JOURNAL OF VIROLOGY, Oct. 2000, p. 9451–9463 Vol. 74, No. 20
0022-538X/00/$04.00+0

Recruitment of Single-Stranded Recombinant Adeno-Associated Virus Vector Genomes and Intermolecular Recombination Are Responsible for Stable Transduction of Liver In Vivo

HIROYUKI NAKAI, THERESA A. STORM, AND MARK A. KAY*

Program in Human Gene Therapy, Departments of Pediatrics and Genetics, Stanford University School of Medicine, Stanford, California 94305

Received 16 March 2000/Accepted 28 July 2000

Recombinant adeno-associated virus (rAAV) vectors stably transduce hepatocytes in experimental animals. Following portal-vein administration of rAAV vectors in vivo, single-stranded (ss) rAAV genomes become double stranded (ds), circularized, and/or concatemerized concomitant with a slow rise and, eventually, steady-state levels of transgene expression. Over time, at least some of the stabilized genomes become integrated into mouse chromosomal DNA. The mechanism(s) of formation of stable ds rAAV genomes from input ss DNA molecules has not been delineated, although second-strand synthesis and genome amplification by a rolling-circle model has been proposed. To begin to delineate a mechanism, we produced rAAV vectors in the presence of bacterial *Pae*R7 or Dam methyltransferase or constructed rAAV vectors labeled with different restriction enzyme recognition sites and introduced them into mouse hepatocytes in vivo. A series of molecular analyses demonstrated that second-strand synthesis and rolling-circle replication did not appear to be the major processes involved in the formation of stable ds rAAV genomes. Rather, recruitment of complementary plus and minus ss genomes and subsequent random head-to-head, head-to-tail, and tail-to-tail intermolecular joining were primarily responsible for the formation of ds vector genomes. These findings contrast with the previously described mechanism(s) of transduction based on in vitro studies. Understanding the mechanistic process responsible for vector transduction may allow the development of new strategies for improving rAAV-mediated gene transfer in vivo.

Recombinant adeno-associated virus (rAAV) vectors can safely transduce various tissues and result in persistent gene expression in vivo (16, 17, 27, 31, 34–36). Determining the mechanism(s) and rate-limiting step(s) for transduction is of great importance in improving on these vectors' ability to transduce a larger proportion of cells in vivo. An unclarified issue is how the stable double-stranded (ds) rAAV genomes are formed from input single-stranded (ss) rAAV genomes.

Following intraportal administration of vector into animals, rAAV-mediated gene expression in the liver slowly increases over a period of 4 to 6 weeks concomitant with a loss of ss rAAV genomes. During this period, some of the rAAV genomes become integrated into host chromosomal DNA as concatemers in a small fraction of the hepatocytes that take up the vector following its infusion into the mouse liver (25, 26, 28, 35). The reason why only a small percentage of hepatocytes become stably transduced is not known, but it appears to be unrelated to the cell cycle status (25). There is increasing evidence that intermolecular recombination of rAAV genomes is involved in the formation of rAAV concatemers in the liver (25, 29), a process recently established in muscle tissue, where heteroconcatemers were shown to be formed (42).

While these studies have provided important information, they do not completely resolve whether the stable ds rAAV genomes result from complex intermolecular recruitment of the input genomes, new DNA synthesis, or some combination of the two. Most of the current data favor the importance of newly synthesized DNA in forming the final rAAV DNA structures. Two separate studies have suggested that second-strand synthesis from the 3′ end of the hairpin structure of the inverted terminal repeats (ITRs) is the rate-limiting step for AAV vector transduction in vitro and in vivo (10, 11). Because integrated concatemers of wild-type AAV genomes are found exclusively in a head-to-tail tandem array in vitro (4, 21, 24, 41), replication of virus genomes has been proposed to occur by a rolling-circle model. Similar structures of rAAV vectors have been found in vivo, suggesting that a similar mechanism may be operative there (5, 12, 26, 28, 35, 37, 39, 40). However, as described here, this mechanism(s) does not seem to contribute to the formation of the stable ds rAAV genomes responsible for persistent transgene expression in quiescent hepatocytes in vivo. We propose that intracellular recruitment of plus and minus ss rAAV genomes, not second-strand synthesis, may be essential for stable in vivo transduction and that intermolecular joining, and not genome amplification, is the major mechanism of rAAV genome concatemerization in hepatocytes in vivo.

MATERIALS AND METHODS

Establishment of a 293 cell line that constitutively expresses the *Pseudomonas aeruginosa Pae*R7 methyltransferase or the *Escherichia coli* Dam methylase. The procedure for cloning a 293 cell line that stably expresses the *Pae*R7 methyltransferase (293PMT) was described previously (30). A 293 cell line that stably expresses the Dam methylase (293damC4) was created as follows. The plasmid pTP166 (22), which contains the gene encoding the *E. coli* Dam methylase, was kindly provided by New England Biolabs. A 0.8-kb *Xba*I-*Pvu*II fragment containing the *dam* gene was excised and cloned into the 3.9-kb fragment of pECFP-Nuc (Clontech, Palo Alto, Calif.) containing the cytomegalovirus immediate-early enhancer-promoter, the simian virus 40 polyadenylation [poly(A)] signal, and a separate neomycin phosphotransferase expression cassette, resulting in the construction of pCMVdamneo. G418-resistant 293 cell clones transfected with pCMVdamneo were screened for *dam* gene expression by digesting the genomic DNA extracted from each clone with *Dpn*I or *Mbo*I. A clone whose genomic

* Corresponding author. Mailing address: Departments of Pediatrics and Genetics, 300 Pasteur Dr., Rm. G305A, Stanford University, Stanford, CA 94305. Phone: (650) 498-6531. Fax: (650) 498-6540. E-mail: markay@stanford.edu.

DNA was *Dpn*I digestible and *Mbo*I indigestible (293damC4) was chosen for subsequent vector production.

Construction of rAAV vectors. Our system for nomenclature of plasmids and rAAV vectors uses a lowercase "p" at the beginning of plasmid names and "AAV" at the beginning of rAAV vector names. AAV-EF1α-LacZ (encoding the *E. coli* β-galactosidase driven by the human elongation factor 1α [EF1α] gene enhancer-promoter) and AAV-EF1α-F.IX (encoding human coagulation factor IX [hF.IX] driven by the EF1α enhancer-promoter) were produced from the vector plasmids pAAV-EF1α-LacZ and pV4.1e-hF.IX, respectively. Construction of these plasmids was described elsewhere (25, 27). AAV.ΔBamHI and AAV.ΔEcoRI were based on the vector plasmids pAAV.ΔBamHI and pAAV.ΔEcoRI, respectively, both of which were derivatives of pV4.1e-hF.IX. pV4.1e-hF.IX has adjacent *Bam*HI and *Eco*RI sites located between the EF1α enhancer-promoter and the hF.IX cDNA. To localize these two sites at the exact center of the rAAV cassette, we trimmed the EF1α enhancer-promoter by removing a 0.3-kb *Spe*I-*Cla*I fragment from its 5′ end, creating pAAV-sδEF1α-F.IX. The sequence shortening was necessary to simplify the Southern blot analyses of vector-injected animals; otherwise, the presence of homo- and heteroconcatemers in head-to-head and tail-to-tail tandem arrays would result in complex DNA patterns that would complicate interpretation. pAAV.ΔBamHI and pAAV.ΔEcoRI lack the unique *Bam*HI or *Eco*RI site, respectively, and were constructed by the following procedure. The unique *Bam*HI or *Eco*RI site in pAAV-sδEF1α-F.IX was destroyed by digestion of pAAV-sδEF1α-F.IX with either *Bam*HI or *Eco*RI, creation of blunt ends with T4 DNA polymerase, and self-ligation with T4 DNA ligase. This manipulation resulted in a 4-bp insertion in the *Bam*HI or *Eco*RI site.

All of the rAAV vectors were prepared by the adenovirus-free triple-plasmid transfection method (3, 23) with pHLP19, an AAV helper plasmid that has been proven to make no recombinant wild-type AAV particles, with a sensitivity of one functional AAV virion in a background of 10^9 rAAV particles (T. Matsushita, S. Godwin, R. Surosky, and P. Colosi, Abstr. 2nd Annu. Meet. Am. Soc. Gene Ther., abstr. 2a, 1999). rAAV vectors methylated by bacterial methylases were produced in 293PMT or 293damC4 cells. The vector purification procedure was outlined elsewhere (3). The physical vector titer was determined by a quantitative dot blot assay (20).

Characterization of methylated rAAV vectors. rAAV DNA was extracted from vector stocks for analysis of methylation status. Briefly, 10^{10} particles of rAAV were resuspended in 100 μl of DNase I buffer (20 mM Tris HCl, 1 mM $MgCl_2$, pH 8.0) containing 100 U of DNase I (Roche Molecular Biochemicals, Indianapolis, Ind.)/ml, incubated at 37°C for 1 h, and then treated with 200 μl of 2-mg/ml solution of proteinase K (Gibco BRL, Gaithersburg, Md.) in proteinase K buffer (10 mM Tris HCl, 10 mM EDTA, 5% sodium dodecyl sulfate, pH 8.0) at 50°C for 1 h. Viral DNA was extracted by the sodium iodide method, using a DNA Extractor kit (Wako Chemicals USA, Richmond, Va.). An additional 10^{10} particles were treated in the same way but without the DNase I digestion step. Denaturation and annealing of the plus and minus ss rAAV genomes were performed as previously described with a modification (1). In brief, rAAV DNA was alkaline denatured with 0.2 N sodium hydroxide, neutralized with ammonium acetate, ethanol precipitated, and resuspended in 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate). Denatured rAAV DNA was annealed by incubation at 70°C for 1 h and slow cooling (1°C/min) to 20°C. For analysis of methylation status with *Pae*R7 methyltransferase, annealed rAAV DNA from 2×10^9 rAAV particles was digested with *Not*I alone or a combination of *Not*I and *Xho*I. For analysis of methylation status with Dam methylase, the same amount of annealed rAAV DNA was digested with *Bgl*II alone, a combination of *Bgl*II and *Dpn*I, a combination of *Bgl*II and *Mbo*I, or *Sau*3AI. The controls were pV4.1e-hF.IX propagated in *E. coli* DH5α (Gibco BRL) or a *dam*$^-$ *dam*$^-$ strain, DM1 (Gibco BRL), and digested with a restriction enzyme(s) in the same manner. *Dpn*I was purchased from Roche, while all other restriction enzymes were from New England Biolabs. The digested samples and controls were electrophoresed on a 1.2% agarose gel, transferred to a nylon membrane (Duralon UV; Stratagene, Ceder Creek, Tex.), hybridized with a 1.9-kb *Xho*I/*Nde*I *lacZ* probe or a 2.3-kb *Bgl*II F.IX probe (see Fig. 1), and exposed to X-ray film.

The possibility of contamination of the vector stocks by ds circular intermediates, which could affect the results, was unlikely because of nuclease treatment and subjection to the stocks two successive CsCl gradients during vector purification. The lack of these contaminants was further verified by digesting DNase I-untreated, unannealed rAAV DNA with *Eco*RI and *Drd*I (in the case of AAV-EF1α-LacZ or AAV-EF1α-LacZ.PMT), *Eco*RI and *Hin*dIII (in the case of AAV-EF1α-F.IX.PMT), or *Xho*I (in the case of AAV-EF1α-F.IX or AAV-EF1α-F.IX.dam), followed by electrophoresis on an alkaline 1.0% agarose gel and Southern blot analysis with a 0.9-kb EF1α probe (see Fig. 1). Double-stranded head-to-tail circular molecules would generate a ~2.8-kb band (in the case of AAV-EF1α-LacZ or AAV-EF1α-LacZ.PMT), a ~3.4-kb band (in the case of AAV-EF1α-F.IX.PMT), or a ~3.2-kb band (in the case of AAV-EF1α-F.IX or AAV-EF1α-F.IX.dam) because they, unlike the other possible forms, would be covalently linked at the 3′ and 5′ ITRs. The absence of DNA bands of the predicted sizes confirmed the absence of ds circular rAAV DNA in the preparation (data not shown).

