmid DNA containing an HCMV/β-galactosidase expression cassette flanked by AAV inverted terminal repeats (ITR) were added to a B1 PCR mixture containing a fixed amount of mouse genomic DNA, and then 5% of the reaction mixture was subjected to quantitative PCR. The highest background amplification (2.6 ± 1.1) over input carryover was observed at $1.5 \times 10^3$ plasmid copies (Table 1). These data were then used to define a baseline for background amplification. We designated a threshold of 2 standard deviations above this background mean (4.8-fold) as diagnostic for net DNA amplification following B1 PCR and, hence, as evidence for integration.

To further validate the B1 PCR assay in even more complex mixtures, we examined amplification of rAAV vector sequences from cells with known rAAV integrants. To generate cells that contained integrated rAAV vector genome sequences in a relevant DNA substrate, C2C12 cells (a transformed mouse myoblast cell line) were transduced with an rAAV vector encoding a fusion gene product (eGFP/zeocin). Under zeocin selection, a pool of zeocin-resistant cells was generated and analyzed. Southern blot analysis on the C2C12 pool showed that the rAAV vector integrated randomly in a head-to-tail tandem array (data not shown). Quantitative PCR analysis further revealed that the C2C12 pool contained on average 4.5 rAAV integrated vector copies/nucleus. A range of rAAV integrated vector copies ($1.5 \times 10^0$ to $1.5 \times 10^4$) were then assayed by serially diluting the DNA into a constant

TABLE 1. Characterization of B1 PCR amplification

| B1 PCR templates | | B1 PCR results | |
|---|---|---|---|
| No. of plasmid copies[a] | No. of C2C12 integrants[b] | Fold amplification (vector)[c] | No. of repeats[d] |
| $1.5 \times 10^4$ | 0 | 2.4 ± 0.8 | 9 |
| $1.5 \times 10^3$ | 0 | 2.6 ± 1.1 | 9 |
| $1.5 \times 10^2$ | 0 | 1.2 ± 0.5 | 9 |
| $1.5 \times 10^1$ | 0 | 0.7 ± 0.7 | 9 |
| $1.5 \times 10^0$ | 0 | 1.0 ± 1.1 | 9 |
| 0 | $1.5 \times 10^4$ | 112.7 ± 45.7 | 6 |
| 0 | $1.5 \times 10^3$ | 223.6 ± 181.5 | 6 |
| 0 | $1.5 \times 10^2$ | 450.9 ± 388.5 | 6 |
| 0 | $1.5 \times 10^1$ | 4.4 ± 3.5 | 6 |
| 0 | $1.5 \times 10^0$ | 4.4 ± 5.7 | 6 |
| $1.5 \times 10^4$ | $1.5 \times 10^4$ | 106.3 ± 39.7 | 6 |
| $1.5 \times 10^4$ | $1.5 \times 10^3$ | 47.1 ± 24.6 | 6 |
| $1.5 \times 10^4$ | $1.5 \times 10^2$ | 16.3 ± 5.9 | 6 |
| $1.5 \times 10^4$ | $7.5 \times 10^1$ | 6.1 ± 1.5 | 6 |
| $1.5 \times 10^4$ | $1.5 \times 10^1$ | 1.5 ± 0.2 | 6 |
| $1.5 \times 10^4$ | $1.5 \times 10^0$ | 1.4 ± 0.3 | 6 |

[a] Number of copies of a plasmid that contains an rAAV/β-galactosidase genome spiked into 20 ng of nontransduced C2C12 DNA.
[b] Number of copies of an rAAV integrated genome (see Results) spiked into a total of 20 ng of C2C12 DNA.
[c] Fold amplification, number of copies amplified in B1 PCR divided by the number of input template copies (see Fig. 1 for a schematic representation). Vector is the rAAV genome being analyzed. Values represent mean fold amplification ± standard deviation.
[d] Number of experimental replicates.

amount (20 ng) of nontransduced (naïve) C2C12 DNA (Table 1). Significant B1 PCR amplification (450-fold) of rAAV vector DNA above input carryover was observed using as few as 150 rAAV integrated vector copies, which was well above the defined cutoff for background amplification (>4.8-fold). We observed that increasing the C2C12 integrated vector copies from $1.5 \times 10^2$ to $1.5 \times 10^4$ (Table 1) did not result in increased amplification in the B1 PCR assay. This was due to the fact that the level of B1 PCR sequence amplification at higher vector template inputs did not completely compensate for the 100-fold increase in the carryover input (denominator).

Because it is possible that rAAV genomes in vivo are integrated and episomal, we next tested the ability of the B1 PCR assay to identify integrated rAAV vector DNA within a competing episomal vector DNA background. The effect of competing episomal vector DNA on B1 PCR assay sensitivity can be readily illustrated by comparing the fold amplification required to observe net vector amplification in the absence of competing episomal DNA to that required in the presence of competing episomal DNA. For example, to detect 150 integrated vector copies in the absence of competing episomal DNA, we would simply need to amplify the integrated sequences by a factor greater than 4.8-fold (our previously defined cutoff value) to score this sample as positive for vector integration. In fact, at the 150-copy level, we observed average amplification of 451-fold over input vector (Table 1). On the other hand, to detect 150 integrated vector copies in the presence of 15,000 episomal copies, the integrated sequences would need to be amplified by a minimum of 385-fold to see net vector amplification. This is because if 150 integrated vector copies were amplified 385-fold in the B1 PCR assay, after amplification there would be a total of 57,700 integrated vector copies plus 15,000 unamplified episomal copies (72,700 total copies). If 5% of this reaction mixture ($0.05 \times 72{,}700 = 3{,}636$) and 5% of the input carryover ($0.05 \times 15{,}150 = 757.5$) were quantified by TaqMan PCR, the fold amplification would be $3{,}636/757.5 = 4.8$. In practice, at the 150-copy level, we observed an average amplification of 16.3-fold over input vector (Table 1). Thus, the effect of the presence of high levels of episomal vector target is to reduce the net amplification, because the input carryover (denominator) is increased (Fig. 1).

