DNA Palindromes with a Modest Arm Length of ≥20 Base Pairs Are a Significant Target for Recombinant Adeno-Associated Virus Vector Integration in the Liver, Muscles, and Heart in Mice[∇]

Katsuya Inagaki,[1] Susanna M. Lewis,[2] Xiaolin Wu,[3] Congrong Ma,[1] David J. Munroe,[3] Sally Fuess,[4] Theresa A. Storm,[4] Mark A. Kay,[4] and Hiroyuki Nakai[1]*

Department of Molecular Genetics & Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261[1]; Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, Toronto, Ontario M5G 1L8, Canada[2]; Laboratory of Molecular Technology, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702[3]; and Departments of Pediatrics and Genetics, Stanford University School of Medicine, Stanford, California 94305[4]

Our previous study has shown that recombinant adeno-associated virus (rAAV) vector integrates preferentially in genes, near transcription start sites and CpG islands in mouse liver (H. Nakai, X. Wu, S. Fuess, T. A. Storm, D. Munroe, E. Montini, S. M. Burgess, M. Grompe, and M. A. Kay, J. Virol. 79:3606-3614, 2005). However, the previous method relied on in vivo selection of rAAV integrants and could be employed for the liver but not for other tissues. Here, we describe a novel method for high-throughput rAAV integration site analysis that does not rely on marker gene expression, selection, or cell division, and therefore it can identify rAAV integration sites in nondividing cells without cell manipulations. Using this new method, we identified and characterized a total of 997 rAAV integration sites in mouse liver, skeletal muscle, and heart, transduced with rAAV2 or rAAV8 vector. The results support our previous observations, but notably they have revealed that DNA palindromes with an arm length of ≥20 bp (total length, ≥40 bp) are a significant target for rAAV integration. Up to ~30% of total integration events occurred in the vicinity of DNA palindromes with an arm length of ≥20 bp. Considering that DNA palindromes may constitute fragile genomic sites, our results support the notion that rAAV integrates at chromosomal sites susceptible to breakage or preexisting breakage sites. The use of rAAV to label fragile genomic sites may provide an important new tool for probing the intrinsic source of ongoing genomic instability in various tissues in animals, studying DNA palindrome metabolism in vivo, and understanding their possible contributions to carcinogenesis and aging.

Adeno-associated virus (AAV) is a small nonpathogenic, replication-defective DNA virus with a single-stranded DNA genome. Recombinant AAV (rAAV) is among the most promising gene delivery vectors for human gene therapy (15, 28). rAAV lacks machinery for virus genome insertion into host chromosomal DNA but does integrate into chromosomes at a low frequency (reviewed in reference 29). Murine leukemia virus preferentially integrates near transcription start sites (59), while human immunodeficiency virus favors transcribed regions for integration (49). As for rAAV, previous studies have shown that rAAV serotype 2 (rAAV2) vector integration occurs preferentially in genes and near gene regulatory sequences in mouse liver (36, 40) and in cultured human cells (32). A recent observation by Miller et al. has suggested that rAAV does not cause chromosomal breaks but integrates at preexisting chromosomal breakage sites (30).

Although rAAV vector integration frequency is considered to be low, it is important to further understand the interactions between rAAV vector and host chromosomal DNA in various tissues in experimental animals. This is because a study has demonstrated increased incidence of liver cancer in rAAV2

vector-treated animals (5), the mechanisms for which remain elusive. In addition, the use of new serotype vectors with robust transduction efficiency, such as rAAV8, can increase vector genome loads in cells, which may pose an increased risk of undesirable genomic alterations in rAAV-transduced cells. Such robust serotype vectors now are widely used for many preclinical studies for gene therapy of various human diseases. Moreover, we have hypothesized that elucidation of the mechanisms of rAAV integration may help in understanding ongoing genomic instability in living animals, which is difficult to investigate with currently available strategies but is important for studies on carcinogenesis and aging. Furthermore, it has been shown that rAAV vectors serve as a powerful tool to study the mechanisms of fundamental biological processes in cells, such as DNA damage responses and DNA repair (11).

To begin to further understand rAAV integration in various tissues of experimental animals, it is essential to establish a system by which rAAV integration sites can be identified on a large scale in nondividing cells with high efficiency and reliability without any cell manipulations. This minimizes possible technical biases and enables identification of rAAV integrations in nonhepatic tissues, for which selection is not easy to perform. However, currently available methods for high-throughput rAAV integration site analysis all rely on cell division either under a selective pressure (40) or without selection (32). Cell division is required for diluting out extrachromosomal rAAV vector genomes with high complexity, which are

* Corresponding author. Mailing address: Department of Molecular Genetics & Biochemistry, University of Pittsburgh School of Medicine, W1244 BSTWR, 200 Lothrop St., Pittsburgh, PA 15261. Phone: (412) 648-8958. Fax: (412) 624-1401. E-mail: nakaih@pitt.edu.
[∇] Published ahead of print on 8 August 2007.
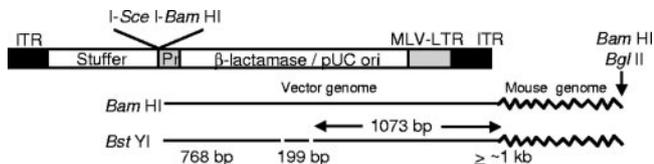
FIG. 1. AAV-ISce I.AO3 vector map. The vector genome consists (left to right) of an AAV2-ITR sequence, a stuffer sequence, an ISceI-BamHI combination site, a shortened Tn*3* prokaryotic promoter (Pr), the β-lactamase gene, the pUC plasmid origin of replication, a portion of MLV long terminal repeat (MLV-LTR), and an ITR. Diagrammed below is a schema of the preparative vector-cellular DNA junction fragments obtained by BamHI and BglII double digestion. Below that are fragments predicted for a diagnostic BstYI digestion used to analyze plasmid clones.

abundantly present in cells in a quiescent state and inhibit efficient isolation of rAAV integrants.

In the present study, we have established a novel high-throughput method to identify rAAV integration sites in non-dividing cells with high efficiency and high reliability independently of cell division, transgene expression, or selection. We have identified a thousand rAAV integration sites in quiescent somatic cells in mouse liver, skeletal muscle, and heart, and we discovered that DNA palindromes are a common target for rAAV integration. DNA palindromes are found prevalently in many organisms, including mammals (24, 58; S. M. Lewis, T. Zheng, S. Chen, T. Alleyne, J. Cheung, T. Chiang, and R. Richard, unpublished data). They have gained attention recently due to accumulating evidence that they have roles in promoting genomic instability in eukaryotes. This DNA motif has been shown to be involved in gene amplification (2, 7, 43, 46, 52–54, 60), nonrandom chromosomal translocations causing human diseases (6, 9, 12, 16–20, 55), genomic instability in animal models (1, 3, 4, 22, 23), retrovirus integration (14), and RAG protein-mediated transposition (45). Thus, the discovery that rAAV integrates preferentially at DNA palindromes not only provides further insights into the mechanisms of rAAV integration but also provides an unprecedented opportunity to study the biological impact and properties of naturally occurring DNA palindromes in tissues of living animals.

MATERIALS AND METHODS

rAAV vector production. rAAV2 and rAAV8 shuttle vectors, AAV2-ISce I.AO3 and AAV8-ISce I.AO3 (Fig. 1), were produced based on plasmid pAAV-ISce I.AO3 (accession number EU022316). The details of plasmid construction are provided upon request. This plasmid carries an ISce I.AO3 cassette between two rAAV2 inverted terminal repeats (ITRs). AAV-ISce I.AO3 genomes were packaged into AAV2 or AAV8 capsids by the triple-plasmid transfection method as previously described (34). Vector titers were determined by a quantitative dot blot assay.