Animal procedures. Six- to 8-week-old female C57Bl/6 mice were obtained from Taconic (Germantown, N.Y.). All animal experiments were performed according to Stanford University's guidelines for animal care. Portal-vein infusions of rAAV vectors, partial hepatectomy, tail vein injection of plasmids, and factor IX enzyme-linked immunosorbent assay were all performed as previously described (19, 27, 43).

Molecular analysis of liver samples. Total DNA was extracted from mouse livers, and 20 μg of DNA was digested with 80 U of a restriction enzyme (or a combination of enzymes) at 37°C for 4 h (for *Bgl*II, *Dpn*I-*Bgl*II, *Mbo*I-*Bgl*II, and *Sau*3AI digestions) or with 100 U overnight (for the other enzyme digestions). The digested DNA was electrophoresed on a 0.8 or 1.2% agarose gel, transferred to a Duralon UV membrane, and hybridized with a ^{32}P-labeled designated probe. The copy number standards (the number of ds rAAV genomes per diploid genomic equivalent) were prepared by adding an equivalent number of pAAV-EF1α-LacZ or pV4.le-hF.IX molecules to total DNA extracted from the livers of naïve C57Bl/6 mice. Some of the samples were heat denatured in 50% formamide and electrophoresed on a 0.8% agarose gel or were alkaline denatured and run on an alkaline 1.0% agarose gel.

To analyze the *Dpn*I cutting activity on hemimethylated DNA under the conditions used for our study, a reconstitution experiment was performed with plasmid pV4.le-hF.IX.hemi, which has one adenine-hemimethylated *Dpn*I recognition site in the sequence. pV4.le-hF.IX.hemi was generated as follows. Plasmid pV4.le-hF.IX was isolated from *dam*-negative *E. coli* (pV4.le-hF.IX.dam−) and adenine methylated by incubation with *Taq*I methylase at a concentration of 5 U per μg of plasmid DNA at 65°C for 1 h. Complete adenine methylation at *Taq*I recognition sites (TCGA) was confirmed by the absence of *Taq*I-cleavable sites after incubation with *Taq*I at a concentration of 10 U per μg of plasmid DNA at 65°C for 1 h. Using this procedure, any *Dpn*I- and *Taq*I-overlapping recognition site (TCGATC) results in hemimethylation of an adenine residue on a GATC site, the *Dpn*I recognition site. Since there is one TCGATC site on pV4.le-hF.IX, pV4.le-hF.IX.hemi should have one adenine-hemimethylated *Dpn*I recognition site in its sequence.

In some studies, the low-molecular-weight (LMW) DNAs were removed from the liver DNA. To do this, 100 μg of total DNA was digested with 300 U of *Kpn*I at 37°C for 4 h and separated on a 0.8% agarose gel. *Kpn*I digestion facilitates DNA separation but does not cut DNA within the vector genome. Large rAAV concatemers present in the *Kpn*I-digested DNA preparations should remain near the beginning of the DNA smear in agarose gel electrophoresis. We recovered the high-molecular-weight (HMW) DNAs of over 17 kb by electroelution, precipitated them with ethanol, and then purified them by chromatography (Micro Bio-Spin 30; Bio-Rad, Hercules, Calif.). All of the gel-fractionated *Kpn*I-digested HMW DNAs recovered from 100 μg of a starting material were subjected to Southern blot analysis following restriction enzyme digestion.

PCR-based detection of mismatch-repaired heteroduplex rAAV vector genomes. PCR was employed to demonstrate mismatch base repair on heteroduplexes formed by AAV.ΔBamHI and AAV.ΔEcoRI genomes. If mismatched sequences at *Bam*HI and *Eco*RI sites on AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes were repaired in hepatocytes, some of the heteroduplexes would convert into AAV.ΔBamHI.ΔEcoRI homoduplexes, which lack both *Bam*HI and *Eco*RI sites.

Ten micrograms of total mouse liver DNA was digested with a combination of *Bam*HI and *Eco*RI (40 U of each) for 4 h to cleave AAV.ΔBamHI or AAV.ΔEcoRI homoduplexes and then purified by standard phenol-chloroform extraction. A 446-bp fragment spanning *Bam*HI and *Eco*RI sites on AAV.ΔBamHI and AAV.ΔEcoRI genomes was amplified by PCR with 1 μg of total DNA as a template, using primers EF1αP21 (5′-CGTCCAGGCACCTCG ATTAGT-3′) and hFIXP21 (5′-ACCTGAATTATACCTCTTTGGCCG-3′). The PCR conditions were 95°C for 2 min followed by 30 cycles of 95°C for 1 min, 60°C for 1 min, and 72°C for 1 min. The PCR products were purified by the use of a QIAquick PCR purification kit (Qiagen, Valencia, Calif.), which can remove primers but not PCR products or template DNA; denatured with 0.2 N sodium hydroxide; ethanol precipitated; and annealed with a ^{32}P-labeled inner primer, hFIXP22 (5′-ATTCAGAATTTTGTTGGCGTTTT-3′). A primer extension reaction was carried out by incubation with nucleosides and the Klenow fragment of DNA polymerase at 37°C for 1 h, generating ^{32}P-labeled homoduplex PCR products. Alternatively, the PCR fragments purified with the QIAquick PCR purification kit were further purified by agarose gel electrophoresis and then heat denatured, annealed with primer hFIXP22, and incubated with nucleosides containing [α-^{32}P]dCTP and *Taq* polymerase at 72°C for 10 min. These ^{32}P-labeled homoduplex PCR products synthesized by the Klenow fragment or *Taq* polymerase were used for the subsequent enzyme digestion to demonstrate the homoduplex products lacking both *Bam*HI and *Eco*RI sites. In brief, ^{32}P-labeled homoduplex PCR products were incubated with *Bam*HI and *Eco*RI in the presence of a 1:1 mixture of linearized pAAV.ΔBamHI and pAAV.ΔEcoRI. This mixture of linearized plasmids showed a 7,330-bp band before the enzyme digestion and exhibited two bands, of 4,564 and 2,766 bp, after the digestion, serving as a monitor of complete enzyme digestion. After 1.2% agarose gel electrophoresis, the gel was dried and bands were detected by using a Molecular Imager System (Bio-Rad).

RESULTS

The majority of persistent ds rAAV genomes contain input-vector genomes. As a molecular tool to help establish whether

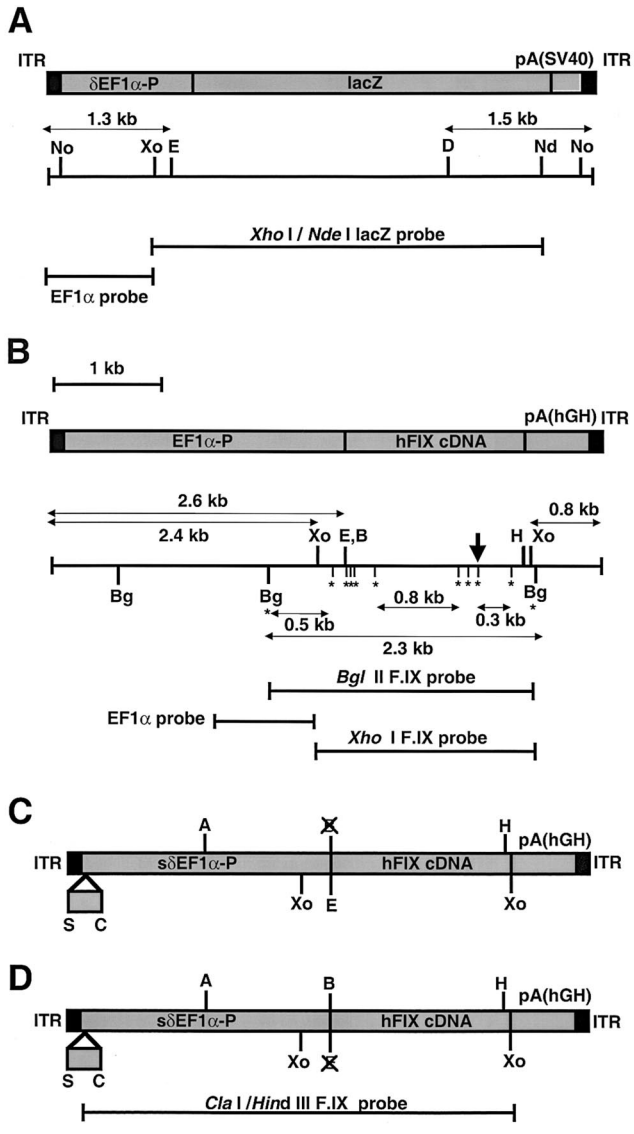


FIG. 1. Structures of the rAAV vectors, restriction enzyme sites, and DNA probes. (A) AAV-EF1α-LacZ and AAV-EF1α-LacZ.PMT. (B) AAV-EF1α-F.IX, AAV-EF1α-F.IX.PMT, and AAV-EF1α-F.IX.dam. The GATC sequences present between the two *Bgl*II sites are indicated by asterisks, and the location of a hemimethylated GATC sequence in pV4.le-hF.IX.hemi is indicated by a perpendicular arrow. (C and D) AAV.ΔBamHI (C) and AAV.ΔEcoRI (D). The 0.3-kb deletion within EF1α-P is shown. A unique *Bam*HI or *Eco*RI site has been destroyed. Abbreviations: EF1α-P, EF1α gene enhancer-promoter; δEF1α-P, a truncated EF1α-P with a 1.3-kb *Spe*I-*Mun*I deletion (28); sδEF1α-P, a truncated EF1α-P with a 0.3-kb *Spe*I-*Cla*I deletion; lacZ, the *E. coli* β-galactosidase gene; pA(SV40), the simian virus 40 early polyadenylation signal [poly(A)]; hFIXcDNA; the human coagulation factor IX cDNA; pA(hGH), the human growth hormone gene poly(A); A, *Afl*II; B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; D, *Drd*I; E, *Eco*RI; H, *Hin*dIII; Nd, *Nde*I; No, *Not*I; S, *Spe*I; Xo, *Xho*I.

genome replication mechanisms were responsible for the concatemeric rAAV DNA structures resulting in stable transduction in vivo, adenine-methylated and nonmethylated DNA forms of two rAAV vectors, either AAV-EF1α-LacZ.PMT or AAV-EF1α-LacZ and either AAV-EF1α-F.IX.PMT or AAV-EF1α-F.IX, were produced (Fig. 1). With methylation of adenine residues at the CTCGAG sequences, the *Xho*I recognition sites became noncleavable (14) and served as a marker of input molecules versus newly synthesized DNA (30). By Southern blot analysis, we confirmed that more than 99% of the AAV-EF1α-LacZ.PMT and AAV-EF1α-F.IX.PMT genomes were methylated (Fig. 2A). In eukaryotic cells, which lack adenine methylases and demethylases, methylated adenine residues in the viral genome can be lost only if second-strand synthesis occurs, generating hemimethylated rAAV duplexes (*Xho*I insensitive), or if replication of rAAV genomes occurs either autonomously or as an integrated proviral genome during cellular DNA replication, resulting in fully unmethylated rAAV DNA duplexes (*Xho*I sensitive). A total of 1.5×10^{11} particles of AAV-EF1α-LacZ.PMT ($n = 3$) or AAV-EF1α-LacZ ($n = 2$) were injected via the portal vein into adult C57Bl/6 mice, and their livers were harvested 6 weeks postinjection. Southern blot analysis of the liver DNA from AAV-EF1α-LacZ.PMT-injected mice showed that a majority of the genomes remained resistant to *Xho*I digestion (Fig. 2B). Because this experiment did not differentiate fully methylated and hemimethylated DNA, we could only conclude that the rAAV genomes did not undergo ds DNA replication.