In light of these observations, a series of episomal and integrated rAAV vector spike experiments were performed to define assay sensitivity in the presence of competing episomal vector DNA (Table 1). Each reaction mixture contained $1.5 \times 10^4$ plasmid copies with different percentages of integrated rAAV vector DNA. We tested dilutions ranging from 50% of the integrated target ($1.5 \times 10^4$ plasmid copies + $1.5 \times 10^4$ integrated copies) down to 0.01% of the integrated target ($1.5 \times 10^4$ plasmid copies + $1.5 \times 10^0$ integrated copies). The B1 PCR results indicated that we were able to detect significant amplification of 75 integrated rAAV vector copies within an episomal background of $1.5 \times 10^4$ plasmid copies (Table 1). Thus, the assay was able to detect rAAV vector integration in as few as 16.7 cell equivalents (assuming 4.5 copies per cell) in a background of 3,333 cells (assuming 6 pg of DNA per cell), even with competing input episomal DNA at levels exceeding $10^4$ vector copies.

To verify that all genomic templates were amplifiable in the B1 PCR, we used the native mouse *gus* locus as an internal control to monitor B1 PCR efficiency. Based on the known genomic sequence of the *gus* gene, a portion of the *gus* locus is flanked by two B1 elements spaced 1.5 kb apart. Significant levels of *gus* locus amplification were observed in all genomic DNA samples (>700-fold; data not shown) following B1 PCR assay analysis.

TABLE 2. B1 PCR of transduced mouse muscle

| B1 PCR templates | | | B1 PCR results | |
|---|---|---|---|---|
| Mouse quadriceps[a] | Wk after injection | No. of vector copies/nucleus[b] | Fold amplification (vector)[c] | No. of repeats[d] |
| B4 | 6 | 35 | 1.1 ± 0.1 | 4 |
| C1 | 6 | 54 | 0.9 ± 0.1 | 4 |
| D3 | 6 | 22 | 1.1 ± 0.3 | 4 |
| E2 | 6 | 18 | 1.0 ± 0.2 | 4 |
| C2 | 12 | 13 | 1.0 ± 0.2 | 4 |
| E3 | 15 | 5 | 1.4 ± 0.2 | 4 |
| E5 | 15 | 8 | 1.5 ± 0.3 | 4 |
| G1 | 24 | 6 | 2.1 ± 0.7 | 4 |
| G2 | 24 | 4 | 1.8 ± 0.1 | 4 |
| H4 | 24 | 5 | 2.2 ± 0.8 | 4 |
| J1 | 32 | 2 | 1.1 ± 0.2 | 4 |
| J2 | 32 | 2 | 1.8 ± 0.5 | 4 |
| I4 | 57 | 2 | 1.5 ± 0.5 | 4 |

[a] Transduced muscles were harvested from individual mice at the indicated time after injection. Total cellular DNA was isolated from the entire muscle, and 20 to 40 ng were used in each reaction.
[b] Number of rAAV genomes per nucleus in the given sample as assayed by TaqMan PCR. A total of $1.5 \times 10^4$ input vector copies were assayed by B1 PCR for each mouse.
[c] Fold amplification, number of copies amplified divided by the number of input copies (see Fig. 1). Vector, rAAV genome analyzed.
[d] Number of experimental replicates.

**Analysis of skeletal muscle for integrated rAAV vector sequences using B1 PCR.** Total cellular DNA was isolated at various times (6 to 57 weeks) postinjection from quadriceps tissue of rAAV/ova-injected animals. The average number of vector copies per nucleus for each muscle sample was determined by quantitative PCR (Table 2). Muscle from the earlier time points contained higher vector copies/nucleus than later time points; this was attributable to the diminution of single-stranded input vector genomes over time. As discussed above, the B1 PCR assay was optimized for the use of $1.5 \times 10^4$ vector copies in 20 to 40 ng of genomic DNA. Therefore, mouse DNA samples having more than $1.5 \times 10^4$ vector copies in 20 ng of genomic DNA were diluted in total cellular DNA isolated from naïve mice to achieve this final copy number. Mouse DNA samples having fewer than $1.5 \times 10^4$ vector copies in 20 ng of genomic DNA (mice J1, J2, and I4) were analyzed using 40 ng of DNA so that a total of $1.5 \times 10^4$ vector copies could be analyzed.

The fold amplification of rAAV vector sequences in all rAAV/ova-injected mouse samples was below the cutoff threshold (Table 2). Based on the assay sensitivity, this meant that there were fewer than 75 integrated rAAV vector copies per $1.5 \times 10^4$ total rAAV vector copies, which means that <0.5% of the total vector DNA signal might have been integrated.

**Use of a novel DNA exonuclease to characterize the form of vector DNA.** Considered together, the failed attempts to identify integrated forms of rAAV vector DNA directly implied