Animal handling. All the animal experiments were performed according to the guidelines for animal care at Stanford University. C57BL/6J mice and DNA-dependent protein kinase catalytic subunit (DNA-PKcs)-deficient C57BL/6J SCID mice were purchased from Jackson Laboratory. Methods for tail and hepatic portal vein injections were as described previously (35).

Southern blotting. Quantitative Southern blot analysis was performed to determine vector genome copy numbers per cell (i.e., double-stranded vector genome copy numbers per diploid genomic equivalent) in each rAAV-transduced tissue as previously described (34). A β-lactamase-specific probe was used for the analysis.

Generation of rAAV vector integration site plasmid libraries. Plasmid rescue formed the basis of the strategy to isolate rAAV integration sites (31, 32, 35, 36, 40). Total DNA was extracted from mouse liver, heart, and lower limb muscle

tissues that had been transduced with rAAV vector. Ten to 30 μg of DNA was incubated at 37°C with ISceI (New England BioLabs [NEB]) at 4 U per μg DNA for 4 h. Additional ISceI enzyme was added at 2 U per μg into each reaction mixture, which was then incubated for another 4 h. After addition of an equal amount of yeast genomic DNA (*Saccharomyces cerevisiae* S288C from the ATCC), the DNA samples were treated with 1 U per μg calf intestinal alkaline phosphatase at 50°C for 1 h. The calf intestinal alkaline phosphatase-treated samples were mixed with a 0.2 volume of 2% low-melting-temperature (LMT) agarose gel in 1× Tris-EDTA (TE) buffer, poured into wells of a 1% LMT agarose gel, and electrophoresed at a low voltage overnight at 4°C. Regions of each lane containing only high-molecular-weight (HMW) DNA were excised and equilibrated first with 1× TE buffer and then with 1× NEB buffer 2 with bovine serum albumin at 4°C. In-gel digestion of HMW DNA was performed with BamHI and BglII (10 U each per μg) (NEB) at 37°C for 2 h. DNA was recovered by dissolving the gel in 1× β-agarase buffer (NEB) at 70°C for 20 min, followed by incubation with β-agarase at 42°C for 1 h, phenol-chloroform extraction, and ethanol precipitation with ammonium acetate and glycogen. The resulting DNA pellets were dissolved in water and quantified by spectrophotometry. The DNA was then self ligated at a concentration of 3 μg DNA in 700 μl of a reaction mixture containing 1,400 U of T4 DNA ligase (NEB) at 16°C overnight. The DNA preparations were purified with phenol-chloroform, followed by isopropanol precipitation with potassium acetate. Linear DNA was removed by dissolving the DNA pellets in water and incubating them with ATP-dependent exonuclease (10 U per μg DNA; Plasmid Safe; Epicenter) for at least 4 h. The DNA was purified again with phenol-chloroform, ethanol precipitated with sodium acetate, and dissolved in water. ElectroMax DH10B *Escherichia coli* (Invitrogen) cells were transformed with 1 to 3 μg of the final DNA product. The resulting plasmid libraries were plated on Luria-Bertani (LB) agar plates containing ampicillin (50 μg/ml).

High-throughput analysis of rAAV provirus plasmid libraries. Plasmid DNA was prepared from each *E. coli* colony with a Perfectprep Plasmid 96 Vac direct bind system (Eppendorf) or with manual minipreps. Each plasmid DNA was digested with BstYI and was separated on 1.2% agarose gels with a control plasmid, pAAV-ISce I.AO3, treated in the same manner. BstYI cuts at the sites generated by BamHI-BamHI, BamHI-BglII, and BglII-BglII cohesive end ligation. Therefore, BstYI digestion yields diagnostic 199- and 768-bp bands (Fig. 1). When a plasmid contains the rAAV-cellular DNA junction sequence, a third band of ≥1 kb should emerge. Therefore, we selected all the plasmids containing at least these three bands for the downstream analysis. Plasmid DNA sequence was determined as previously described with a 3730x DNA analyzer (Applied Biosystems) and sequencing primer OriP2 (40). In some cases, a sequencing primer, 36-39BamHI-1 (5′-CGACACGGAAATGTTGAATACTCAT-3′), also was used.

PCR amplification and subsequent cloning of rAAV-labeled DNA palindromes into a plasmid. Four representative DNA palindromes labeled by rAAV integrations (palindrome coordinates chr 10: 98632057, chr 11: 44726916, chr 15: 99702254, and chr 19: 11606158; see Table 3) were amplified by PCR in a 50-μl reaction mixture containing 0.5 μg naive mouse liver genomic DNA, 2× Pfx amplification buffer, 1 mM MgCl₂, 0.2 mM of each deoxynucleoside triphosphate, 0.4 μM each of forward and reverse primers, and 1 U of Platinum Pfx DNA polymerase (Invitrogen). PCR cycles were 2 min at 95°C and 34 cycles of 15 s at 95°C, 30 s at 60°C, and 30 s at 68°C, and subsequently 5 min at 68°C. The primer combinations were 49-102Pal10D1-1F1 and 49-102Pal10D1-1R1 for DNA palindrome chr 10: 98632057, 49-102Pal11B1.1-1F1 and 49-102Pal11B1.1-1R2 for DNA palindrome chr 11: 44726916, 49-102Pal15F1-1F1 and 49-102Pal15F1-1R1 for DNA palindrome chr 15: 99702254, and 49-102Pal19A-1F1 and 49-102Pal19A-1R1 for DNA palindrome chr 19: 11606158.

PCR primer sequences are the following: 49-102Pal10D1-1F1, 5′-AGCCGG GATCCTTACCTCACATTT-3′; 49-102Pal10D1-1R1, 5′-CCGCTACCAAAGC ATAAGCCGTTT-3′; 49-102Pal11B1.1-1F1, 5′-GGTGAAACACAGCTACCC TTGTGA-3′; 49-102Pal11B1.1-1R2, 5′-ACCATGTTCCAGATAAGAGGGCG T-3′; 49-102Pal15F1-1F1, 5′-TGCTGCACAAGACTAAGGACCAGA-3′; 49-102Pal15F1-1R1, 5′-GGCTGCCTCTGCATCTTGAATGTT-3′; 49-102Pal19A-1F1, 5′-ACAAAGGAAGACTGGGCAAATGGG-3′; and 49-102Pal19A-1R1, 5′-TACGGCAAGATGGTTCCTTGGAGT-3′.

The PCR products were treated with T4 polynucleotide kinase (NEB), inserted into the unique EcoRV site of pBluescript KS II(−) (Stratagene) by DNA ligation with T4 DNA ligase, and then introduced into ElectroMax DH10B *E. coli* for cloning. The transformed bacteria were plated on LB agar plates containing ampicillin (50 μg/ml).

The stability of DNA palindromes upon cloning in bacteria was assessed in the following manner. Plasmid DNA was recovered from each colony, digested with a combination of EcoRI and HindIII, which excised the cloned PCR products

containing each DNA palindrome, and analyzed by 2% agarose gel electrophoresis. In addition, plasmid DNA was sequenced with an M13 reverse primer (5′-GGAAACAGCTATGACCATG-3′).