In a second experiment, 2.4×10^{11} (group 1; $n = 6$) and 7.2×10^{11} (group 2; $n = 3$) particles of AAV-EF1α-F.IX.PMT were infused via the portal vein into adult C57Bl/6 mice. Similar to what we had observed previously, the levels of human factor IX in the plasma slowly increased over a period of 6 weeks, reaching levels (means ± standard deviations) of 875 ± 234 ng/ml (for 2.4×10^{11} particles) and 1,915 ± 150 ng/ml (for 7.2×10^{11} particles), compared to a level of 893 ± 394 ng/ml in the control mice that received 2.4×10^{11} particles of AAV-EF1α-F.IX (group 3; $n = 4$). This confirmed that methylation of the vector did not affect its ability to express a transgene product. To demonstrate that some of the methylated vector likely represented integrated DNA forms in vivo, at 6 weeks post-vector administration, a partial (two-thirds) hepatectomy was performed on six mice to stimulate conditions under which most hepatocytes would divide once or twice ($n = 2$ in each group), resulting in some proportion of the nonintegrated DNA being diluted and/or lost. Liver tissue removed at the time of partial hepatectomy was analyzed by Southern blotting. The results showed that there were 0.76 ± 0.23 (groups 1 and 3) or 2.06 ± 0.69 (group 2) ds rAAV genomes per diploid genome equivalent (Fig. 2C), and the rAAV genomes appeared to be predominantly head-to-tail molecules consisting of either concatemers or circular monomeric forms (data not shown). The rAAV genome copy number and concatemeric arrangements were consistent with our previously published results (26, 28, 35). After liver regeneration, there was a 5- to 10-fold drop in the number of rAAV genome copies per diploid genomic equivalent, confirming the presence of nonintegrated episomal ds rAAV genomes (Fig. 2C). In nonregenerating livers, virtually all of the ds rAAV genomes in the AAV-EF1α-F.IX.PMT-injected mice were not cut with *Xho*I, while as expected, proviral *Xho*I-digestible genomes appeared when hepatocytes underwent cell division in vivo (Fig. 2D). The lack of fully unmethylated DNA was not consistent with the previously proposed rolling-circle model, which requires synthesis of both DNA strands leading to head-to-tail concatemerization of rAAV genomes (24). The origin of the very small number of *Xho*I-digestible genomes present in some of the nonhepatectomized mice was not known.

HMW large concatemers are formed by intermolecular ligation of input rAAV genomes. Because the majority of the rAAV proviral genomes did not arise from ds DNA replication after in vivo administration, we hypothesized that the stable proviral concatemers were formed by intermolecular ligation of input rAAV genomes. This hypothesis is consistent with the recent finding of heteroconcatemers from two different rAAV genomes in mouse muscle and liver tissues (25, 29, 42). To test

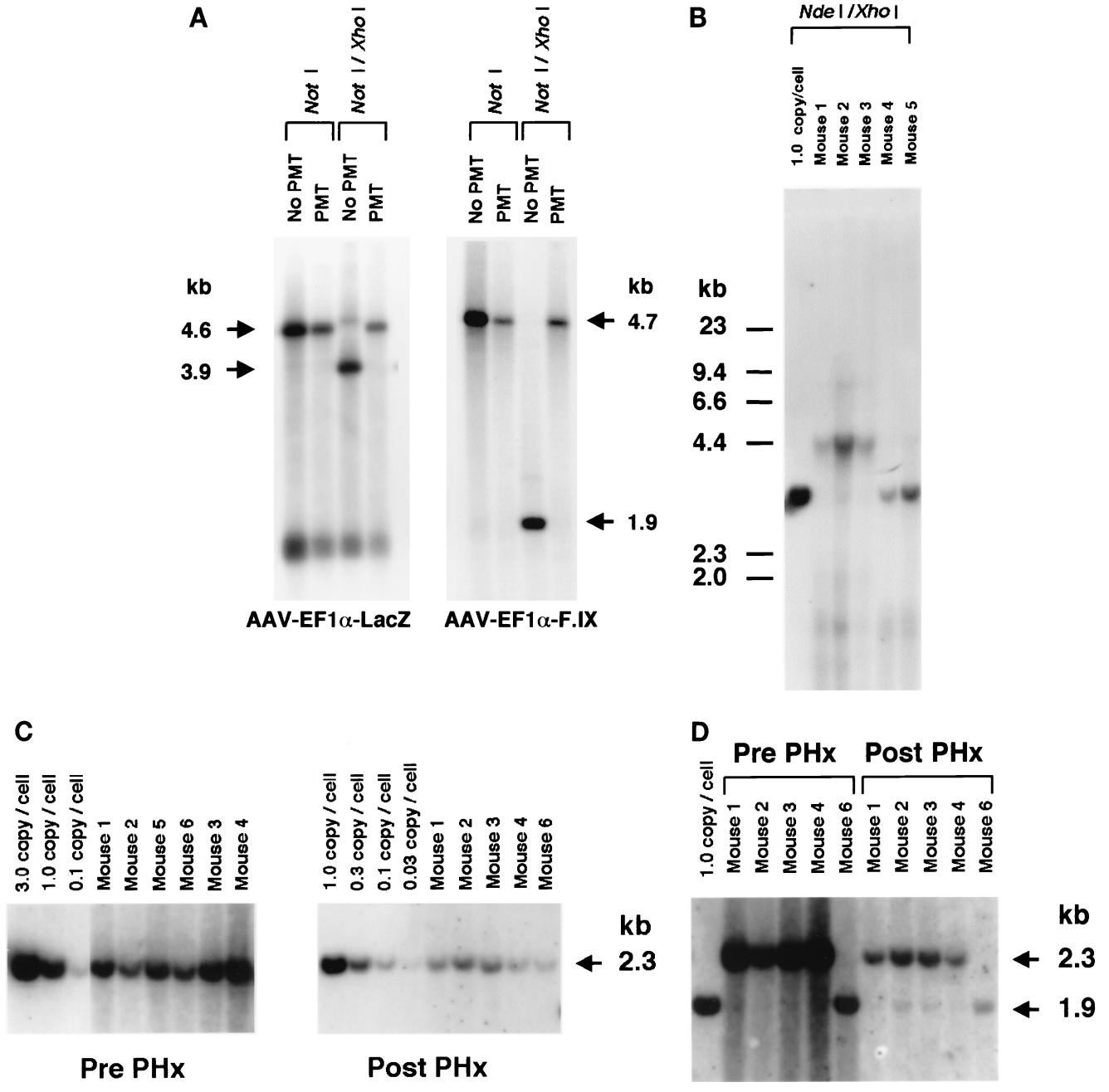


FIG. 2. Southern blot analysis of the AAV.PMT vectors and transduced mouse livers. (A) Determination of methylation status at the *Xho*I sites on AAV-EF1α-LacZ (no PMT or not methylated) and AAV-EF1α-LacZ.PMT vectors (left panel) and on AAV-EF1α-F.IX (no PMT) and AAV-EF1α-F.IX.PMT vectors (right panel). ss viral DNA corresponding to 2.0×10^9 particles was extracted from purified stocks, annealed, and digested with the enzyme(s) indicated above the lanes prior to Southern analysis using *Xho*I-*Nde*I *lacZ* and *Xho*I F.IX probes (left and right panels, respectively). Note that no *Xho*I-digestible genomes were observed in the PMT lanes. (B) Methylation status of rAAV genomes from the livers of mice 6 weeks after AAV-EF1α-LacZ or AAV-EF1α-LacZ.PMT injection. The liver DNA was digested with *Nde*I and *Xho*I prior to Southern analysis using the *Xho*I-*Nde*I *lacZ* fragment as a probe. Faint *Xho*I-digestible genomes were observed, but most of the genomes remained uncut in the AAV-EF1α-LacZ.PMT-injected mice (mice 1, 2, and 3), while all the genomes were cut with *Xho*I in AAV-EF1α-LacZ-injected mice (mice 4 and 5). One-unit-length 4.4-kb bands represent head-to-tail molecules because *Nde*I results in a single cut in the absence of *Xho*I digestion. (C) The number of vector genomes per cell (diploid genomic equivalent) before and after two-thirds partial hepatectomy (PHx) in the mice injected with AAV-EF1α-F.IX or AAV-EF1α-F.IX.PMT. Mice 1 and 2 were injected with 2.4×10^{11} particles of AAV-EF1α-F.IX.PMT, mice 3 and 4 were injected with 7.2×10^{11} particles of AAV-EF1α-F.IX.PMT, and mice 5 and 6 were injected with 2.4×10^{11} particles of AAV-EF1α-F.IX. The liver DNA was digested with *Bgl*II and probed with a *Bgl*II F.IX fragment. (D) Methylation status of rAAV genomes from livers of mice injected with AAV-EF1α-F.IX or AAV-EF1α-F.IX.PMT before and after two-thirds hepatectomy. The liver DNA was digested with *Bgl*II and *Xho*I and probed with a *Bgl*II F.IX fragment. Virtually all of the genomes remained uncut by *Xho*I before hepatectomy, but *Xho*I-digestible genomes were present after the procedure. The mouse DNAs used were the same as those employed for the experiment shown in panel C.

our hypothesis, we constructed two vectors that differed by only 8 bases and altered a unique restriction enzyme recognition site in the center of the viral vector genome. The strategic placement of these sites provided a relatively simple means to analyze the results (see Materials and Methods). AAV.ΔBamHI, which has an *Eco*RI site but lacks a *Bam*HI site (Fig. 1C), and AAV.ΔEcoRI, which has a *Bam*HI site but lacks the *Eco*RI site (Fig. 1D), were mixed at various ratios (group 4, 1:0; group 5, 3:1; group 6, 1:1; and group 7, 0:1), and a total of 2.8×10^{11} particles were injected into adult C57Bl/6 mice via the portal vein ($n = 3$/group). These mixed vectors were all functional since there was no significant difference in plasma hF.IX levels at 6 weeks after injection among these four groups (means ± standard deviations, 7,828 ± 2,323 ng/ml, 7,587 ± 666 ng/ml, 7,002 ± 3,114 ng/ml, and 10,099 ± 4,956 ng/ml in groups 4, 5, 6, and 7, respectively). Six weeks after rAAV injection, the mice were sacrificed and total DNA was extracted from their livers. When undigested DNA was examined by Southern blot analysis using a vector-specific probe, all 12 mice from the four groups had an HMW DNA signal, presumed to represent integrated rAAV concatemers (26, 28) and a band between the positions of the ss vector and 1 unit length of the ds rAAV genome, representing supercoiled ds circular monomers (8) (data not shown). To analyze for intermolecular joining in concatemers, the liver DNA was digested with either *Bam*HI or *Eco*RI and subjected to Southern blot analysis. If concatemeric HMW rAAV genomes originated from amplification of a single rAAV genome, either *Bam*HI or *Eco*RI digestion of liver DNA from animals that received both AAV.ΔBamHI and AAV.ΔEcoRI vectors (group 5 or 6) would result in a superimposable pattern, as with DNA from animals in groups 4 and 7 that received only AAV.ΔBamHI or AAV.ΔEcoRI vector. In contrast, if a concatemer comprised both AAV.ΔEcoRI and AAV.ΔBamHI vectors, a DNA ladder consisting of 2-, 3-, or >3-unit (rAAV genome)-length bands would be expected (Fig. 3A). As shown in Fig. 3B and 3C, the *Bam*HI- or *Eco*RI-digested liver DNA formed a ladder in mice injected with both vectors (groups 5 and 6), indicating the presence of heteroconcatemers.

Since rAAV genomes are ss, DNA preparations from transduced tissues may form noncovalently linked multimers observed as ladders upon electrophoresis on a neutral agarose gel. Although most of the ss genomes should not remain in the livers 6 weeks after injection, we cannot completely rule out the possibility that the ladders observed after either *Bam*HI or *Eco*RI digestion simply represent low levels of noncovalently linked multimers originating from ss rAAV genomes, rather than covalently linked heteroconcatemers. To exclude this possibility, denatured DNA samples were electrophoresed on a 0.8% agarose gel and subjected to Southern blot analysis. DNA ladders were observed in samples from mice injected with both AAV.ΔEcoRI and AAV.ΔBamHI vectors, but not in samples from mice injected with either vector alone, confirming the presence of covalently linked heteroconcatemers (Fig. 3D).