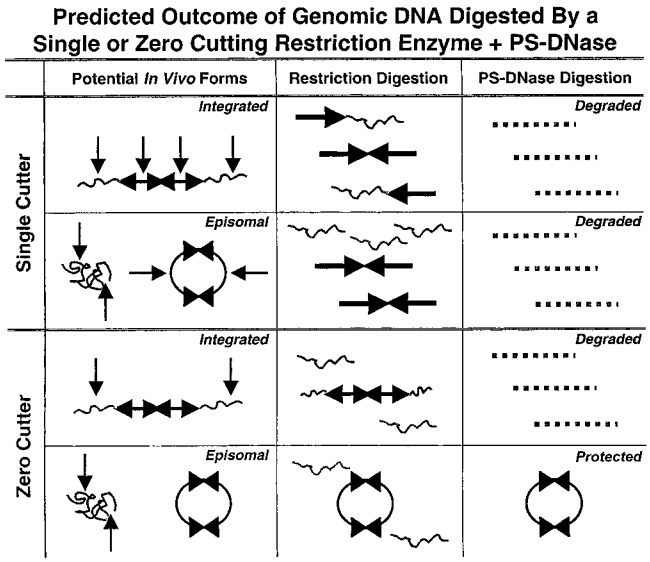


FIG. 2. Schematic of the PS-DNase assay for episomal rAAV genomes. The predicted outcomes of treating rAAV-transduced muscle DNA with selected restriction enzymes in combination with PS-DNase are shown in this diagram. In the case of treatment with an enzyme that cuts once within the genome (Single Cutter), subsequent treatment with PS-DNase leads to degradation of all rAAV genomes, regardless of their forms (episomal or integrated). In the second case, when the total DNA is cut with an enzyme that does not cut within the rAAV genome (Zero Cutter), after treatment with PS-DNase, integrated forms are degraded while episomal forms are protected.

that the high-molecular-weight vector DNA which persisted in muscle over long periods of time existed as episomal concatemers. To investigate this possibility, we developed an assay using a novel DNA exonuclease (Plasmid-Safe, henceforth called PS-DNase) to selectively analyze circular DNA. PS-DNase selectively hydrolyzes linear double-stranded DNA and linear and closed-circular single-stranded DNA but not double-stranded, closed circular supercoiled or nicked DNA. This nuclease is routinely used in the course of plasmid DNA purification to selectively remove contaminating bacterial genomic DNA from circular plasmid DNA preparations.

Given these unique properties, we reasoned that PS-DNase might be useful to discriminate between integrated (susceptible to digestion) and episomal (resistant to digestion) rAAV vector DNA that was associated with high-molecular-weight genomic DNA (Fig. 2). When used in combination with specific restriction enzymes, we predicted that episomal rAAV DNA would be protected from degradation by PS-DNase and that this protection (or degradation) would be evident on Southern blots.

Before proceeding to analyses of DNA from transduced muscle, we wanted to confirm the predicted specificities of PS-DNase. We examined linear, open circular, and supercoiled plasmid DNA by Southern blot hybridization. As expected, linear plasmid DNA was completely hydrolyzed by PS-DNase while open circular and supercoiled plasmid DNAs were protected from PS-DNase digestion (Fig. 3). Because of the extensive secondary structure present in the AAV ITR, it was important to verify that PS-DNase could digest through the ITR. Accordingly, we purified single-stranded rAAV vector DNA and demonstrated that PS-DNase could completely hydrolyze rAAV vector DNA, either in single-stranded form (data not shown) or reannealed in a double-stranded form (Fig. 4, lanes 2 and 3). We also tested PS-DNase activity on replicating monomeric (Rfm) and dimeric (Rfd) replicative forms and showed that they were likewise susceptible to digestion (Fig. 4, lanes 4 and 5).

To extend our characterization of PS-DNase activity, we also asked whether PS-DNase could hydrolyze rAAV vector DNA which was known to be integrated. Total cellular DNA was isolated from the C2C12 pool described earlier, digested with a restriction enzyme (*Eco*RI) that does not cut within the rAAV vector genome, and subsequently treated with PS-DNase to degrade all linear double-stranded DNA. Quantitative PCR analysis revealed that >99% of the rAAV vector signal was eliminated, indicating that PS-DNase can hydrolyze concatameric, integrated rAAV vector DNA (data not shown). Moreover, as an internal control, we performed quantitative PCR using a primer-probe pair specific to the endogenous murine *gus* gene and showed that >99% of the genomic DNA was also degraded (data not shown).

Total cellular muscle DNA was isolated from two animals at 10 or 15 weeks after rAAV injection with either rAAV/gus or rAAV/ova, respectively. The resulting Southern blot hybridization patterns from the two animals were nearly identical before and after treatment with the indicated enzymes; data from the 15-week animal are shown in Fig. 5. Uncut genomic DNA contained several rAAV vector DNA hybridizing species, including high-molecular-weight forms, a monomeric vector form, and residual input single-stranded vector DNA. When genomic DNA was digested with a restriction enzyme that cut once within the rAAV vector, the high-molecular-weight and monomeric bands were reduced to a prominent band which (based upon size) corresponded to a unit length head-to-tail fragment (Fig. 5, lane 2). This was consistent with the rAAV vector DNA being organized in concatameric head-to-tail arrays. When this same linearized genomic DNA was exposed to PS-DNase, complete loss of signal was observed on both the

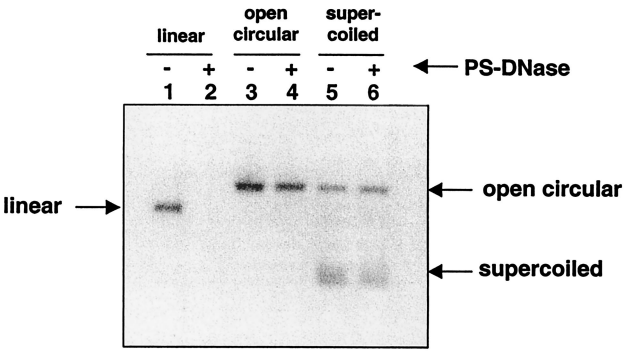


FIG. 3. Specificity of PS-DNase on plasmid DNA substrates. Southern blot analysis was performed on 1 ng of the indicated plasmid DNA templates with (+) or without (−) PS-DNase treatment. Linear plasmid DNA (lanes 1 and 2) was completely hydrolyzed in the presence of PS-DNase (lane 2). As expected, open circular (lanes 3 and 4) and supercoiled DNA (lanes 5 and 6) forms were resistant to PS-DNase treatment (lanes 4 and 6).