Bioinformatics. We performed in silico digestion of the mouse genomic DNA with BamHI and BglII in the following manner. The entire sequence of each mouse chromosome (University of California Santa Cruz [UCSC] mm8 and NCBI Build 36, February 2006 freeze) was searched for BamHI (GGATCC) and BglII (AGATCT) sites using a Perl script. The fragment length was calculated from one site to the next site found.

For mapping of the rAAV integration sites, we first determined the breakage site of the rAAV vector genome using the BLAST 2 sequences (bl2seq) program of NCBI by comparing the vector genome sequence and isolated plasmid sequence. The rAAV-flanking DNA then was used as the query against the public mouse genome database (UCSC mm8 and NCBI Build 36) as previously described (40, 59). If rAAV-flanking sequences were not identified in the mouse genome, we searched them against human genome sequences and the sequences of the plasmids used for vector production, i.e., pHLP19 (AAV2 helper plasmid), p5E18-VD2-8 (AAV8 helper plasmid), pladeno5 (adenovirus helper plasmid), and pAAV-ISce I.AO3 (AAV-ISce I.AO3 vector plasmid). This was necessary because human genomic DNA and these plasmid DNAs, which are irrelevant to the rAAV vector genome, often were found incorporated in rAAV vector genomes by illegitimate recombination at the time of rAAV vector production in human embryonic kidney 293 cells, as previously reported (31, 32, 35, 37, 40). As a measure of illegitimate events occurring during the plasmid rescue procedure, we also searched isolated rAAV provirus plasmid DNA against the yeast genome. When precise breakpoints could not be determined due to microhomology at integration sites, we defined the break sites of the rAAV vector and the flanking mouse genome sequence such that the microhomology was included in both sides. Computer-simulated random integration sites (1,000, 10,000, and 30,000) were generated as previously described (40, 59). Random numbers were generated by a rand() function in the Perl program, with the srand() function for seed as previously described (59). To investigate the spectrum of rAAV and random integration sites in the mouse genome, we downloaded coordinates of RefSeq genes, CpG islands, and genomic locations of micro-RNA (miRNA) for the February 2006 mouse genome freeze from the UCSC genome project website and analyzed integration site data as previously described (40, 59).

For palindrome analyses, the flanking 500-bp genomic sequences centromeric and telomeric to each integration site were extracted from the mouse genome database and were searched for the presence of DNA palindromes. The size of the examined window (i.e., 1 kb) took into consideration that rAAV integration can be accompanied by chromosomal sequence deletion and that, in mice, over 80% of these are less than 1 kb (40). Each 1-kb sequence was aligned against itself using the NCBI bl2seq program (version BLAST-2-2-8) with a setting of (-p blastn, -G 5, -E 2, -q -2, -r 1, -e 10.0, -w 11). All the self-complementary sequences identified by the program, the minimum being 12 bp in total length (6 bp in arm length), were collected and again searched with a newer version of the NCBI bl2seq program (version BLAST-2-2-14) using the same settings for confirmation.

In our analysis, palindromes were defined according to the following criteria: (i) inverted repeats are present and spaced ≤5 bp apart; (ii) the arm length is ≥6 bp; (iii) mismatches are minimized such that overall self complementarity with any spacer included is ≥80%; and (iv) no sequence gaps occur in the self-aligned region. If two or more palindromes occurred within the 1-kb window of sequence around the rAAV or random integration site, the longest palindrome (first priority) and closest proximity (second priority) to each integration site was designated the integration-labeled palindrome. When the longest palindromic regions had multiple alternative self alignments, which was particularly the case with long (AT)n-containing palindromes, the palindrome showing the longest alignment with self complementarity of ≥90% was taken, if present, as the integration-labeled palindrome. rAAV orientation was defined according to whether the vector was incorporated in a plus or minus orientation relative to the numbering of the mouse reference sequence. An orientation was randomly assigned to each simulated integration event using a rand() function. A total of 20 independent random integration data sets that included an orientation parameter were generated for the simulated insertion sites.

The coordinates of the center of each palindrome were defined as follows. For palindromes covering an odd number of base pairs (spacer included), the coordinate of the center-most position was used. For palindromes covering an even number of base pairs, where both of two positions are central, the base pair closest to the integration site was taken as the palindrome's central coordinate.

The coordinates of the rAAV integration sites from our previous study (accession numbers EI173586 to EI174306) (40) were updated, and the rAAV provirus-flanking DNA sequences were reanalyzed according to the February 2006 freeze (UCSC mm8 and NCBI Build 36). The palindrome analysis was performed as described above.

For comparison, rAAV integration sites identified in nonselected human cultured cells (32) were retrieved from GenBank. Of 1,172 submitted sequences (accession numbers DU709854 to DU711025), 815 were chosen according to the following criteria: (i) we excluded sequences with plasmid-derived or undefined rAAV flanks; (ii) we excluded sequences with insertions of over 20 bp in length at junctions; (iii) we excluded sequences in which the breakpoint in the rAAV genome was too close to the sequence primer and therefore we could not determine the breakpoints; (iv) we excluded rAAV integrations into the human rRNA gene repeats; and (v) we included only integration sites that mapped to a unique site in the human genome (hg 18) with identity of 95% or more.

For dinucleotide repeat analyses, we determined the lengths and positions of integration-labeled dinucleotide repeat tracts present in the same 1-kb sequence windows (±500 bp from rAAV and random integration sites) as described for the palindrome analyses.

Statistics. The experimentally determined palindrome-labeling and integration frequencies were compared to those expected from 1,000, 10,000, or 30,000 randomly occurring integrations. The biological significance of various parameters of interest was assessed by determining the statistical significance of detected biases using the two-tailed χ^2 test. For cases in which values in a contingency table were five or less, the two-tailed Fisher's exact probability test was used. Twenty-nine integrations landing in the rRNA gene repeats in the present study were excluded from the statistical analyses.

Nucleotide sequence accession numbers. Sequences of the AAV vector plasmid pAAV-ISce I.AO3 and of the rAAV vector genome-host cellular DNA junction have been deposited in GenBank under the accession numbers EU022316 (for pAAV-ISce I.AO3) and ER934559 to ER935499 and ER935831 (for rAAV integration junction sequences).

RESULTS

Experimental approach. AAV-ISce I.AO3 (Fig. 1) is a 3.1-kb rAAV shuttle vector designed for the purpose of isolating rAAV integration sites from nondividing cells by a plasmid rescue technique. The salient features of this shuttle vector relevant to the present study are the following: (i) it carries a 1.7-kb DNA sequence containing a short prokaryotic promoter-driven ampicillin resistance (β-lactamase) gene and plasmid origin of replication (ori) (Amp/Ori cassette) primarily derived from pUC plasmid (35); (ii) it carries an ISceI-BamHI combination site at the 5′ end of the Amp/Ori cassette; and (iii) it has only one BamHI and ISceI site and lacks a BglII site. The unique BamHI site serves in the recovery of rAAV provirus sequences in bacteria by plasmid rescue, whereas the role of the unique ISceI site adjacent to the BamHI site is for physical removal of extrachromosomal and/or concatemeric rAAV genomes from sample DNA. The mouse genome does not contain ISceI endonuclease recognition sites; therefore, an rAAV provirus genome can remain as HMW DNA following ISceI digestion if it has become linked to host chromosomal DNA. Accordingly, rAAV provirus-host chromosomal DNA junction fragments can be separated from other rAAV vector-vector junction fragments by DNA fractionation. We truncated the DNA fragment containing the prokaryotic promoter driving the β-lactamase gene by 74 to 54 bp (nucleotide positions 3896 to 3949 of the Tn3 transposon; accession number V00613). This modification further suppressed the vector-vector recombinant background in the libraries while retaining expression of the β-lactamase gene in *E. coli*. AAV-ISce I.AO3 carries a portion of the Moloney murine leukemia virus (MLV) long terminal repeat sequence (207 nucleotides; nucleotide positions 7575 to 7781 of MLV; accession number NC 001501) incorporated so that an established linear amplification-mediated PCR technique (48) could be used to isolate rAAV inte-

gration sites. However, this feature was not exploited in the present study.