While these results are consistent with the joining of input rAAV genomes into concatemers, the more prominent 1-unit-length DNA fragment signal (Fig. 3B and C) may have formed from amplification of a single vector, resulting in concatemers. Alternatively, a 1-unit-length fragment may have originated from ds circular molecules that contaminated the DNA preparations. If there was no genome amplification of single molecules, and if all rAAV genomes were equally recruited to form the HMW concatemers, digestion of DNA from group 6 animals (AAV.ΔBamHI/AAV.ΔEcoRI = 1:1) with either *Bam*HI or *Eco*RI should result in single- and double-unit-length DNA signals of similar intensity. To distinguish between these different possibilities, we removed the LMW rAAV DNA forms, which could contribute to a 1-unit-length band when digested with *Bam*HI or *Eco*RI. To do this, *Kpn*I-digested liver DNA (*Kpn*I does not cut the vector sequences) was electrophoresed on an agarose gel, and genome-sized DNAs of over 17 kb were recovered. Southern blot analysis using this fractionated HMW DNA, representing the HMW concatemers, showed a dramatically decreased intensity of the 1-unit-length band, to a level close to the 1:1 ratio of 1- to 2-unit-length fragments in DNA from a group 6 animal (Fig. 3E). In addition, as expected, when the mice received more AAV.ΔBamHI than AAV.ΔEcoRI (group 5), the 1-unit-length band was fainter than the 2-unit-length band when the DNA was digested with *Bam*HI (Fig. 3E). These findings are consistent with our hypothesis that concatemers observed in the rAAV-transduced liver result from recruitment of input rAAV genomes, rather than from a single rAAV genome as would be expected if a rolling-circle model was solely operational (24). Although this coinjection experiment could not totally exclude the possibility that dimer or larger concatemers were formed first by the recruitment of rAAV genomes followed by replication, these amplified genomes, if present, could not account for the majority of the rAAV DNA because genome amplification was not detected in the methylation studies described above.

Previous studies, including our own, have claimed that the head-to-tail configuration is the major form for rAAV concatemers (5, 17, 26, 28, 37, 39). Because linking of rAAV genomes may occur randomly in terms of their orientation, this is in potential conflict with the evidence provided above demonstrating the absence of vector genome amplification by the rolling-circle mechanism. To reevaluate the prevalence of head-to-tail linking in rAAV concatemers, we did a Southern blot analysis using *Kpn*I digested, gel-fractionated HMW DNA containing only rAAV concatemers, with no LMW circular monomers present. Although ds rAAV vector genomes were found predominantly in a head-to-tail array when total DNA was used for the analysis, the prevalence of head-to-tail molecules was lower in the HMW DNA fraction, and there were more head-to-head and tail-to-tail forms in concatemers than we had previously observed (Fig. 3F and data not shown). Similarly, when transduced hepatocytes underwent cell division after partial hepatectomy at 6 weeks postinjection (groups 1 and 2), episomal forms (mainly circular monomers) were lost, resulting in smaller numbers of head-to-tail molecules (data not shown).

Thus, we conclude that rAAV concatemers are primarily formed by intermolecular ligation of input rAAV genomes and not solely by rolling-circle amplification of a single genome. Furthermore, the linkage orientations in concatemers are relatively random.

Recruitment of ss genomes, and not second-strand DNA synthesis, is responsible for formation of persistent ds rAAV genomes. Another interesting finding of coinjection of two vectors identical with the exception of a single restriction site (AAV.ΔBamHI and AAV.ΔEcoRI) was the presence of ds rAAV genomes in concatemers and circular monomers that could not be cut with either *Bam*HI or *Eco*RI. When the liver DNA from AAV.ΔBamHI-injected mice (group 4) or AAV.ΔEcoRI-injected mice (group 7) was digested with *Eco*RI or *Bam*HI, respectively, all of the ds rAAV genomes, including HMW concatemers, circular monomers, and ds linear monomers, were cleaved, resulting in a single-unit or faint half-unit fragments in addition to smear signals representing integrated DNA (Fig. 3B and C). However, when the liver DNA from mice injected with both vectors was digested with a combination of *Bam*HI and *Eco*RI, some *Bam*HI- and *Eco*RI-nondi-

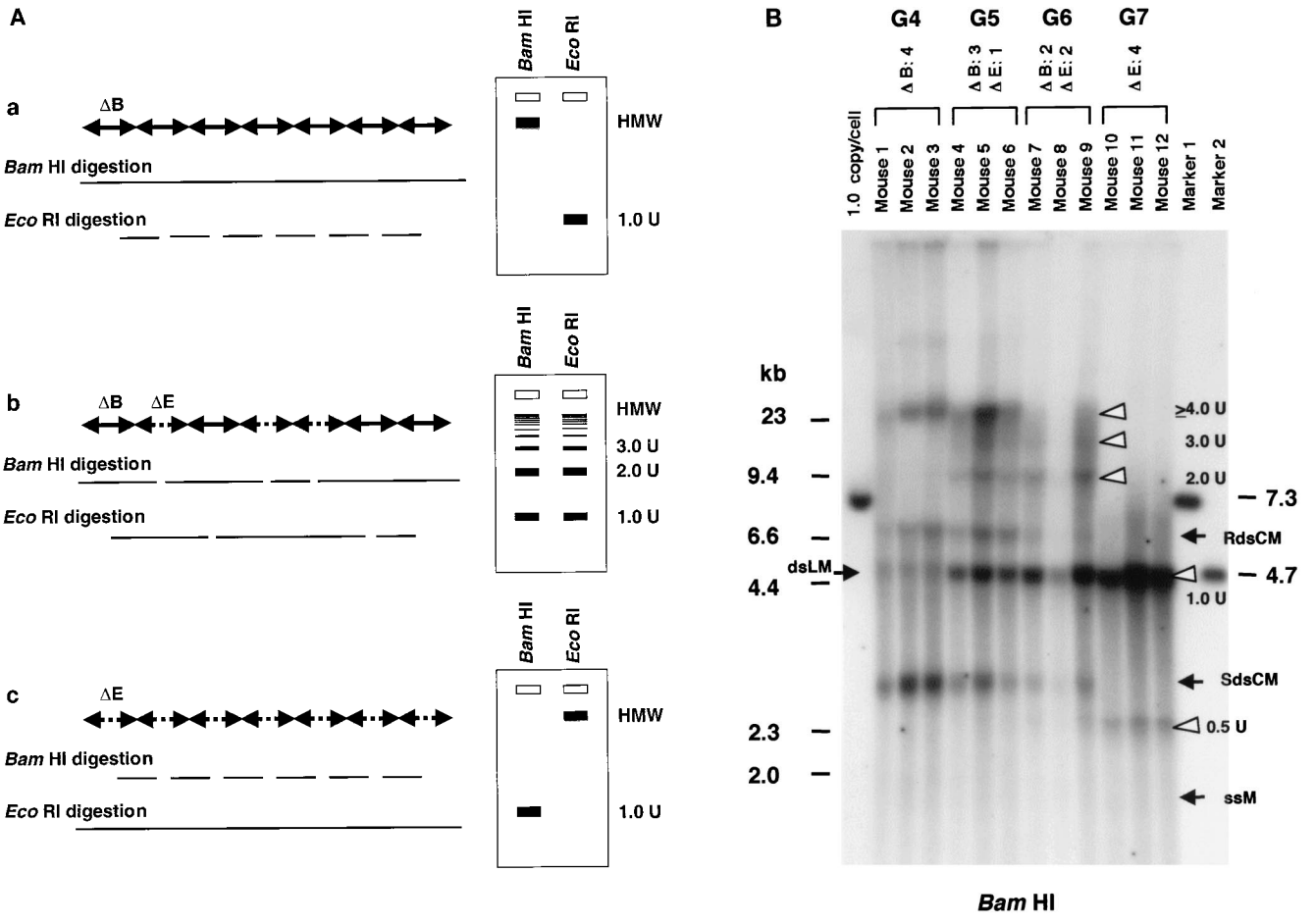


FIG. 3. Southern blot analyses of liver DNA from mice transduced with AAV.ΔBamHI and/or AAV.ΔEcoRI. (A) Schematic representation of HMW rAAV concatemer structural analysis using AAV.ΔBamHI and AAV.ΔEcoRI. Possible concatemer structures are shown on the left, and the corresponding Southern blot results after *Bam*HI or *Eco*RI digestion are shown on the right. The solid line and dotted line with arrowheads represent AAV.ΔBamHI (ΔB) and AAV.ΔEcoRI (ΔE) genomes, respectively. If the HMW concatemers consist of multiple copies from a single rAAV genome amplified by a rolling-circle mechanism (schemes a and c), Southern blot results for liver DNA from the mice injected with both vectors at a 1:1 ratio should show a mixture of scheme a and c products, 1-unit-length fragments, and a genomic-DNA-sized signal for either the *Bam*HI or *Eco*RI digest. If the recruitment of input rAAV genomes is the mechanism (scheme b), a ladder consisting of *n*-unit fragments would be observed for either enzyme digest. In the actual DNA samples from both vector-injected mice, the presence of *Bam*HI- and *Eco*RI-nondigestible rAAV genomes (see Results) would complicate the interpretation of each *n*-unit-sized fragment in the ladder. However, a ladder should never be observed in a rolling-circle amplification mechanism of concatemerization. (B and C) *Bam*HI and *Eco*RI digestion, respectively. *Bam*HI in group 4 (G4) and *Eco*RI in group 7 (G7) served as noncutter controls. Discrete genomic-DNA-sized signals in DNA indicate the presence of large concatemers. The ladder formation consisting of *n*-unit fragments in groups 5 (G5) and 6 (G6) indicates the presence of heteroconcatemers. Supercoiled ds circular monomer (SdsCM) increases and relaxed ds circular monomer (RdsCM) decreases in size with restriction enzyme digestion to the 1.0-unit-size position when cut with *Bam*HI or *Eco*RI, while ds linear monomer (dsLM) moves down to the 0.5-unit-size position when cut with either enzyme. The position of the ss monomeric rAAV genome is indicated as ssM, although it was not observed. Open arrowheads indicate positions of n-unit-length framents. Marker 1 was a 7.3-kb marker; marker 2 was a 4.7-kb marker representing 1 unit length. The *Cla*I-*Hin*dIII F.IX probe was used for these blots. (D) Southern blot analysis demonstrating the presence of covalently linked heteroconcatemers. *Bam*HI-digested total mouse liver DNAs were prepared and electrophoresed on a 0.8% gel in the same manner as that used for the experiment shown in panel B, except for the inclusion of heat denaturation of digested mouse liver DNA in 50% formamide before the samples were loaded on the gel. Ladders were observed in mice injected with both vectors (groups 5 and 6), as indicated by open arrowheads, but not in a mouse injected with a single AAV.ΔEcoRI vector (group 7), demonstrating the presence of covalently linked heteroconcatemers. Black arrows indicate head-to-head or tail-to-tail molecules that could be cut with *Bam*HI, resulting in 1.0-unit-sized ss molecules that folded back to 0.5-unit-sized ds molecules with a hairpin structure by intramolecular annealing during electrophoresis under neutral conditions. This was confirmed by alkaline gel electrophoresis, since no >1.0-unit-sized fragment was observed in group 7 mice after *Bam*HI digestion (data not shown). Although we demonstrated the formation of ladders, we could not define some of the bands because of the folding-back phenomenon observed with head-to-head or tail-to-tail molecules. Marker 1, *Bam*HI-digested naïve-mouse liver DNA (20 μg) with 1.0 copy of pV4.le-hF.IX per cell, serving as a monitor for complete digestion and as an approximate 0.5-unit-length size marker. Marker 2, *Pvu*II-digested pAAV.ΔBamHI added to *Bam*HI-digested naïve-mouse liver DNA (20 μg) at 1.0 copy per cell, serving as an approximate 1.0-unit-length size marker. The size markers on the left represent the positions of ds DNA fragments. (E) Southern blot analysis of fractionated HMW DNA (≥17 kb) following *Kpn*I digestion. The HMW DNAs of one or two mice from groups G5, G6, and G7 were digested with *Bam*HI and subjected to Southern blot analysis with a *Cla*I-*Hin*dIII F.IX probe. The intensities of 1.0- and 2.0-unit-length bands were almost the same in G6, and the intensity of the 2.0-unit-length band was higher than that of the 1.0-unit-length band in G5, which is suggestive of random recruitment of input rAAV genomes in a concatemer. (F) Southern blot analysis of total DNA (left panel) or *Kpn*I-digested gel-fractionated HMW DNA (right panels) to establish linking orientations in HMW concatemers. Liver samples from two mice in group 4 (mice 2 and 3) were chosen for this analysis. The purified HMW DNA was digested with *Afl*II or *Hin*dIII (see Fig. 1C for their cleavage site locations) and subjected to Southern blot analysis with the *Cla*I-*Hin*dIII F.IX probe. Only the results of *Afl*II digestion are shown here. Open and closed arrowheads indicate positions of head-to-tail (H-T) and tail-to-tail (T-T) molecules, respectively. Although the bands were fuzzy and covered with the enriched smear hybridization indicative of integrated genomes, the head-to-tail form was not the predominant form of HMW concatemers, unlike the situation for total DNA. (G) Southern blotting of DNA from mice in groups 5 and 6 to look for BamHI- and EcoRI-nondigestible ds rAAV genomes. *Bam*HI-*Eco*RI-*Xho*I triple digestion generated a 1.7-kb band when rAAV genomes had either a *Bam*HI or an *Eco*RI site, while a 1.9-kb band was seen when they lacked both sites. All four mice analyzed had a 1.9-kb band in addition to the major 1.7-kb band. The 2.0-unit-length band, RdsCM, and SdsCM, which remained uncut when treated with *Bam*HI-*Eco*RI, indicate the presence of *Bam*HI- and *Eco*RI-nondigestible ds rAAV genomes in concatemers and circular monomers. Incomplete enzyme digestion can be ruled out from the complete digestion of the 1.0-copy-number standards (20 μg of naïve-mouse genomic DNA containing a 1:1 mixture of pAAV.ΔBamHI and pAAV.ΔEcoRI equivalent to 1.0 copy per cell). Marker 1 is the 1.9-kb *Xho*I-*Xho*I fragment. The *Xho*I F.IX probe was used.