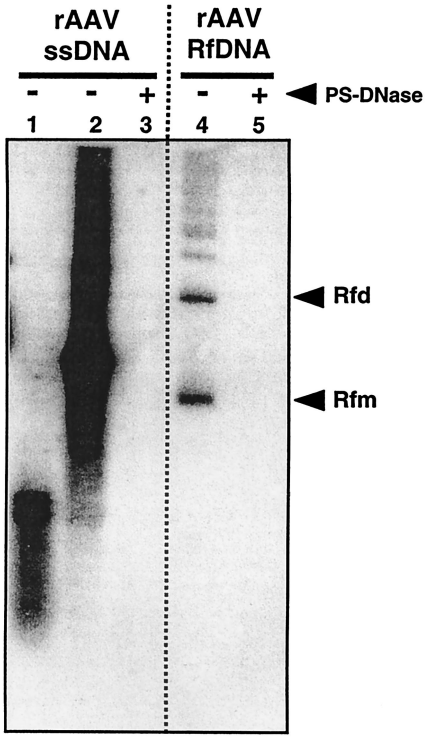


FIG. 4. Specificity of PS-DNase on rAAV replicative forms. Southern blot analyses of rAAV vector forms (1 ng) treated with (+) or without (−) PS-DNase are shown. Recombinant AAV single-stranded genomes were isolated from a highly purified rAAV preparation, and the DNA was denatured by boiling (note the presence of single-stranded genomes of both polarities) (lane 1). A similarly treated DNA sample was allowed to anneal (lane 2) and shows a prominent band corresponding to double-stranded unit length vector, in addition to a smear of complex annealed forms. All reannealed forms were susceptible to PS-DNase treatment (lane 3). The results of an analysis of the susceptibility of cellular rAAV replicative forms (Rf DNA) to PS-DNase digestion are shown in lanes 4 and 5. Monomeric (Rfm), dimeric (Rfd), and higher-order (lane 4) forms were completely hydrolyzed by PS-DNase coincubation (lane 5).

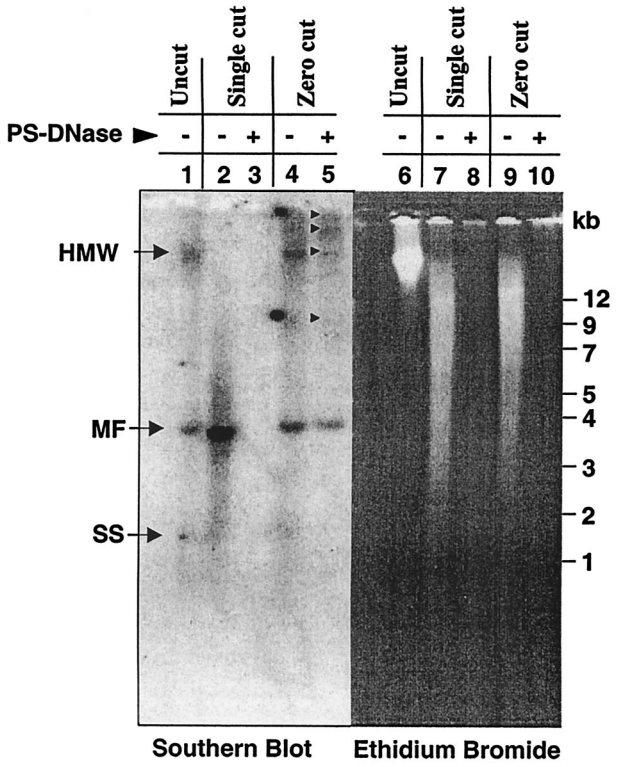


FIG. 5. Southern blot analysis of rAAV vector forms in skeletal muscle following PS-DNase treatment. Total cellular DNA was isolated from skeletal muscle at 15 weeks following injection with rAAV/ova. In the left panel, a Southern blot hybridized with an HCMV vector-specific probe is shown. In the right panel, the identical gel stained with ethidium bromide is shown. Lanes 1 and 6, uncut DNA; lanes 2 and 7, *Pst*I-digested DNA (Single cut [in the rAAV genome]); lanes 3 and 8, *Pst*I-digested DNA after PS-DNase digestion; lanes 4 and 9, *Eco*RI-digested DNA (Zero cut [within the rAAV genome]); lanes 5 and 10, *Eco*RI-digested DNA after PS-DNase digestion. PS-DNase completely degraded linearized rAAV vector DNA (lanes 3 and 8) but did not degrade *Eco*RI-digested (Zero cut) rAAV vector DNA (lane 5, arrowheads), even though the cellular genomic DNA was completely degraded (lane 10).

Southern blot (Fig. 5, lane 3) and the same gel stained with ethidium bromide (Fig. 5, lane 8).

In contrast, when the same genomic DNA samples were digested with a restriction enzyme that does not cut within the rAAV vector, the pattern of the hybridization signal was very similar to that observed with uncut DNA (Fig. 5, lane 4). Importantly, after PS-DNase treatment, all rAAV vector forms (except the single-stranded input form) were still present (Fig. 5, lane 5). PS-DNase activity on linear DNA substrates was confirmed by the total loss of ethidium bromide-stained DNA (Fig. 5, lane 10). Thus, the resistance of the monomeric and high-molecular-weight rAAV DNA vector forms to PS-DNase indicated that these vector forms were double-stranded, circular episomes. Furthermore, the resistant high-molecular-weight vector signal appeared as distinct bands of 12 kb and higher (Fig. 5, lanes 4 and 5), suggesting that these rAAV vector forms represent different-sized episomal concatemers.