In rAAV-transduced cells, rAAV genomes exist in various double-stranded DNA forms. These include extrachromosomal double-stranded circular monomers, double-stranded circular and linear concatemers, and integrated monomeric and concatemeric genomes, all of which exhibit various rearrangements (35, 36, 38, 40, 41). This multiplicity of structure creates significant complexity in DNA samples from rAAV-transduced cells, where only a small portion of vector genomes actually integrates into host chromosomal DNA (41). In proliferating cells, the extrachromosomal forms can be diluted out by cell division, but substantial dilution does not occur in quiescent cells in animal tissues, adding greatly to the challenge of identifying rAAV integration events in the latter context. To isolate many integration sites in nondividing cells, we developed a new methodology, for which the important features include the following. (i) Abundant extrachromosomal circular monomer genomes and concatemeric genomes are removed by ISceI digestion and size fractionation. (ii) Circular ligation products are prepared free of recombinogenic linear DNA before transformation of bacteria. (iii) Artifactual recombination events between rAAV vector genomes and mouse chromosomal DNA generated either in vitro or in bacteria is monitored by addition of yeast genomic DNA. The yeast genomic DNA mixed with sample DNA serves as a tag sequence of unwanted recombination, which we can measure as the proportion of rescued plasmids bearing rAAV genomes flanked with yeast genomic DNA.

Representativeness of the rAAV integration site data set in our study. To minimize technical biases, the choice of the appropriate restriction enzyme(s) employed in plasmid rescue was critical. To construct libraries best representing vector integration sites, digestion of sample DNA with a restriction enzyme(s) must generate DNA fragments in appropriate sizes so that a vast majority of digested DNA fragments are small enough to transform bacteria efficiently.

In the present study, we opted to use BamHI and BglII double digestion, in which BamHI cuts the vector genome once. To verify that the choice of BamHI and BglII double digestion was appropriate, in silico restriction enzyme digestion of the mouse genome was performed. A histogram of all 1,395,386 BamHI-BglII restricted fragments revealed that 90 and 99% of the fragments are ≤4.5 and ≤9.5 kb in length, respectively (Fig. 2). To perform a similar analysis of BamHI-BglII-restricted fragments flanking random integration sites, we multiplied the size of each of 1,395,386 BamHI-BglII-restricted fragments with a random number between 0 and 1. This analysis revealed that 90 and 99% of random integration-flanking BamHI-BglII-restricted fragments should be ≤2.5 and ≤6.0 kb in length, respectively (Fig. 2). Because the maximum predicted size of rAAV vector genomes contained in rAAV integration junction plasmids is, in theory, 2.2 kb, 90 and 99% of rAAV provirus plasmid sizes should be ≤4.7 and ≤8.2 kb, respectively.

The relationship between the size of rAAV integration junction plasmids and bacterial transformation efficiency was investigated by selecting eight AAV-ISce I.AO3 vector integration junction plasmids representing sizes that ranged between 2.2 and 10.1 kb, mixing them at an equal molar ratio, and
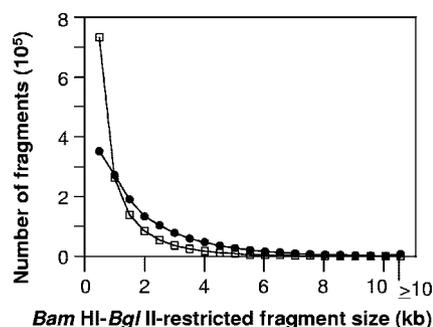

FIG. 2. Histogram of in silico BamHI-BglII-digested mouse genomic DNA fragments. In silico-digested DNA fragments were grouped based upon size in 500-bp increments. ●, BamHI-BglII-digested mouse genomic DNA; □, BamHI-BglII-digested mouse genomic DNA flanked with computer-simulated random integrants (see the text).

transforming DH10B *E. coli*. The plasmids were obtained from our collection of rAAV integration junction plasmids. Two hundred seven colonies were randomly analyzed among the resulting transformants. The 10.1-, 8.1-, 7.0-, 4.8-, 3.8-, 3.3-, 2.5-, and 2.2-kb plasmids were present in the following percentages, respectively: 1.4, 11.6, 9.7, 18.4, 21.2, 12.1, 16.9, and 8.7%. This indicated that plasmids of 8.1 kb in size or smaller could be efficiently retrieved in bacteria, but that a 10.1-kb plasmid had a much lower transforming potential. This test established that virtually all the rAAV integration junction plasmids fell within a size range that was able to transform bacteria efficiently. Therefore, the data set we have generated reflected a representative collection of rAAV integration events.

The reliability of our plasmid rescue strategy was assessed by monitoring illegitimate intermolecular recombination between rAAV vector genomes and yeast genomic DNA. Among 1,269 plasmid clones obtained from our libraries, no such recombination events were observed, demonstrating the high fidelity of the method.

A survey of rAAV integration sites in quiescent somatic cells from mouse liver, skeletal muscle, and heart. rAAV2 and rAAV8 vectors with AAV-ISce I.AO3 genomes were introduced into C57BL/6J and DNA-PKcs-deficient C57BL/6J SCID male mice by injection via the tail vein or the portal vein. The purpose of studying both wild-type and DNA-PKcs-deficient mice was to investigate whether DNA-PKcs activity influences rAAV integration patterns in vivo. A recent study by Song et al. suggested that DNA-PKcs inhibits rAAV integration in cultured cells and in mouse liver (51). Serotypes 2 and 8 were chosen because they primarily transduce hepatocytes, striated myofibers, and cardiomyocytes in the liver, skeletal muscle, and heart, respectively (10, 34). These cell types are presumed to be mostly quiescent in adult animals. A dose of 7.2×10^{12} or 5.0×10^{10} vector genomes (vg) of AAV8-ISce I.AO3 per mouse was infused via the tail vein, and 3.0×10^{11} vg of AAV2-ISce I.AO3 per mouse was infused via the portal vein (Table 1). According to our previous studies, a tail vein injection of 5.0×10^{10} vg/mouse of rAAV8 vector or a portal vein injection of 3.0×10^{11} vg/mouse of rAAV2 vector transduces 5 to 10% of hepatocytes in the liver (10, 34, 39). A tail

TABLE 1. Summary of results from high-throughput rAAV integration site analysis

Mouse	Vector	Vector dose (10^{10})	Route[b]	Tissue	No. of clones in each category				Efficiency (%)[a]
					Total sequenced	Mouse hit	rRNA gene repeat	Mapped to a unique site	
B6	AAV8	720	TV	Liver	151	25	1	24	16.6
B6	AAV8	5	TV	Liver	230	218	11	201	94.8
B6	AAV2	30	PV	Liver	84	80	2	76	95.2
B6	AAV8	720	TV	Muscle	21	17	1	16	81.0
B6	AAV8	720	TV	Heart	93	86	1	83	92.5
B6 total					579	426	16	400	73.6
SCID	AAV8	720	TV	Liver	229	166	5	160	72.5
SCID	AAV8	5	TV	Liver	74	65	0	63	87.8
SCID	AAV2	30	PV	Liver	86	82	2	79	95.3
SCID	AAV8	720	TV	Muscle	189	161	5	148	85.2
SCID	AAV8	720	TV	Heart	112	97	1	91	86.6
SCID total					690	571	13	541	82.8
Total					1,269	997	29	941	78.6

[a] Efficiency is derived from the following calculation: (number of clones containing a sequence from the mouse genome)/(number of total clones sequenced) × 100.
[b] TV, tail vein; PV, portal vein.

vein injection of 7.2×10^{12} vg/mouse of rAAV8 vector can transduce virtually all the hepatocytes, cardiomyocytes, and striated myofibers (10, 34, 39). Liver, skeletal muscle, and heart tissues were harvested 6 weeks postinjection to identify rAAV integration sites in each of these tissues.