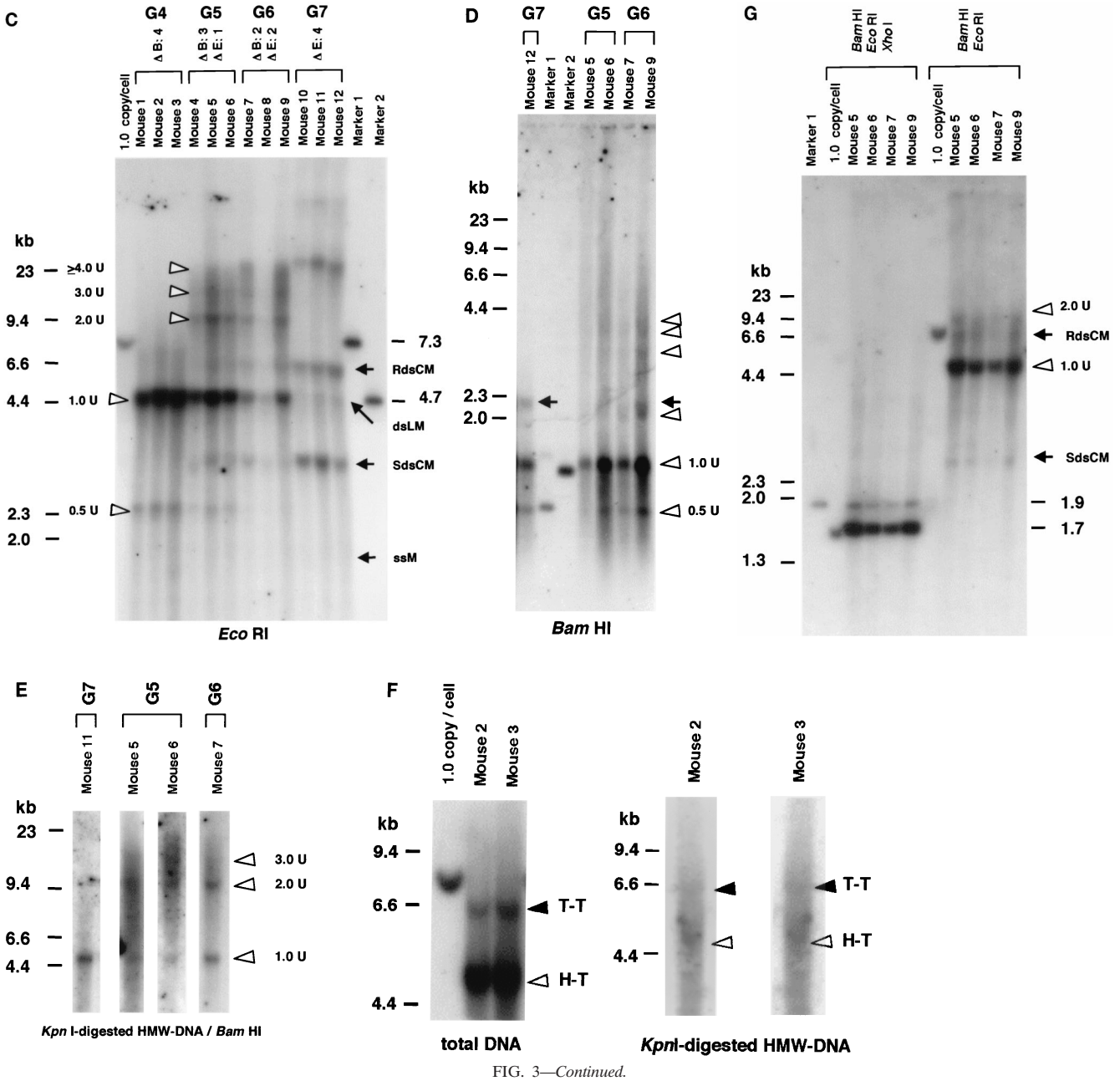


FIG. 3—*Continued.*

gestible ds rAAV genomes were observed (Fig. 3G). Although these *Bam*HI- and *Eco*RI-nondigestible ds genomes may have originated from AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes generated during DNA extraction, at least some of them were formed in hepatocytes. This is because *Bam*HI-*Eco*RI double digestion revealed that some of the *Bam*HI- and *Eco*RI-nondigestible genomes consisted of supercoiled and relaxed ds circular monomers and concatemers of two or more rAAV genomes (Fig. 3G). These molecular species could not be formed during DNA extraction.

To obtain definitive evidence that annealed input plus and minus ss rAAV genomes contributed to stable liver transduction, we produced an rAAV vector (AAV-EF1α-F.IX.dam) in cells that express the *E. coli* Dam methylase. This vector had exactly the same sequence as that of AAV-EF1α-F.IX except that all of the adenine residues at GATC sites were methylated (Fig. 1B). Three restriction endonucleases, *Dpn*I, *Mbo*I, and *Sau*3AI, were used to distinguish annealed plus and minus ss rAAV genomes (fully methylated) from ds rAAV duplexes formed by second-strand synthesis (hemimethylated). This approach has been widely used to monitor DNA replication in eukaryotic cells and to distinguish full methylation from hemimethylation (2, 6, 15, 18). However, because *Dpn*I can also partially cut fully unmethylated DNA under conditions of high enzyme-to-DNA ratios (Fig. 4B, lane 6) and others reported *Dpn*I cutting activity on hemimethylated DNA (38), we tested the extent to which *Dpn*I cleaved hemimethylated GATC sites under the conditions used for these analyses.

We generated pV4.le-hF.IX.hemi, a plasmid that contained one adenine-hemimethylated *Dpn*I recognition site, by using