**Episomal rAAV vector genomes are transcriptionally active.** Since the PS-DNase and B1 PCR experiments both indicated that the majority of rAAV vector DNA persists in an episomal form, we were interested in whether these forms were transcriptionally active and contributed to the observed transgene expression. Total cellular DNA was isolated from rAAV/ova-inoculated skeletal muscle at 15 weeks postinjection and digested with a no-cut restriction enzyme (*Eco*RI) and was then treated with PS-DNase to degrade linear genomic fragments. By performing TaqMan analysis with a *gus* primer-probe set, we verified that >99.9% of the genomic DNA signal had been eliminated. The PS-DNase-resistant rAAV vector DNA (episomal form) was used as a template for circle PCR, in which PCR primers designed to amplify away from each other were positioned at a unique *Pst*I restriction site. The products of the PCR were cloned into a TA cloning vector, and bacterial colonies were screened using a vector-specific hybridization probe. Positive clones ranged in size from 0.3 to 3.2 kb in length, and 15 clones (>3.0 kb) were chosen for further analysis. To determine whether the rescued rAAV sequences were capable of expressing the ovalbumin transgene, the inserts were released from the cloning vector with *Pst*I digestion and

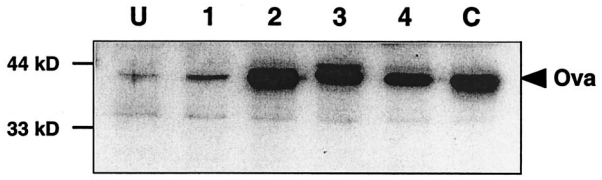


FIG. 6. Gene expression from in vivo rAAV vector forms. A circle PCR approach was used to clone the predicted in vivo monomeric circular vector form. PCR primers, overlapping the unique *Pst*I site at their 3′ ends, were used to amplify rAAV vectors from PS-DNase-treated muscle DNA. The resulting PCR products were cloned into a TA-cloning vector for further analysis. To determine whether the cloned vectors were transcriptionally active, they were excised from the TA vector with *Pst*I, self-ligated, and transfected into HeLa cells. Cell lysates were analyzed by Western blotting with an anti-ova-horseradish peroxidase antibody. Four independently rescued clones (lanes 1 to 4) that express chicken ovalbumin protein are shown. Lane U, untransfected cell lysate; lane C, lysate from cells transfected with a DNA expression plasmid encoding chicken ovalbumin.

self-ligated to recapitulate a circular episomal form. The ligated rAAV vector DNA was directly transfected into HeLa cells, and ovalbumin expression was confirmed by Western blot analysis in 6 of the 15 clones analyzed (4 are shown in Fig. 6). Sequence analysis of the clones revealed that they all possessed rearranged ITR elements, with various degrees of deletion. Deletion of a portion of the ova expression cassette was observed in three clones. Although the forms of the rearrangements and sequence deletions are similar to those of previously published in vivo ITR (15), we cannot formally rule out the possibility that the observed sequence alterations occurred as a result of the circle PCR process or upon passage through bacteria.

## DISCUSSION

To account for the observed long-lived gene expression in mouse muscle mediated by rAAV vectors, our original hypothesis was that rAAV vector genomes integrated in vivo. This prediction was primarily supported by the association of rAAV sequences with high-molecular-weight DNA on Southern blots (7, 8). Contrary to our predictions, we were unable to recover a single clone that represented vector-host cell DNA junctions from muscle tissue with very high vector copy numbers. As a secondary and complementary approach to the same question, we devised a PCR-based strategy modeled after an assay designed to determine the level of proviral human immunodeficiency virus DNA integration (3). In that human immunodeficiency virus assay, integration frequency was determined by comparing the amplification profile of unknown samples to a standard curve that was developed from the amplification profile of a mixed population of proviral integrants. Our approach was similar in concept, but the requirement to discriminate between episomal and proviral integrants necessitated a two-step design. In the first step, we used a single B1-specific (Alu-like) primer to amplify B1 amplicons, some of which contained integrated rAAV DNA sequences. The second step relied on quantitative PCR to determine the level of amplification of rAAV vector sequences (Fig. 1).

The B1 PCR assay is by its nature inefficient. B1 elements comprise about 1% of the total mouse genome and are present on average every 15 kb (25). The actual distances between B1 elements differ, and shorter B1 amplicons are expected to amplify more efficiently than longer amplicons. Hence, in certain instances large B1 amplicons containing integrated vector sequences might not be detected in the assay. In fact, our data directly support this proposition. We detected an rAAV integration event in 1 out of every 196 rAAV/eGFP-zeo-transduced C2C12 cells assayed. However, it is important to note that the C2C12 rAAV-transduced pool allowed us to control for the inefficiency of rAAV detection and define the limits of sensitivity for the assay. The integrated pool was composed of mouse muscle (C2C12) cells that were transduced with an rAAV vector at a high MOI ($10^4$ DRP/cell). This potentially mimicked in vivo transduction conditions and, following drug selection, generated a population of cells that had the rAAV vector genome randomly integrated into the C2C12 muscle cell genome. Thus, by subjecting this pool of DNA to the B1 PCR assay, we experimentally defined the limits of detection on this heterogeneous mixture of integrants (relative to B1 amplicon position and length) following random rAAV vector integration into host genomic DNA.

Importantly, none of the mice that had received rAAV injections into muscle had detectable integrated vector, and considering the limit of detection, this indicated that at least 99.5% of the rAAV vector in skeletal muscle was not integrated. Thus, two independent experimental approaches (direct cloning and B1 PCR) failed to detect integrated forms of rAAV vector DNA.

These negative findings compelled us to consider an alternative hypothesis, namely, that rAAV vector DNA was persisting as extrachromosomal high-molecular-weight forms. To characterize these high-molecular-weight molecules, we used a novel DNase (PS-DNase) that digests linear molecules but not double-stranded episomes. Based on the predicted activity of PS-DNase and fragment analysis after restriction enzyme digestion, the high-molecular-weight moieties were shown to be concatameric episomes. In addition to the large episomes, we also characterized an abundant circular, double-stranded monomeric vector form. The circular nature of these molecules was predicted by their resistance to PS-DNase digestion and confirmed by direct cloning and sequencing from tissue. Sequence analyses revealed structures with deleted sequences within the viral ITR, an observation consistent with previous in vivo and in vitro data showing significant rearrangement of ITR elements following rAAV transduction (15, 32, 38, 40, 43, 57, 60). Other investigators have also reported the isolation of smaller circular rAAV vector forms from skeletal muscle which were demonstrated to be capable of directing transgene expression (13, 61). Yang et al. (61) also reported that rescued rAAV circular intermediates isolated from skeletal muscle at 120 days postinfection were able to express the transgene. The biochemistry of rAAV episome formation remains undefined and is to be the subject of future investigations.