In total, rAAV-flanking DNA sequences were determined for 1,269 rAAV integration plasmid clones from the various tissue libraries (Table 1). Of these, 997 clones were independent rAAV-mouse genomic DNA integration junctions. The remaining clones contained junctions of rAAV-human genomic sequences, rAAV-helper plasmid junctions, or rAAV-rAAV junctions as previously observed (31, 32, 35, 36, 40). The efficiency of isolation of rAAV vector integration sites was calculated as the number of plasmid clones that carried rAAV vector integration sites divided by the total number of clones for which the junction sequences were determined. This efficiency consistently fell within 73 to 95%, except for a single sample from C57BL/6J liver transduced with 7.2×10^{12} vg/mouse (17% efficiency) (Table 1). This particular mouse sample contained many rearranged vector genomes originating from concatemers with high complexity, which presumably resulted in the disruption of the ISceI site, and therefore could not be quantitatively removed from mouse genomic DNA by physical separation (data not shown). The breakpoints of rAAV genomes were distributed in the inner half of the AAV ITR and its flanking internal vector sequences, as we have observed previously (32, 40).

rAAV integrates preferentially in genes and near gene regulatory sequences in hepatic and nonhepatic mouse tissues. We investigated rAAV integration site preference in mouse liver, skeletal muscle, and heart. Consistent with previous reports (32, 40), the rRNA gene repeat is a preferred site for rAAV integration (29/997 = 2.9%, compared to a predicted 0.3% frequency from a random integration model) (40) (Table 1). Excluding these rRNA gene integrations, a total of 941 rAAV integration sites mapped to unique sites in the mouse genome. We compared rAAV integration sites to 10,000 computer-simulated random integrations (Table 2). The results

were consistent with our previous observation, i.e., preferential integration into RefSeq genes, near transcription start sites, and into or near CpG islands (32, 40). However, the bias for transcription start sites and CpG islands in the present study (in which integration sites were identified in quiescent cells) was less pronounced than in our previous study (in which integration sites were isolated after in vivo selection of integrants) (40) (Table 2). Apart from this difference, the trend for integration into RefSeq genes was quite comparable between the two studies. We also investigated rAAV integration in or near miRNA-coding regions. None and 2 of 941 rAAV integrations occurred within the region ±1 and ±5 kb from genomic locations of miRNA, respectively.

DNA palindromes were found at or near rAAV integration sites at unusually high frequencies. DNA palindromes are a major component of the genomic sequences that potentially form unstable non-B DNA structures (reviewed in references 18, 23, and 57). Therefore, we hypothesized that DNA palindromes might be favored sites for rAAV integration in host chromosomal DNA. To test this hypothesis, we identified DNA palindromes with an arm length of ≥6 bp with a Perl script based on NCBI's bl2seq program as described in Materials and Methods. For the purposes of the study, a palindrome was defined as an inverted repeat sequence with an arm length of ≥6 bp, a spacer of ≤5 bp, and overall self complementarity of ≥80%.

We compared the frequency of integration in the vicinity of DNA palindromes for the working data set of 941 rAAV integrations to those for 1,000 and 30,000 computer-simulated random integrations. A DNA palindrome found within 500 bp of either side of an integration site that fulfilled our criteria was defined as having been integration labeled. We set a 1-kb window for the analysis, because rAAV integrations are frequently accompanied by chromosomal sequence deletions (40). A total of 369 rAAV integrations labeled a DNA palindrome with an arm length of ≥6 bp. Of the 1,000 computer-simulated integrations, 318 random integrations labeled a palindrome with an arm length of ≥6 bp. Of the 30,000 computer-

TABLE 2. Locations of unselected rAAV integration sites, computer-simulated random integration sites, in-vivo-selected rAAV integration sites, and break-prone palindromes in various mouse tissues[d]

Mouse type and site or palindrome examined	Vector	Vector dose (10^10)	Route	Tissue	rAAV vector genome load (ds-vg/dge) in cells[a]	No. of sites analyzed	Frequency[b] (%)				
							RefSeq gene	Tx ± 1 kb	Tx ± 5 kb	CpG ± 1 kb	CpG ± 5 kb
rAAV integration sites, no selection											
B6	AAV8	720	TV	Liver	2,291 ± 548	24	41.7	8.3	20.8	8.3	8.3
B6	AAV8	5	TV	Liver	11 ± 4	201	*51.2*	5.0	*16.4*	5.5	*14.9*
B6	AAV2	30	PV	Liver	4.7 ± 2.1	76	*63.2*	2.6	*13.2*	6.6	*14.5*
B6	AAV8	720	TV	Muscle	6.1 ± 2.1	16	43.8	0.0	0.0	0.0	18.8
B6	AAV8	720	TV	Heart	25 ± 5	83	*50.6*	3.6	7.2	3.6	8.4
B6 total						400	*52.5*	4.3	*13.5*	5.3	*13.3*
SCID	AAV8	720	TV	Liver	2,223 ± 101	160	*48.1*	0.6	*11.3*	2.5	*11.9*
SCID	AAV8	5	TV	Liver	8.0 ± 3.5	63	*61.9*	4.8	*12.7*	9.5	*14.3*
SCID	AAV2	30	PV	Liver	7.4 ± 3.6	79	*51.9*	10.1	*22.8*	8.9	*19.0*
SCID	AAV8	720	TV	Muscle	9.9 ± 0.1	148	*55.4*	2.7	*16.2*	3.4	*13.5*
SCID	AAV8	720	TV	Heart	28 ± 10	91	*53.8*	12.1	*18.7*	7.7	*12.1*
SCID total						541	*53.2*	5.0	*15.7*	5.4	*13.7*
Combined total, no selection						941	*52.9*	4.7	*14.8*	5.3	*13.5*
Random simulation						10,000	26.8	1.3	6.6	1.5	5.8
rAAV integration sites, in vivo selection[c]											
HTI	AAV2	30	PV	Liver		283	*54.6*	*25.0*	*43.7*	*34.5*	*46.8*
Palindromes (arm length ≥20 bp)											
Random simulation						399	23.1	0.3	4.3	0.5	5.5
rAAV labeled (break prone)						134	*53.7*	4.5	*12.7*	3.0	*11.2*