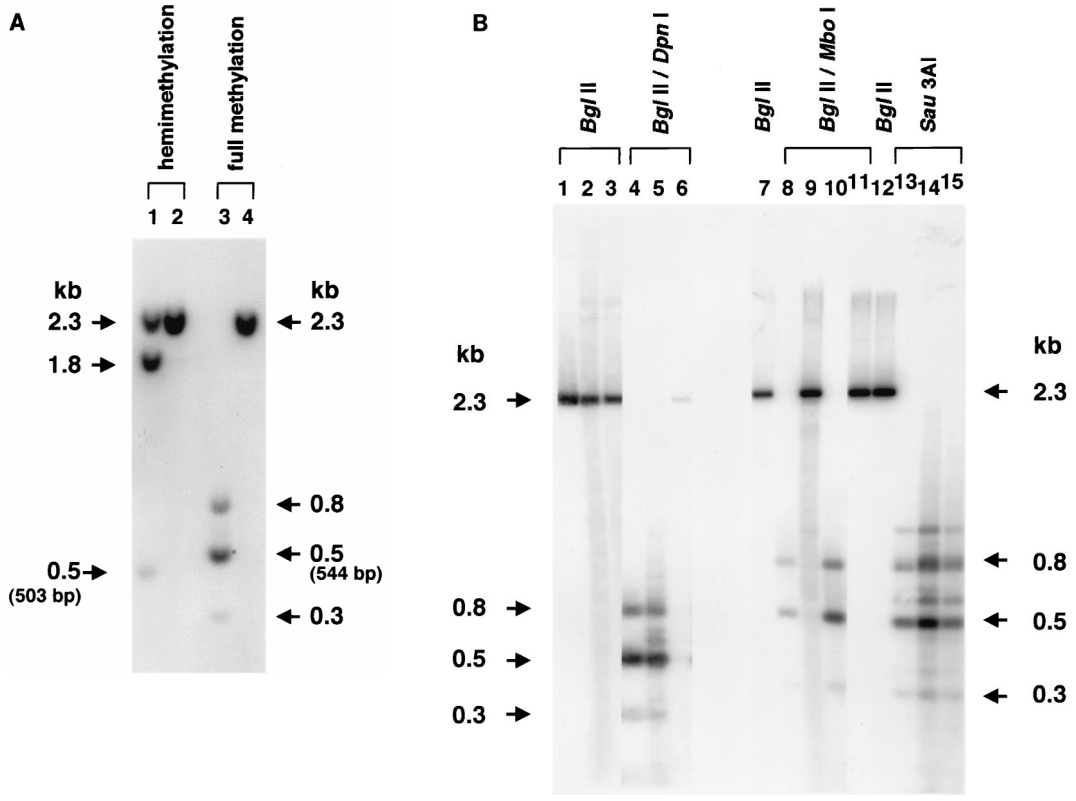


FIG. 4. (A) *Dpn*I cutting activity on hemimethylated DNA. Plasmid pV4.le-hF.IX.hemi, which carried one adenine-hemimethylated *Dpn*I recognition sequence (see Fig. 1B for its location), was synthesized with *Taq*I methylase. Nineteen copies of this plasmid (lanes 1 and 2) or a control fully methylated plasmid, pV4.le-hFIX.dam+ (pV4.le-hF.IX propagated in a *dam*-positive strain of *E. coli*), per cell were added to 20 μg of naïve-mouse genomic DNA and digested with a combination of *Dpn*I and *Bgl*II (lanes 1 and 3) or *Bgl*II only (lanes 2 and 4) under the same conditions used for the analyses shown in Fig. 4C, and the digest was subjected to Southern blot analysis with the *Bgl*II F.IX probe. Nineteen copies of pV4.le-hF.IX.hemi per cell represents the same molar ratio of hemimethylated *Dpn*I sites as that in 1.0 copy of ds vector genomes per cell. *Dpn*I completely digested the fully methylated plasmid, while about 50% activity was observed on hemimethylated plasmid. (B and C) Southern blot analysis of the AAV-EF1α-F.IX.dam vector. (B) Analysis of methylation status at the GATC sites on the AAV-EF1α-F.IX and AAV-EF1α-F.IX.dam vectors by Southern blotting. ss viral DNAs corresponding to 2.0×10^9 particles were annealed, digested with the enzyme(s) indicated above the lanes, and electrophoresed on a 1.2% agarose gel along with the markers (*Bgl*II-digested [lanes 1 and 12], *Bgl*II- and *Dpn*I-digested [lane 4], *Bgl*II- and *Mbo*I-digested [lane 11], or *Sau*3AI-digested [lane 13] pV4.le-hF.IX.dam+ or *Bgl*II-digested [lane 7] or *Bgl*II- and *Mbo*I-digested [lane 8] pV4.le-hF.IX.dam− [pV4.le-hF.IX propagated in a *dam*-negative strain of *E. coli*]). The DNA blot was probed with the *Bgl*II F.IX probe. Lanes 2, 5, 9, and 14, AAV-EF1α-F.IX.dam; lanes 3, 6, 10, and 15, AAV-EF1α-F.IX. *Dpn*I-digestible genomes are clearly observed in AAV-EF1α-F.IX.dam (lane 5). Under the conditions used here for *Dpn*I digestion (10 U of *Dpn*I for 2.0×10^9 ss rAAV genomes or 5.4 ng of viral DNA), *Dpn*I showed some cutting activity on fully unmethylated ds DNA; this resulted in a faint 2.3-kb band in lane 6 due to partial digestion of the *Dpn*I sites, but no discrete bands (except for the 2.3-kb band). However, this activity was completely repressed in the presence of mouse genomic DNA at a concentration of 0.25 μg per unit of enzyme (see panel C). (C) Methylation status of ds rAAV genomes in mouse liver at 6 weeks after injection of AAV-EF1α-F.IX.dam. The liver DNA was extracted and electrophoresed on a 1.2% agarose gel after digestion with the enzyme(s) indicated above the lanes, and the blot was analyzed with the *Bgl*II F.IX probe. Mice 1 to 3 received AAV-EF1α-F.IX.dam vector, and mice 4 and 5 received AAV-EF1α-F.IX vector. Most of the ds rAAV genomes in mice 1 to 3 remain fully methylated. The 1.0-copy-number standards were prepared using pV4.le-hF.IX.dam+ except for the right-most lane, which contained pV4.le-hF.IX.dam−.

*Taq*I methylase. Since the average rAAV vector copy number in the livers of injected mice was about 1.0 (containing 19 *Dpn*I sites) per diploid equivalent (shown below), to maintain equimolar amounts of hemimethylated *Dpn*I sites, mouse genomic DNA containing hemimethylated ds rAAV vector genomes at 1.0 copy per cell was reconstituted by adding 19 copies of pV4.le-hF.IX.hemi per cell to naïve-mouse liver DNA and digested with *Bgl*II and *Dpn*I under the same conditions used for the analysis for vector-injected liver samples. We also did the same analysis with mouse genomic DNA containing 1.0 copy of pV4.le-hF.IX.hemi per cell. As a control, pV4.le-hF.IX.dam+ (dam methylated) was added to naïve-mouse liver DNA and digestion was performed in the same manner. As shown in Fig. 4A, in both reconstitution experiments, *Dpn*I completely cleaved fully methylated DNA, while ~50% cutting activity on hemimethylated DNA was observed under the conditions used for this study.

Methylation of the adenine residues at GATC sites (the *Dpn*I, *Mbo*I, and *Sau*3AI recognition sites) within AAV-EF1α-F.IX.dam vector genomes was confirmed by a *Dpn*I resistance assay (Fig. 4B). A total of 2.4×10^{11} particles of AAV-EF1α-F.IX.dam were injected into mice ($n = 3$) in the same manner as described above, and the liver DNA was collected 6 weeks postinjection. There were no significant differences in the hF.IX expression level (mean ± standard deviation, 1,372 ± 446 ng/ml) or the vector copy number (1.02 ± 0.17 copy/diploid genomic equivalent) at 6 weeks in the AAV-EF1α-F.IX.dam-injected mice and those of the AAV-EF1α-F.IX- or AAV-EF1α-F.IX.PMT-injected mice described earlier. The methylation status of the GATC sequences was analyzed by digestion of the liver DNA with *Dpn*I-*Bgl*II, *Mbo*I-*Bgl*II, or *Sau*3AI. Virtually all of the ds rAAV genomes in AAV-EF1α-F.IX.dam-injected mice were cut with *Dpn*I but not with *Mbo*I. As expected, DNA from the mice injected with AAV-EF1α-F.IX remained uncut by *Dpn*I but was cleaved by *Mbo*I (Fig. 4C). Since the cutting activity on hemimethylated DNA is

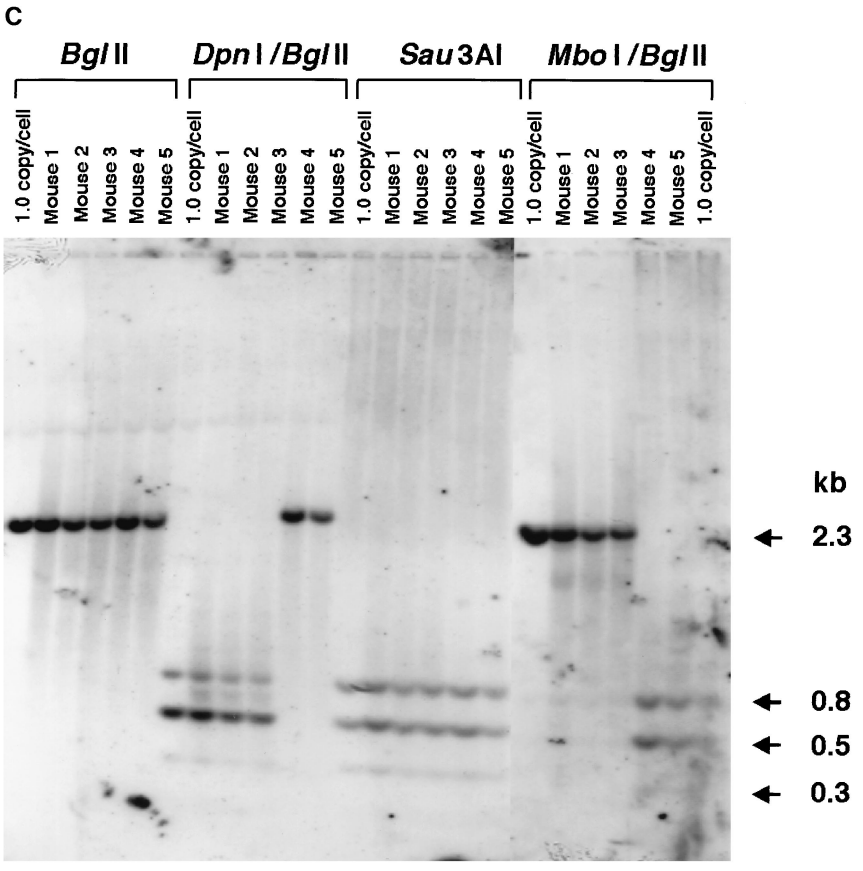


FIG. 4—*Continued.*

~50% under these conditions, the majority of the ds vector genomes should be fully methylated to obtain a complete digestion pattern with *Dpn*I as observed in Fig. 4C. If all of the ds vector genomes had been generated by second-strand synthesis, the chance of getting completely digested fragments with *Dpn*I would have been ~50% × ~50% = ~25% per fragment. This would have resulted in DNA ladders consisting of multiple partially digested fragments, and not the pattern observed in the blot shown in Fig. 4C. Because the majority of detectable ds rAAV genomes remained fully methylated, this further demonstrated that ds rAAV genomes responsible for gene expression had been formed by annealing of input complementary ss rAAV genomes.

Mismatch repair on heteroduplex rAAV genomes: additional support for annealing of complementary ss rAAV vector genomes in hepatocytes. As discussed above, when AAV.ΔBamHI and AAV.ΔEcoRI vectors were introduced into the mouse liver at a 1:1 ratio, some *Bam*HI- and *Eco*RI-nondigestible ds genomes were observed (Fig. 3G). We originally assumed that they were the result of formation of annealed AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes. However, the predicted AAV.ΔBamHI homoduplex/AAV.ΔBamHI-AAV.ΔEcoRI heteroduplex/AAV.ΔEcoRI homoduplex ratio was 1:2:1. Therefore, the intensities of the 1.7-kb band (either *Bam*HI- or *Eco*RI-digestible ds genomes, representing homoduplexes) and 1.9-kb band (*Bam*HI- and *Eco*RI-nondigestible ds genomes, representing heteroduplexes) should have been the same. However, the actual quantity of BamHI- and EcoRI-nondigestible ds genomes was smaller than expected (Fig. 3G). Since efficient repair of base mismatches on heteroduplex DNA in mammalian cells had been reported (13), we hypothesized that some of the AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes were converted into homoduplexes by mismatch repair (i.e., either *Bam*HI- or *Eco*RI-digestible homoduplexes, *Bam*HI- and *Eco*RI-digestible homoduplexes, or *Bam*HI- and *Eco*RI-nondigestible homoduplexes).

To demonstrate the presence of *Bam*HI- and *Eco*RI-nondigestible homoduplexes formed from mismatch-repaired AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes in hepatocytes, we developed a PCR-based assay as described in Materials and Methods and the legend to Fig. 5A. We analyzed two DNA preparations. Sample 1 consisted of the total liver DNA from a mouse injected with both AAV.ΔBamHI and AAV.ΔEcoRI at a 1:1 ratio (mouse 9, in group 6), while sample 2 was a 1:1 mixture of total liver DNA from a mouse injected with only AAV.ΔBamHI (mouse 1, in group 4) and a mouse injected with only AAV.ΔEcoRI (mouse 10, in group 7). Sample 2, which served as a negative control, contained both AAV.ΔBamHI and AAV.ΔEcoRI genomes but no mismatch-repaired genomes. When the PCR products purified by the use of a Qiagen PCR purification kit were used as a template for a primer extension reaction with a ^{32}P-labeled inner primer (hFIXP22), in addition to the expected 422-bp product we obtained a truncated homoduplex product, due to premature termination of the polymerase reaction, and faint smeared bands (Fig. 5B, lanes 5, 7, 9, and 11). The faint smeared products that annealed with the inner primer were presumed to be rAAV genomes amplified independently of the PCR primer, as reported by Snyder et al. (33). In the negative control (sample 2), all of the homoduplexes were digested with