Thus, considered together, these data argue that very little (if any) rAAV vector DNA was integrated in transduced mouse muscle and that the vast majority of the vector sequences persist as large and small concatameric episomes. In retrospect, these findings are not surprising, because the experimental system did not predict integration. The rAAV vectors used in these studies did not contain the AAV *rep* gene,

and the mouse host cell DNA did not contain the AAV S1 locus (45). Furthermore, even when conditions are permissive for in vivo site-specific integration, this unusual biologic phenomenon is not a common event (22). However, we cannot assert that rAAV vectors never integrate. In fact, for many common modalities of in vivo exogenous DNA transfer, random integration of introduced DNA (via cellular repair or recombination mechanisms) has been shown to occur. Integration has been documented to occur at low levels for nonintegrating vectors such as naked plasmid DNA and recombinant adenovirus vectors (20, 58)

An important question then arises: what is the potential risk (mutagenic or oncogenic) associated with such low-frequency events? For plasmid DNA gene transfer following intramuscular injection, theoretical calculations based on experimental data have allowed investigators to assert that the integration frequency (mutation rate) is at least 3,000 times below the spontaneous mutation rate (31, 35). When these same sorts of considerations are applied to the present rAAV data, we can derive similar numbers. Based on assay sensitivity, we would have detected a maximum level of 75 integrated vector copies in 20 ng of genomic DNA, which corresponds to a maximum of 3,750 integration events per microgram of DNA (assuming integration of a single vector copy per site). Moreover, assuming that 1 μg of mouse DNA is equivalent to $1.67 \times 10^5$ genomes and that there are $7.5 \times 10^4$ genes per mouse genome (35), there would then be a maximum of $2.9 \times 10^{-7}$ integrated rAAV vector copies/gene. Comparison of this mutation rate to the generally accepted spontaneous mutation rate of a single base in the mammalian genome ($1 \times 10^{-5}$ mutations/gene) (11) reveals that the maximum mutation rate induced by rAAV integration in this study is at least 33 times lower than the spontaneous mutation rate. This value is likely an overestimation of the actual rate, because rAAV vectors typically integrate at a single locus in head-to-tail tandem arrays. Thus, assuming a minimum of two to four rAAV vectors integrated into a single site, the maximum mutation rate is further refined to a level at least 66 to 132 times lower than that of the spontaneous mutation rate. Although these numbers are slightly lower than those estimated for plasmid DNA, it should be noted that rAAV vectors uniformly have a higher initial transduction frequency than plasmid DNA, which leads to a much higher residual vector copy number in muscle (7).

Alternatively, rAAV integration can be viewed as an insertional mutagenic event due to possible tumor suppressor gene inactivation. To derive an estimate of the maximum number of rAAV integration events per injected muscle, we can assume that there are 1.6 million nuclei per quadriceps muscle (12, 24), yielding 9.6 μg of genomic DNA. Therefore, our data predict that the maximum number of integration events per injected muscle would be $3.6 \times 10^4$ integration events (3,750 integration events per microgram). The probability that integration of an intracellular DNA molecule could inactivate a tumor suppressor gene is estimated to be at most $10^{-6}$ to $10^{-9}$ (49). Thus, the probability of a single mutagenic event occurring in a tumor suppressor gene (worst-case scenario) in the entire injected muscle mass is at most between 0.0036% and 3.6%. Moreover, since only a small percentage of cells in the body can actually become oncogenic and since carcinogenesis itself is a multistep process, the possibility that rAAV integration occurs at the maximum defined levels determined herein appears to be remote.

In conclusion, our data suggest that rAAV genomes persist in mouse muscle as transcriptionally active large and small concatameric episomes. When considered in the context of potential and theoretical integration events, rAAV vectors show similarities to plasmid DNA in muscle. The nature of rAAV vector genome persistence in tissues other than muscle and species other than mice remains to be defined. It is worth noting that rAAV vectors currently in use in human clinical trials have been very well tolerated, even at high doses given by different routes over extended periods of time (23, 26, 52, 53).

## ACKNOWLEDGMENTS

We thank the CCRI Viral Vector Core Laboratory for rAAV vector production, Michael C. Soult for expert technical assistance, Ruju Chen for assistance with animal studies, and Richard Peluso for insightful suggestions.

This work was funded by grants from the National Institutes of Health.