[a] Vector genome loads in cells are expressed as means ± standard deviations ($n = 3$ to 5 for the mice injected with a low vector dose) or means ± |mean-each value| ($n = 2$ for the mice injected with a high vector dose). The vertical bars indicate the absolute value of a number.
[b] Percentages in italics are significantly higher than those from the random integration model ($P < 0.05$ by χ^2 test or Fisher's exact test).
[c] Only the results from analysis of the right (3′) side of rAAV-cellular DNA junctions are presented.
[d] Abbreviations: ds-vg/dge, double-stranded vector genome copy numbers per diploid genomic equivalent; Tx, transcription start site; B6, C57BL/6J mice; SCID, DNA-PKcs-deficient C57BL/6J SCID mice; HTI, hereditary tyrosinemia type I mice; TV, tail vein injection; PV, portal vein injection.

simulated integrations, 1,327 labeled a palindrome with an arm length of ≥11 bp. We categorized DNA palindromes based on their arm length and compared the palindrome-labeling frequency in each category between the experimental and simulated groups. This analysis unambiguously revealed that 1-kb regions containing DNA palindromes are indeed hot spots for rAAV vector integration in the liver, skeletal muscle, and heart (Fig. 3). For example, in SCID mouse heart, 30% of all rAAV integrations were in the vicinity of DNA palindromes with an arm length of ≥20 bp, whereas a 1.3% frequency was observed in the random simulation (Fig. 3J). The observation holds for both wild-type mice and DNA-PKcs-deficient SCID mice (Fig. 3C, D, E, F, H, I, and J). The effect was diminished in a sample for which a high-vector-genome load in cells had been measured at the time tissue was harvested (Fig. 3B and Table 2). Notably, preferential labeling of palindromes was not observed in our previous study (36, 40) (Fig. 3K and L).

DNA palindromes with an arm length of ≥20 bp (total length, ≥40 bp) have a biological impact on rAAV integration. The 369 rAAV-labeled DNA palindromes with an arm length range between 6 and 72 bp and 1,645 random integration-labeled palindromes with an arm length range between 6 and 86 bp were compared to determine the minimum length of palindromes with biological significance. In the analysis, rAAV- and simulation-labeled palindromes were sorted according to arm length, and each set was displayed in plots showing the labeling frequency as a function of arm length.

The difference became significant at a palindrome arm length of 19 to 20 bp or more (Fig. 3M). Thus, we conclude that naturally occurring palindromes as short as ~40 bp (arm length of ~20 bp) have an impact in vivo, attracting rAAV genomes for their integration. This is a significant finding, in that it demonstrates that DNA palindromes much shorter than well-characterized unstable AT-rich palindromes with an arm length of 150 to 300 bp (18) can have some impact in vivo.

DNA palindromes with an arm length of ≥20 bp are in fact the primary target for integration. The aforementioned observation that rAAV integration occurred preferentially near DNA palindromes per se does not necessarily indicate that rAAV integrated into DNA palindromes themselves. It might be possible that DNA palindromes merely attracted rAAV vector genomes by some mechanism and allowed them to integrate in their vicinity but not necessarily into DNA palindromes. To investigate which is likely the case, we analyzed the positional relationship between rAAV integration sites and the nearby labeled palindromes. For this, the location of each of the 369 rAAV-labeled palindromes was scored within the defined 1-kb region extending ±500 bp of each integration site. The histogram collating the palindrome locations for the rAAV integration site collection showed a unimodal distribution pattern with a peak corresponding to the rAAV integration site (Fig. 4A). Importantly, this distribution pattern is primarily attributed to that of 134 DNA palindromes with arm lengths of ≥20 bp (Fig. 4B) and not to 235 of those with arm
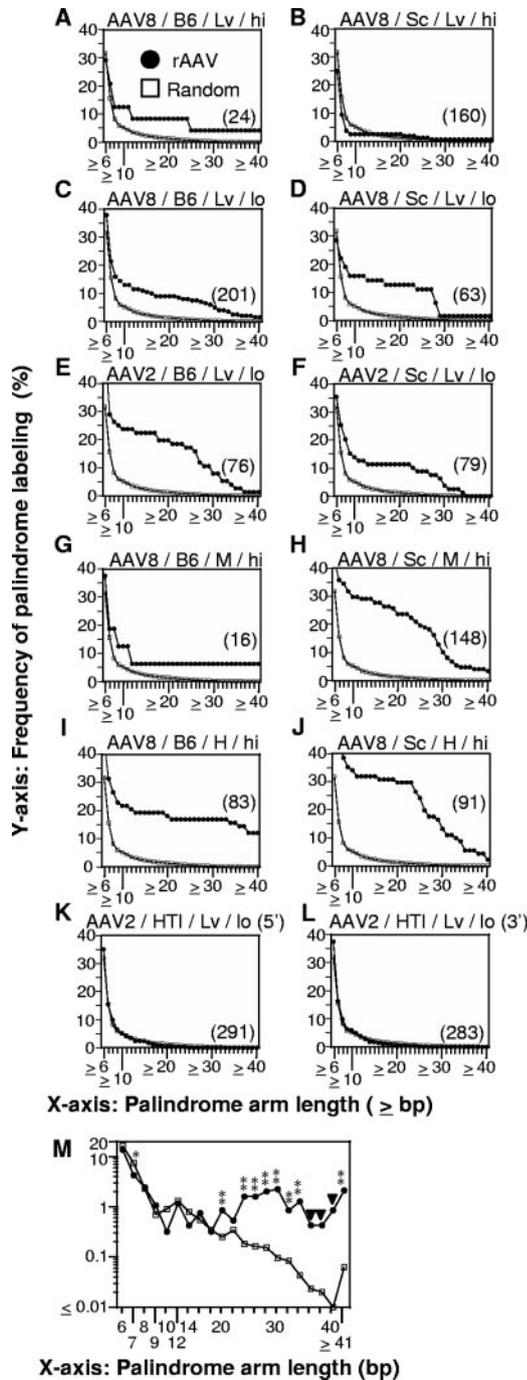
FIG. 3. Frequency of palindrome labeling by rAAV integration compared to that of labeling by random integration. Palindromes at or near rAAV and random integration sites are categorized based on their arm length. (A to L) Frequency of palindrome labeling as a percentage of total integrations is plotted for each size category. Experimental variables are given above each graph: rAAV vector serotype (AAV2 or AAV8)/mouse strain (B6, C57BL/6J; Sc, SCID; HTI, hereditary tyrosinemia type I mouse)/tissue type (Lv, liver; M, skeletal muscle; H, heart)/vector dose (hi, high dose; lo, low dose; see Tables 1 and 2). The number in parentheses in each panel indicates the total number of rAAV integration sites analyzed in each group. For panels A to J, 3′ rAAV integration sites were isolated without selection, while integration sites were isolated after in vivo selection for panels K and L, in which 5′ and 3′ rAAV integration sites are separately displayed. (M) Palindrome labeling frequency data from panels A to J are com-
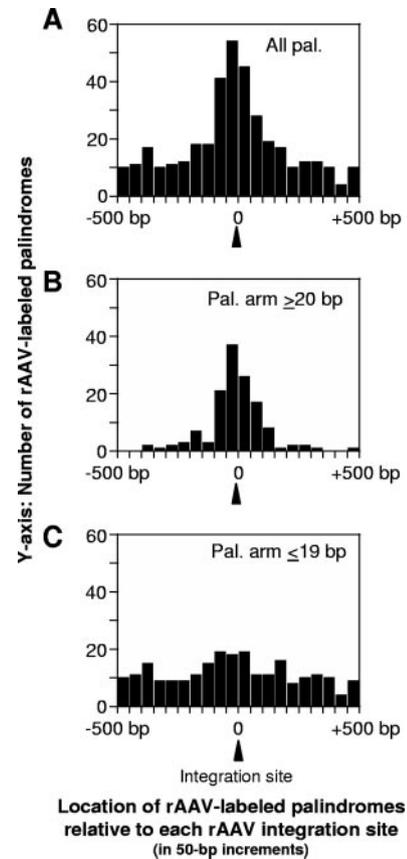
FIG. 4. Positional relationship between rAAV integration sites and DNA palindromes (pal.). The positions of rAAV-labeled DNA palindromes are displayed relative to their associated rAAV integration sites. A 1-kb sequence window represents ±500 bp centromeric (minus) and telomeric (plus) to each rAAV integration site (centered at the 0-bp position). The histogram gives the number of palindromes located within each 50-bp increment relative to the rAAV integration sites (i.e., positions −500 to −451, −450 to −401, and so on). An exception is the 51-bp center window from positions −50 to 0, which includes an rAAV integration site. (A) rAAV-labeled DNA palindromes with an arm length of ≥6 bp (369 palindromes in all). (B) rAAV-labeled DNA palindromes with an arm length of ≥20 bp (134 in all). (C) rAAV-labeled DNA palindromes with an arm length between 6 and 19 bp (235 in all).