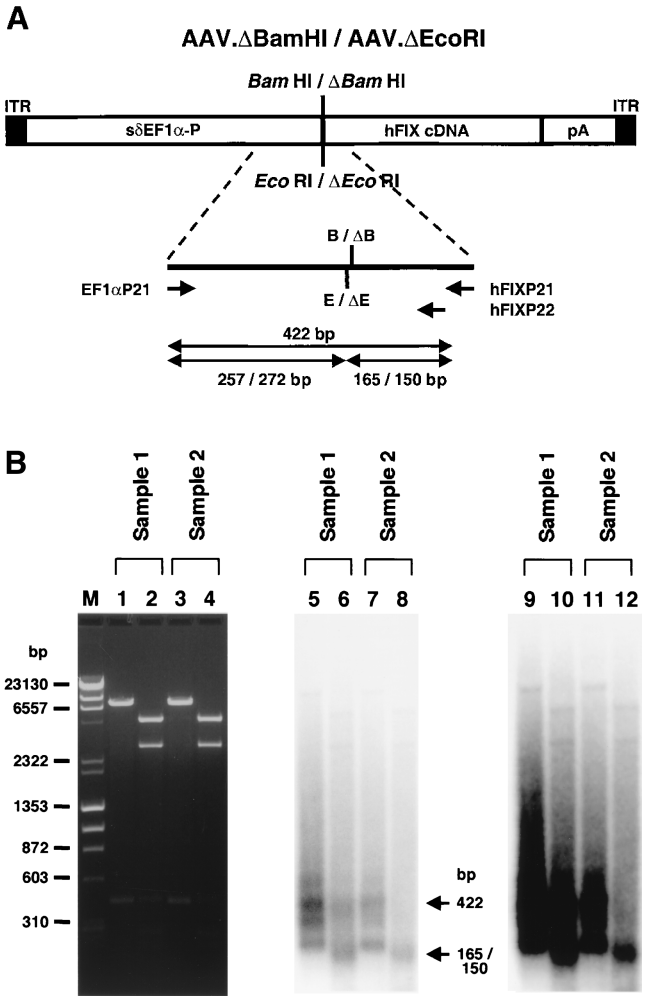
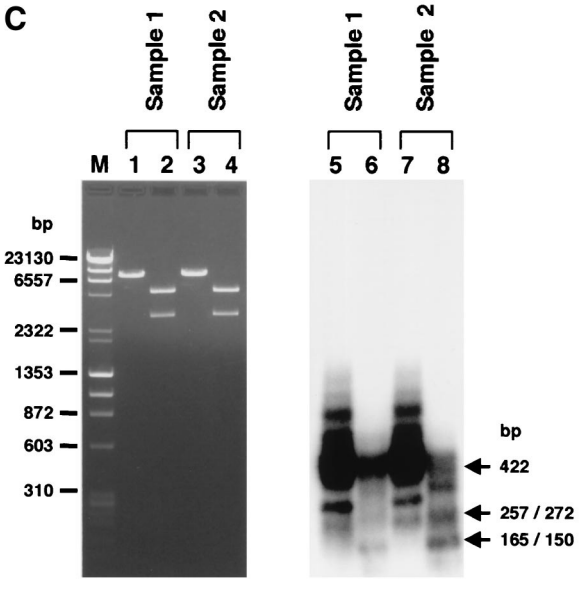


FIG. 5. PCR-based detection of mismatch repair of heteroduplex rAAV genomes in mice injected with a combination of AAV.ΔBamHI and AAV.ΔEcoRI vectors. (A) Primer locations and lengths of the products. EF1αP21 and hFIXP21 were used for PCR, and hFIXP22 was used for primer extension of the PCR products. The 422-bp product was cut into 257- and 165-bp fragments at the *Eco*RI-Δ*Eco*RI site and into 272- and 150-bp fragments at the BamHI-ΔBamHI site. pA, polyadenylation signal; B and ΔB, *Bam* HI and Δ*Bam*HI sites, respectively; E and ΔE, *Eco*RI and Δ*Eco*RI sites, respectively. (B) Total liver DNA from mouse 9 (group 7), which was injected with both AAV.ΔBamHI and AAV.ΔEcoRI vectors (sample 1), and a 1:1 mixture of two liver DNAs from mouse 1 (group 4) and mouse 10 (group 7), which were injected with only one vector (sample 2), were digested with a combination of *Bam*HI and *Eco*RI and amplified by PCR with primer set EF1αP21-hFIXP21. Radioactive homoduplexes were synthesized by primer extension with a ^{32}P-labeled hFIXP22 primer, digested with *Bam*HI and *Eco*RI, and electrophoresed on a 1.2% agarose gel. The products were stained with ethidium bromide (left panel), and radioactive products were detected by a Molecular Imager System (middle and right panels). Lanes with odd numbers contained undigested products, while lanes with even numbers contained completely digested products. Complete *Bam*HI-*Eco*RI double digestion was confirmed by cleavage of linearized plasmid DNA of a 1:1 mixture of pAAV.ΔBamHI and pAAV.ΔEcoRI added in the reaction mixtures (lanes 2 and 4). In sample 2 (which serves as a negative control with no mismatch-repaired genomes), all of the homoduplex products were cleavable with either *Bam*HI or *Eco*RI, generating 165- and 150-bp band (lanes 8 and 12), while some of the homoduplexes in sample 1 remained uncut (lanes 6 and 10). In the rightmost panel, the signals from smear homoduplexes were emphasized to demonstrate that radioactive homoduplex rAAV genomes amplified independently of the PCR primers in sample 1 (lane 10), but not in sample 2 (lane 12), also contained mismatch-repaired genomes. (C) An analysis identical to that of panel B was carried out with gel-purified PCR products as a template for the primer extension reaction. A considerable number of radioactive homoduplex genomes remained uncut (lane 6). Sample 2 served as a negative control with no *Bam*HI- and *Eco*RI-nondigestible radioactive homoduplexes (lane 8).



either *Bam*HI or *Eco*RI (Fig. 5B, lanes 8 and 12) as expected. In contrast, some of the homoduplexes in sample 1 remained uncut under conditions that gave complete enzyme digestion (Fig. 5B, lanes 6 and 10). A similar result was obtained when gel-purified PCR fragments was used for the primer extension followed by a complete digestion with *Bam*HI and *Eco*RI (Fig. 5C). These findings indicated that some BamHI- and EcoRI-nondigestible homoduplexes arose when mismatched bases on AAV.ΔBamHI-AAV.ΔEcoRI heteroduplexes were repaired in hepatocytes and converted into AAV.ΔBamHI.ΔEcoRI genomes.

Alternative mechanisms, such as recombination between AAV.ΔBamHI homoduplexes and AAV.ΔEcoRI homoduplexes, may have been operational because there were ds circular forms in hepatocytes that could participate in this kind of intermolecular recombination. To address this issue, we injected C57Bl/6 mice with 25 μg of plasmids (i.e., either pAAV.ΔBamHI or pAAV.ΔEcoRI, or a combination of the two at a 1:1 ratio) through the tail vein as described in Materials and Methods (n = 5 each). All of the mice expressed hF.IX at high levels initially (2,680 to 19,340 ng/ml, with an average ± standard deviation of 11,938 ± 6,223 at day 1); hF.IX levels dropped thereafter, but it was still expressed at levels of 150 to 1,620 ng/ml (average, 1,321 ± 824) at 6 weeks postinjection. Liver samples were taken from some of the mice 6 weeks postinjection and analyzed for the generation of BamHI- and EcoRI-nondigestible genomes by Southern blotting. As shown in Fig. 6, there were no BamHI- and EcoRI-nondigestible genomes present in the mice injected with both plasmids, demonstrating that this kind of recombination, if it occurs, is a rare event.

DISCUSSION

The different molecular fates of genomes. Two independent studies have concluded that second-strand synthesis is a rate-limiting step for rAAV transduction, because adenovirus proteins, UV irradiation, or genotoxic agents, which facilitate second-strand synthesis of ss rAAV genomes, can greatly augment transduction in vitro and in vivo (10, 11). However, this does not necessarily mean that the ds rAAV genomes formed via second-strand synthesis under these conditions are the mole-

cules that contribute to persistent transgene expression in vivo. In the two previous studies, cultured proliferating 293 and HeLa cells were examined at 24 h postinfection for molecular structure of the rAAV genomes. Replicative forms (monomeric or dimeric replicative forms [Rfm and Rfd, respectively]) were observed in the presence, but not in the absence, of adenovirus proteins or genotoxic treatment. In addition, in a similar study in muscle, Duan et al. detected no Rfm or Rfd molecular forms in the absence of adenovirus proteins but did observe transgene expression (7). Fisher et al. (11) analyzed mouse hepatocyte transduction by rAAV with or without adenovirus coinfection, but they examined the transduced tissue only at 3 days postinjection and not at later time points when transgene expression reached a plateau. Thus, there has been no definite evidence that transcriptionally active ds rAAV genomes responsible for persistent transgene expression at later time points are attributable to molecules formed by second-strand synthesis. In contrast, there has been only one report claiming that second-strand synthesis occurs in muscle in vivo in the absence of adenovirus proteins or genotoxic treatment (37). However, tissues were examined up to 3 days postinjection, when transgene expression would be expected to be minimal. Moreover, their experimental design could not differentiate between second-strand synthesis and head-to-head intermolecular recombination, a mechanism more consistent with what was observed in our study.

Recently, several observations have suggested that there are independent mechanisms for short-term augmentation of transgene expression by adenovirus proteins and for long-term persistent transgene expression in vivo. First, Snyder et al. reported that conversion of ss rAAV genomes to ds genomes by second-strand synthesis mediated by adenovirus coinjection resulted in only temporary augmentation of transgene expression in the mouse liver and had no effect on long-term expression (35), suggesting that Rfm and Rfd molecular forms could not participate in the ds rAAV genomes responsible for persistent gene expression. Second, Duan et al. showed that the formation of ds rAAV circles represented a pathway different from that of second-strand synthesis mediated by adenovirus E4orf6 (7). Although they assumed that ds rAAV circles were formed by second-strand synthesis (9), there was no definitive experimental evidence to support this. Because our data are most consistent with the formation of stable ds rAAV genomes by the recruitment of complementary ss rAAV genomes in the mouse liver in the absence of adenovirus proteins or genotoxic treatment, we hypothesize that in hepatocytes, replicative forms of rAAV genomes (Rfm and Rfd) generated by second-strand synthesis, if present, have a short half-life and may not participate in the molecular forms responsible for long-term transgene expression.

A novel model for rAAV transduction in vivo. Our studies are not consistent with the more widely held views on the mechanism of rAAV-mediated transduction, which include second-strand synthesis and/or a rolling-circle model. We propose a model (Fig. 7) in which annealing of the plus and minus strands of the vector genomes is a crucial step toward achieving stable transduction, although we do not know whether this is a rate-limiting step. The annealed ds rAAV genomes may become transcriptionally active and/or form molecularly stable circles or integrated concatemers by inter- or intrahomologous recombination between two ITRs (8, 25, 26, 28, 37, 42). There is some indication, but no definitive proof, that the circular molecules represent preintegrative intermediates and/or may be at least in part responsible for persistent gene expression (8). Moreover, there is now evidence that heterodimeric ds molecules exist in the muscle and liver when two different

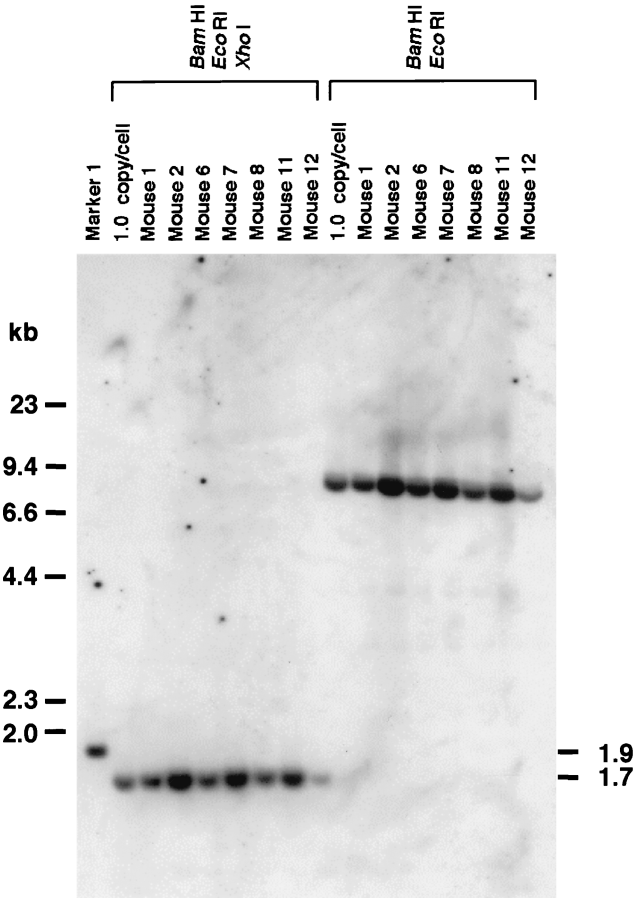


FIG. 6. Southern blot analysis to detect *Bam*HI- and *Eco*RI-nondigestible plasmids in mouse livers transfected with plasmids. Twenty-five micrograms of either pAAV.ΔBamHI (mice 1 and 2) or pAAV.ΔEcoRI (mice 11 and 12), or a combination of both (mice 6, 7, and 8), was infused through the tail vein as described in Materials and Methods, and liver samples were analyzed at 6 weeks postinjection in the same manner as that for the experiment shown in Fig. 3G. There were no *Bam*HI- and *Eco*RI-nondigestible genomes detected in these mice. 1.0-copy/cell standards, marker 1, and the probe used were the same as for the experiment shown in Fig. 3G.

rAAV vectors are injected (25, 29), and it appears that these molecules are formed in the muscle by intermolecular recombination of ds rAAV circles (42). Taken together, these results indicate that we may have previously overestimated the number of head-to-tail rAAV molecules in concatemers in vivo by underestimating the contribution of supercoiled or relaxed circular monomeric genomes to the head-to-tail molecules in earlier genomic Southern blot analyses.