## REFERENCES

1. **Afione, S. A., C. K. Conrad, W. G. Kearns, S. Chunduru, R. Adams, T. C. Reynolds, W. B. Guggino, G. R. Cutting, B. J. Carter, and T. R. Flotte.** 1996. In vivo model of adeno-associated virus vector persistence and rescue. J. Virol. **70:**3235–3241.
2. **Bertran, J., J. L. Miller, Y. Yang, A. Fenimore-Justman, F. Rueda, E. F. Vanin, and A. W. Nienhuis.** 1996. Recombinant adeno-associated virus-mediated high-efficiency, transient expression of the murine cationic amino acid transporter (ecotropic retroviral receptor) permits stable transduction of human HeLa cells by ecotropic retroviral vectors. J. Virol. **70:**6759–6766.
3. **Butler, S. L., M. S. Hansen, and F. D. Bushman.** 2001. A quantitative assay for HIV DNA integration in vivo. Nat. Med. **7:**631–634.
4. **Chen, S. J., J. Tazelaar, A. D. Moscioni, and J. M. Wilson.** 2000. In vivo selection of hepatocytes transduced with adeno-associated viral vectors. Mol. Ther. **1:**414–422.
5. **Chirmule, N., K. Propert, S. Magosin, Y. Qian, R. Qian, and J. Wilson.** 1999. Immune responses to adenovirus and adeno-associated virus in humans. Gene Ther. **6:**1574–1583.
6. **Clark, K. R., X. Liu, J. P. McGrath, and P. R. Johnson.** 1999. Highly purified recombinant adeno-associated virus vectors are biologically active and free of detectable helper and wild-type viruses. Hum. Gene Ther. **10:**1031–1039.
7. **Clark, K. R., T. J. Sferra, and P. R. Johnson.** 1997. Recombinant adeno-associated viral vectors mediate long-term transgene expression in muscle. Hum. Gene Ther. **8:**659–669.
8. **Clark, K. R., T. J. Sferra, W. Lo, G. Qu, R. Chen, and P. R. Johnson.** 1999. Gene transfer into the CNS using recombinant adeno-associated virus: analysis of vector DNA forms resulting in sustained expression. J. Drug Target. **7:**269–283.
9. **Clark, K. R., F. Voulgaropoulou, D. M. Fraley, and P. R. Johnson.** 1995. Cell lines for the production of recombinant adeno-associated virus. Hum. Gene Ther. **6:**1329–1341.
10. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell **9:**91–99.
11. **Cole, J., and T. R. Skopek.** 1994. International Commission for Protection Against Environmental Mutagens and Carcinogens. Working paper no. 3. Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. Mutat. Res. **304:**33–105.
12. **Danko, I., and J. A. Wolff.** 1994. Direct gene transfer into muscle. Vaccine **12:**1499–1502.
13. **Duan, D., P. Sharma, J. Yang, Y. Yue, L. Dudus, Y. Zhang, K. J. Fisher, and J. F. Engelhardt.** 1998. Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J. Virol. **72:**8568–8577.
14. **Felgner, P. L., and G. M. Ringold.** 1989. Cationic liposome-mediated transfection. Nature **337:**387–388.
15. **Fisher, K. J., K. Jooss, J. Alston, Y. Yang, S. E. Haecker, K. High, R. Pathak, S. E. Raper, and J. M. Wilson.** 1997. Recombinant adeno-associated virus for muscle-directed gene therapy. Nat. Med. **3:**306–312.
16. **Flannery, J. G., S. Zolotukhin, M. I. Vaquero, M. M. LaVail, N. Muzyczka, and W. W. Hauswirth.** 1997. Efficient photoreceptor-targeted gene expression in vivo by recombinant adeno-associated virus. Proc. Natl. Acad. Sci. USA **94:**6916–6921.
17. **Flotte, T. R., S. A. Afione, C. Conrad, S. A. McGrath, R. Solow, H. Oka, P. L. Zeitlin, W. B. Guggino, and B. J. Carter.** 1993. Stable in vivo expression of