lengths of ≤19 bp (Fig. 4C). Twenty-six of 134 rAAV-labeled palindromes with arm lengths of ≥20 bp had rAAV integration within the palindromes. No specific pattern was observed for 399 similarly analyzed randomly labeled palindromes with arm lengths of ≥20 bp (data not shown). This observation strongly indicates that DNA palindromes themselves were in fact the

bined and plotted as a function of palindrome arm length. For this analysis, labeling frequency of palindromes with the same arm length is compared between rAAV and random integrations. For palindromes with an arm length of ≥11 bp, points represent palindrome arm lengths in increments of 2 bp (i.e., 11 to 12 bp, 13 to 14 bp, and so on). Asterisks and solid triangles indicate statistical significance compared to results for random integrations ($P < 0.001$ [two asterisks] and $0.001 \leq P < 0.01$ [one asterisk] by χ^2 test; $P < 0.001$ [closed triangle] by Fisher's exact test).
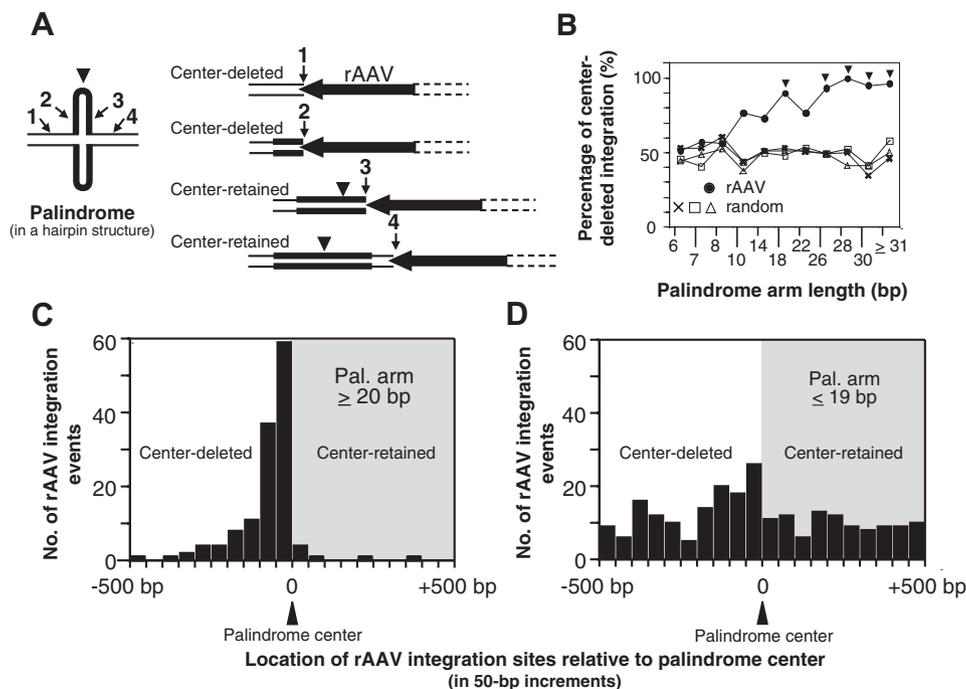
FIG. 5. Patterns of palindrome (Pal.) center retention/deletion associated with rAAV integration. (A) Definition of center-deleted and center-retained rAAV integration. (B) Percentage of center-deleted rAAV integrations of total rAAV integrations is plotted as a function of palindrome arm length and compared to that for random integration. To increase the power of analysis, palindromes with different arm lengths are combined in the following manner: 9 to 10 bp (shown as 10 in the figure), 11 to 14 bp (14), 15 to 18 bp (18), 19 to 22 bp (22), 23 to 26 bp (26), 27 to 28 bp (28), 29 to 30 bp (30), and 31 bp or more ($\geq$31). We compared the observed frequency to that of 20 independently generated random integration data sets (see Materials and Methods) by the χ^2 test or Fisher's exact test. Solid triangles indicate statistical significance ($P < 0.05$; Fisher's exact test) for all 20 random data set comparisons. Three representative random integration data sets are included. (C and D) Positional relationship between center-deleted or center-retained rAAV integrations and the palindrome symmetry center. rAAV integration sites are shown as histograms, giving their locations in 50-bp increments across a 1-kb region centered on the palindrome symmetry center. An exception is the 51-bp center window from positions -50 to 0, which includes the palindrome symmetry center. Center-deleted or center-retained rAAV integrations are displayed in the left (clear background) or right (gray background) portion of the 1-kb window as labeled. (C) rAAV-labeled DNA palindromes with an arm length of $\geq$20 bp (134 palindromes in all). (D) rAAV-labeled DNA palindromes with an arm length between 6 and 19 bp (235 in all).

primary target for rAAV integration. rAAV-cellular DNA junctions found near but outside the DNA palindromes most likely resulted from deletions of host chromosomal DNA associated with integrations (40) that would be within the DNA palindrome if no cellular DNA sequence deletions had occurred. In addition, the contrasting positional distribution patterns between palindromes with arm lengths of $\geq$20 and $\leq$19 bp give an independent confirmation, along with the statistical evidence shown in Fig. 3M, that a biological impact becomes apparent only once palindromes have a length of roughly $\geq$40 bp.