In our model, the step required for in vivo transduction is presumed to be the annealing of two complementary ss rAAV genomes in the nucleus. This implies that a single particle is not capable of stably transducing quiescent cells in vivo, which may explain the requirement of a relatively high multiplicity of infection for a single transduction event in the case of rAAV. Presumably, rAAV genomes of a single polarity can transduce quiescent cells in vivo only when adenovirus helper functions or equivalent factors, such as treatment with genotoxic agents, are provided, since Samulski et al. reported that wild-type AAV genomes of a single polarity are sufficient for a permissive infection to proceed in adenovirus type 5-infected 293 cells (32).

Our present studies eliminate the notion that second-strand

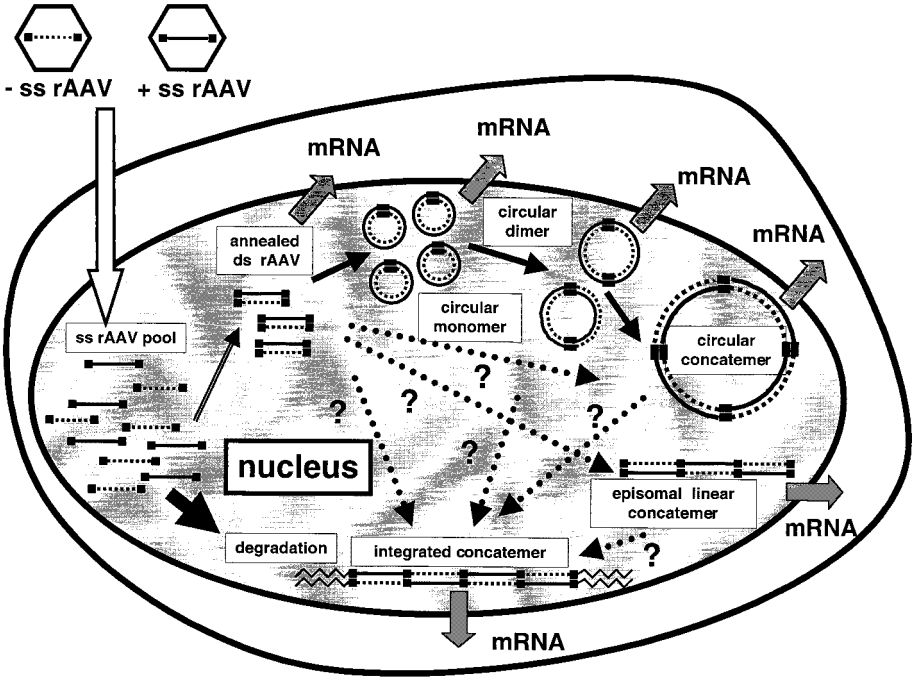


FIG. 7. Schematic representation of rAAV transduction in mouse hepatocytes. rAAV virions containing either a plus or minus ss genome enter the nucleus, creating a pool of ss rAAV genomes. Plus- and minus-strand rAAV genomes combine to form ds linear rAAV genomes. This step is presumed to be crucial for rAAV transduction in vivo, and most of the ss rAAV genomes that fail to partner are gradually degraded over a period of 1 to 2 months. Successfully annealed ds rAAV genomes are converted into a more stable form (i.e., ds circular monomer). Some of these ds circular monomers combine with each other, forming ds circular dimers or larger concatemers by intermolecular recombination through the ITRs (42). Some of the ds linear monomer may also form episomal linear concatemers. It is not yet known how integrated concatemers are formed, but possible pathways are indicated by dotted arrows with question marks. All of the ds forms—i.e., circular monomers, dimers, and larger concatemers; ds linear monomers and concatemers; and integrated concatemers—may be capable of expressing transgene (26, 42). However, which form(s) contributes to persistent transgene expression from hepatocytes is not known.

DNA synthesis or the rolling-circle mechanism is responsible for the formation of stable ds vector genomes in the liver, which has important ramifications for further study. As-yet-unidentified cellular proteins are believed to be required for stable transduction. The lack of rAAV DNA synthesis during the process of transduction changes the likely candidate cellular proteins required for the process to occur. Finally, if the recruitment of ss AAV genomes is a rate-limiting step (after the vector enters the nucleus) in stable transduction in vivo, the ability to design vectors with enhanced abilities to anneal in vivo may result in enhanced transduction efficiencies.

ACKNOWLEDGMENTS

We thank Sally Fuess and Leonard Meuse for technical assistance.

This work was supported by NIH grants ROI HL53682 and HL64274.

REFERENCES

1. **Berns, K. I., and J. A. Rose.** 1970. Evidence for a single-stranded adenovirus-associated virus genome: isolation and separation of complementary single strands. J. Virol. **5:**693–699.
2. **Blackburn, E. H., W. C. Pan, and C. C. Johnson.** 1983. Methylation of ribosomal RNA genes in the macronucleus of *Tetrahymena thermophila*. Nucleic Acids Res. **11:**5131–5145.
3. **Burton, M., H. Nakai, P. Colosi, J. Cunningham, R. Mitchell, and L. Couto.** 1999. Coexpression of factor VIII heavy and light chain adeno-associated viral vectors produces biologically active protein. Proc. Natl. Acad. Sci. USA **96:**12725–12730.
4. **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.
5. **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.
6. **Classon, M., M. Henriksson, J. Sumegi, G. Klein, and M.-L. Hammarskjold.** 1987. Elevated c-myc expression facilitates the replication of SV40 DNA in human lymphoma cells. Nature **330:**272–274.
7. **Duan, D., P. Sharma, L. Dudus, Y. Zhang, S. Sanlioglu, Z. Yan, Y. Yue, Y. Ye, R. Lester, J. Yang, K. J. Fisher, and J. F. Engelhardt.** 1999. Formation of adeno-associated virus circular genomes is differentially regulated by adenovirus E4 ORF6 and E2a gene expression. J. Virol. **73:**161–169.
8. **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.
9. **Duan, D., Z. Yan, Y. Yue, and J. F. Engelhardt.** 1999. Structural analysis of adeno-associated virus transduction circular intermediates. Virology **261:**8–14.
10. **Ferrari, F. K., T. Samulski, T. Shenk, and R. J. Samulski.** 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. **70:**3227–3234.
11. **Fisher, K. J., G. P. Gao, M. D. Weitzman, R. DeMatteo, J. F. Burda, and J. M. Wilson.** 1996. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J. Virol. **70:**520–532.
12. **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.
13. **Folger, K. R., K. Thomas, and M. R. Capecchi.** 1985. Efficient correction of mismatched bases in plasmid heteroduplexes injected into cultured mammalian cell nuclei. Mol. Cell. Biol. **5:**70–74.
14. **Gingeras, T. R., and J. E. Brooks.** 1983. Cloned restriction/modification system from *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA **80:**402–406.
15. **Hays, J. B., E. J. Ackerman, and Q. S. Pang.** 1990. Rapid and apparently error-prone excision repair of nonreplicating UV-irradiated plasmids in *Xenopus laevis* oocytes. Mol. Cell. Biol. **10:**3505–3511.
16. **Herzog, R. W., J. N. Hagstrom, S. H. Kung, S. J. Tai, J. M. Wilson, K. J. Fisher, and K. A. High.** 1997. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. Proc. Natl. Acad. Sci. USA **94:**5804–5809.
17. **Herzog, R. W., E. Y. Yang, L. B. Couto, J. N. Hagstrom, D. Elwell, P. A. Fields, M. Burton, D. A. Bellinger, M. S. Read, K. M. Brinkhous, G. M.**

Podsakoff, T. C. Nichols, G. J. Kurtzman, and K. A. High. 1999. Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector. Nat. Med. **5:**56–63.
18. **Hughes, P., A. Landoulsi, R. Kern, and M. Kohiyama.** 1989. Dam methylated and hemimethylated oriC plasmids are replicated symmetrically; a novel and general test of replication symmetry. Mol. Gen. Genet. **217:**278–280.
19. **Kay, M. A., Q. Li, T. J. Liu, F. Leland, C. Toman, M. Finegold, and S. L. Woo.** 1992. Hepatic gene therapy: persistent expression of human α1-antitrypsin in mice after direct gene delivery in vivo. Hum. Gene Ther. **3:**641–647.
20. **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.
21. **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.
22. **Marinus, M. G., A. Poteete, and J. A. Arraj.** 1984. Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene **28:**123–125.
23. **Matsushita, T., S. Elliger, C. Elliger, G. Podsakoff, L. Villarreal, G. J. Kurtzman, Y. Iwaki, and P. Colosi.** 1998. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. **5:**938–945.
24. **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.
25. **Miao, C. H., H. Nakai, A. R. Thompson, T. A. Storm, W. Chiu, R. O. Snyder, and M. A. Kay.** 2000. Nonrandom transduction of recombinant adeno-associated viral vectors in mouse hepatocytes in vivo: cell cycling does not influence hepatocyte transduction. J. Virol. **74:**3793–3803.
26. **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.
27. **Nakai, H., R. W. Herzog, J. N. Hagstrom, J. Walter, S. H. Kung, E. Y. Yang, S. J. Tai, Y. Iwaki, G. J. Kurtzman, K. J. Fisher, P. Colosi, L. B. Couto, and K. A. High.** 1998. Adeno-associated viral vector-mediated gene transfer of human blood coagulation factor IX into mouse liver. Blood **91:**4600–4607.
28. **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.
29. **Nakai, H., T. Storm, and M. Kay.** 2000. Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. Nat. Biotechnol. **18:**527–532.
30. **Nelson, J. E., and M. A. Kay.** 1997. Persistence of recombinant adenovirus in vivo is not dependent on vector DNA replication. J. Virol. **71:**8902–8907.
31. **Russell, D. W., and M. A. Kay.** 1999. Adeno-associated virus vectors and hematology. Blood **94:**864–874.
32. **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.
33. **Snyder, R. O.** 2000. AAV and RT-PCR: true or false? Mol. Ther. **1:**389–390.
34. **Snyder, R. O., C. Miao, L. Meuse, J. Tubb, B. A. Donahue, H. F. Lin, D. W. Stafford, S. Patel, A. R. Thompson, T. Nichols, M. S. Read, D. A. Bellinger, K. M. Brinkhous, and M. A. Kay.** 1999. Correction of hemophilia B in canine and murine models using recombinant adeno-associated viral vectors. Nat. Med. **5:**64–70.
35. **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.
36. **Snyder, R. O., S. K. Spratt, C. Lagarde, D. Bohl, B. Kaspar, B. Sloan, L. K. Cohen, and O. Danos.** 1997. Efficient and stable adeno-associated virus-mediated transduction in the skeletal muscle of adult immunocompetent mice. Hum. Gene Ther. **8:**1891–1900.
37. **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.
38. **Weiner, M. P., G. L. Costa, W. Schoettlin, J. Cline, E. Mathur, and J. C. Bauer.** 1994. Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. Gene **151:**119–123.
39. **Xiao, W., S. C. Berta, M. M. Lu, A. D. Moscioni, J. Tazelaar, and J. M. Wilson.** 1998. Adeno-associated virus as a vector for liver-directed gene therapy. J. Virol. **72:**10222–10226.
40. **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.
41. **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.
42. **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.
43. **Yant, S. R., L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak, and M. A. Kay.** 2000. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat. Genet. **25:**35–41.