the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector. Proc. Natl. Acad. Sci. USA **90:**10613–10617.
18. **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.
19. **Graham, F. L., and L. Prevec.** 1991. Manipulation of adenovirus vectors. Methods Mol. Biol. **7:**109–128.
20. **Harui, A., S. Suzuki, S. Kochanek, and K. Mitani.** 1999. Frequency and stability of chromosomal integration of adenovirus vectors. J. Virol. **73:**6141–6146.
21. **Hermonat, P. L., and N. Muzyczka.** 1984. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA **81:**6466–6470.
22. **Hernandez, Y. J., J. Wang, W. G. Kearns, S. Loiler, A. Poirier, and T. R. Flotte.** 1999. Latent adeno-associated virus infection elicits humoral but not cell-mediated immune responses in a nonhuman primate model. J. Virol. **73:**8549–8558.
23. **High, K. A.** 2001. Gene transfer as an approach to treating hemophilia. Circ. Res. **88:**137–144.
24. **Hohlfeld, R., and A. G. Engel.** 1994. The immunobiology of muscle. Immunol. Today **15:**269–274.
25. **Kass, D. H., M. E. Raynor, and T. M. Williams.** 2000. Evolutionary history of B1 retroposons in the genus Mus. J. Mol. Evol. **51:**256–264.
26. **Kay, M. A., C. S. Manno, M. V. Ragni, P. J. Larson, L. B. Couto, A. McClelland, B. Glader, A. J. Chew, S. J. Tai, R. W. Herzog, V. Arruda, F. Johnson, C. Scallan, E. Skarsgard, A. W. Flake, and K. A. High.** 2000. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. Nat. Genet. **24:**257–261.
27. **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.
28. **Koeberl, D. D., I. E. Alexander, C. L. Halbert, D. W. Russell, and A. D. Miller.** 1997. Persistent expression of human clotting factor IX from mouse liver after intravenous injection of adeno-associated virus vectors. Proc. Natl. Acad. Sci. USA **94:**1426–1431.
29. **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.
30. **Lebkowski, J. S., M. M. McNally, T. B. Okarma, and L. B. Lerch.** 1988. Adeno-associated virus: a vector system for efficient introduction and integration of DNA into a variety of mammalian cell types. Mol. Cell. Biol. **8:**3988–3996.
31. **Ledwith, B. J., S. Manam, P. J. Troilo, A. B. Barnum, C. J. Pauley, T. G. Griffiths, II, L. B. Harper, C. M. Beare, W. J. Bagdon, and W. W. Nichols.** 2000. Plasmid DNA vaccines: investigation of integration into host cellular DNA following intramuscular injection in mice. Intervirology **43:**258–272.
32. **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.
33. **Malik, A. K., P. E. Monahan, D. L. Allen, B. G. Chen, R. J. Samulski, and K. Kurachi.** 2000. Kinetics of recombinant adeno-associated virus-mediated gene transfer. J. Virol. **74:**3555–3565.
34. **Malik, P., S. A. McQuiston, X. J. Yu, K. A. Pepper, W. J. Krall, G. M. Podsakoff, G. J. Kurtzman, and D. B. Kohn.** 1997. Recombinant adeno-associated virus mediates a high level of gene transfer but less efficient integration in the K562 human hematopoietic cell line. J. Virol. **71:**1776–1783.
35. **Martin, T., S. Parker, R. Hedstrom, T. Le, S. Hoffman, J. Norman, P. Hobart, and D. Lew.** 1999. Plasmid DNA malaria vaccine: the potential for genomic integration after intramuscular injection. Hum. Gene Ther. **10:**759–768.
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. **Miao, C. H., R. O. Snyder, D. B. Schowalter, G. A. Patijn, B. Donahue, B. Winther, and M. A. Kay.** 1998. The kinetics of rAAV integration in the liver. Nat. Genet. **19:**13–15.
38. **Miller, D. G., E. A. Rutledge, and D. W. Russell.** 2002. Chromosomal effects of adeno-associated virus vector integration. Nat. Genet. **30:**147–148.
39. **Monahan, P. E., and R. J. Samulski.** 2000. Adeno-associated virus vectors for gene therapy: more pros than cons? Mol. Med. Today **6:**433–440.
40. **Nakai, H., Y. Iwaki, M. A. Kay, and L. B. Couto.** 1999. Isolation of recombinant adeno-associated virus vector-cellular DNA junctions from mouse liver. J. Virol. **73:**5438–5447.
41. **Nakai, H., T. A. Storm, and M. A. Kay.** 2000. Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. J. Virol. **74:**9451–9463.
42. **Nakai, H., S. R. Yant, T. A. Storm, S. Fuess, L. Meuse, and M. A. Kay.** 2001. Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J. Virol. **75:**6969–6976.
43. **Rutledge, E. A., and D. W. Russell.** 1997. Adeno-associated virus vector integration junctions. J. Virol. **71:**8429–8436.
44. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. **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.
46. **Snyder, R. O., C. H. Miao, G. A. Patijn, S. K. Spratt, O. Danos, D. Nagy, A. M. Gown, B. Winther, L. Meuse, L. K. Cohen, A. R. Thompson, and M. A. Kay.** 1997. Persistent and therapeutic concentrations of human factor IX in mice after hepatic gene transfer of recombinant AAV vectors. Nat. Genet. **16:**270–276.
47. **Song, S., J. Embury, P. J. Laipis, K. I. Berns, J. M. Crawford, and T. R. Flotte.** 2001. Stable therapeutic serum levels of human alpha-1 antitrypsin (AAT) after portal vein injection of recombinant adeno-associated virus (rAAV) vectors. Gene Ther. **8:**1299–1306.
48. **Song, S., P. J. Laipis, K. I. Berns, and T. R. Flotte.** 2001. Effect of DNA-dependent protein kinase on the molecular fate of the rAAV2 genome in skeletal muscle. Proc. Natl. Acad. Sci. USA **98:**4084–4088.
49. **Temin, H. M.** 1990. Overview of biological effects of addition of DNA molecules to cells. J. Med. Virol. **31:**13–17.
50. **Tratschin, J. D., I. L. Miller, M. G. Smith, and B. J. Carter.** 1985. Adeno-associated virus vector for high-frequency integration, expression, and rescue of genes in mammalian cells. Mol. Cell. Biol. **5:**3251–3260.
51. **Vincent-Lacaze, N., R. O. Snyder, R. Gluzman, D. Bohl, C. Lagarde, and O. Danos.** 1999. Structure of adeno-associated virus vector DNA following transduction of the skeletal muscle. J. Virol. **73:**1949–1955.
52. **Wagner, J. A., A. H. Messner, M. L. Moran, R. Daifuku, K. Kouyama, J. K. Desch, S. Manley, A. M. Norbash, C. K. Conrad, S. Friborg, T. Reynolds, W. B. Guggino, R. B. Moss, B. J. Carter, J. J. Wine, T. R. Flotte, and P. Gardner.** 1999. Safety and biological efficacy of an adeno-associated virus vector-cystic fibrosis transmembrane regulator (AAV-CFTR) in the cystic fibrosis maxillary sinus. Laryngoscope **109:**266–274.
53. **Wagner, J. A., T. Reynolds, M. L. Moran, R. B. Moss, J. J. Wine, T. R. Flotte, and P. Gardner.** 1998. Efficient and persistent gene transfer of AAV-CFTR in maxillary sinus. Lancet **351:**1702–1703.
54. **Wallace, B., and C. Miyada.** 1987. Oligonucleotide probes for the screening of recombinant DNA libraries, p. 432–442. *In* S. L. Berger and A. R. Kimmel (ed.), Guide to molecular cloning techniques. Harcourt Brace Janovich Publishers, San Diego, Calif.
55. **Walsh, C. E., J. M. Liu, X. Xiao, N. S. Young, A. W. Nienhuis, and R. J. Samulski.** 1992. Regulated high level expression of a human gamma-globin gene introduced into erythroid cells by an adeno-associated virus vector. Proc. Natl. Acad. Sci. USA **89:**7257–7261.
56. **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.
57. **Wu, P., M. I. Phillips, J. Bui, and E. F. Terwilliger.** 1998. Adeno-associated virus vector-mediated transgene integration into neurons and other nondividing cell targets. J. Virol. **72:**5919–5926.
58. **Wurm, F. M., and C. J. Petropoulos.** 1994. Plasmid integration, amplification and cytogenetics in CHO cells: questions and comments. Biologicals **22:**95–102.
59. **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.
60. **Yang, C. C., X. Xiao, X. Zhu, D. C. Ansardi, N. D. Epstein, M. R. Frey, A. G. Matera, and R. J. Samulski.** 1997. Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. J. Virol. **71:**9231–9247.
61. **Yang, J., W. Zhou, Y. Zhang, T. Zidon, T. Ritchie, and J. F. Engelhardt.** 1999. Concatamerization of adeno-associated virus circular genomes occurs through intermolecular recombination. J. Virol. **73:**9468–9477.