To further investigate the relationship between rAAV integration sites and palindromes, a reciprocal analysis was performed in which rAAV integration sites were mapped within a 1-kb window around rAAV-labeled palindromes ($\pm$500 bp from rAAV-labeled palindromes). Here, we discriminated between center-deleted and center-retained integration events as defined in Fig. 5A. In center-deleted integration, the genomic DNA flanking the rAAV insertion did not contain the palindrome symmetry center, whereas in center-retained integration the symmetry center was retained. The significance of this is that in cases of center-retained integration, one can be certain that the palindrome symmetry center was not disrupted by

rAAV integration, whereas for center-deleted integration, symmetry center disruption cannot be definitely determined. A statistical analysis of 369 rAAV-labeled and 1,645 random integration-labeled palindromes revealed a significant center-deleted rAAV integration bias when the palindrome arm length was 15 to 18 bp or more (Fig. 5B). This again provides additional evidence that palindromes with an arm length of $\geq$20 bp have special attributes. By mapping locations of the 134 sites of rAAV integration events that labeled DNA palindromes with an arm length of $\geq$20 bp within the 1-kb window, it became clear that rAAV integrations were almost exclusively center-deleted events (Fig. 5C). Even more notable was the strong enhancement of integrations for the positions closest to (and before) the palindrome symmetry center (Fig. 5C). Neither trend was observed when 235 rAAV-labeled palindromes with an arm length of $\leq$19 bp were displayed (Fig. 5D). One may argue that palindromic sequences were deleted in bacteria, resulting in an artifactual bias toward center-deleted integration. Although we cannot totally exclude this possibility in all the integration events presented here, a majority of palindrome-labeling integration events should correctly present the center-deleted or center-retained events based on the observations presented in the last section of Results.
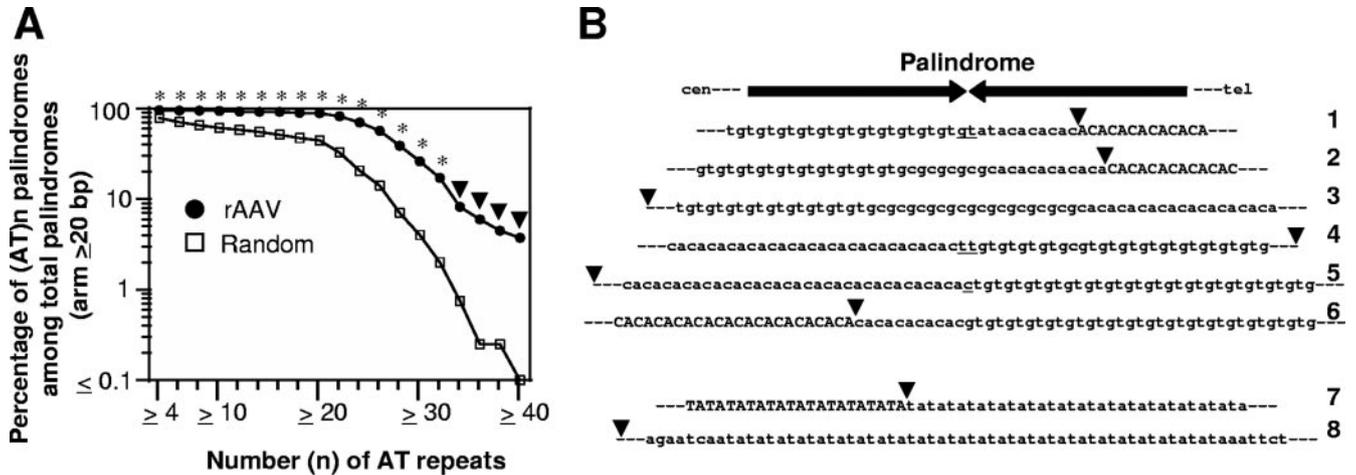
FIG. 6. Significance of (AT)n and (AC)n (GT)n palindromes in rAAV-labeled palindromes susceptible to breakage. (A) Percentage of rAAV-labeled (AT)n palindromes among the 134 rAAV-labeled palindromes with an arm length of ≥20 bp is plotted as a function of n. Random integration-labeled palindromes are likewise plotted. Asterisks and solid triangles indicate statistical significance between rAAV and random integrations (∗, $P < 0.001$ by χ^2 test; closed triangle, $P < 0.001$ by Fisher's exact test). (B) DNA sequences of six (AC)n (GT)n palindromes (numbered 1 to 6) labeled by rAAV integration. Black arrowheads indicate the positions at which joining to the rAAV vector genome occurred. Uppercase letters denote the portion of the palindrome sequence that was retained in each junction, while lowercase letters denote the nonretained portion. Underlined nucleotides are central spacer sequences. All six (AC)n (GT)n palindromes show a center-deleted pattern. Two representative (AT)n palindromes are shown as 7 and 8 for comparison. Coordinates of the palindromes numbered 1 to 8 (UCSC mm8) are chr 10: 38176562, chr 5: 118212140, chr 6: 7083048, chr 15: 29735040, chr 8: 120556025, chr 16: 69258562, chr X: 150903099, and chr 15: 3434443, respectively.

Diverse palindromic sequences are rAAV integration targets. The palindromes that were labeled by rAAV in a nonrandom fashion were predominantly those containing a long stretch of AT (TA) dinucleotide repeats [i.e., (AT)n palindromes]. To investigate whether (AT)n palindromes are a biologically significant type of DNA palindrome susceptible to breakage, we compared the degrees of prevalence of (AT)n palindromes between 134 rAAV-labeled and 399 random integration-labeled DNA palindromes with arm lengths of ≥20 bp. The analysis revealed that rAAV-labeled palindromes have a significantly higher proportion of (AT)n palindromes than random integration-labeled palindromes (Fig. 6A). This demonstrates that (AT)n palindromes are a significant type of DNA palindrome that constitute the targets for rAAV vector integration.

However, DNA palindromes labeled by rAAV in a nonrandom fashion were not exclusively (AT)n palindromes. Six of 134 rAAV-labeled palindromes with an arm length of ≥20 bp were comprised of (AC)n on one arm and (GT)n on the other [i.e., (AC)n (GT)n palindromes] (Fig. 6B). Statistical analysis revealed that the rAAV labeling frequency of this type of non-AT-rich palindrome with an arm length of ≥20 bp (6 of 941) was significantly higher than the random labeling frequency (27/30,000) ($P < 0.0001$; χ^2 test). In addition, all of the (AC)n (GT)n palindromes were labeled in a center-deleted pattern (Fig. 6B). Thus, not only AT-rich palindromes but also other types of palindromes constitute targets for rAAV vector integration.

Nonpalindromic dinucleotides are not rAAV integration targets. An important question to address in the present case, given that rAAV-labeled palindromes were predominantly (AT)n palindromes (Fig. 6A), is whether it is a palindrome or simple dinucleotide repeat nature that is operative in creating a platform for rAAV integration. To address this, we took all possible dinucleotide repeats [i.e., (AT)n or (TA)n, (CG)n or (GC)n, (AC)n or (GT)n, and (AG)n or (CT)n] and measured their labeling frequency in the experimental or randomly assigned data sets. Here, labeling frequency is the frequency of rAAV integration events at or near dinucleotide repeat tracts of interest (i.e., whether or not at least one dinucleotide repeat tract is found within 500 bp of each integration site). The rAAV labeling frequency was significantly higher than that of random labeling only for (AT)n (Fig. 7A). This was not the case for nonpalindromic dinucleotide repeats (AC or GT)n or
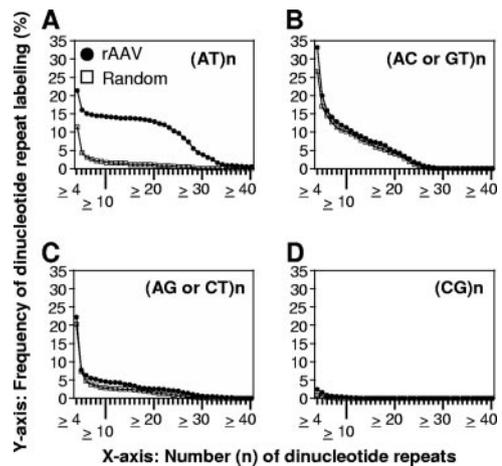
FIG. 7. Analyses of dinucleotide labeling by rAAV integration. Relative frequency of rAAV-labeling of the four possible types of dinucleotide repeats among total integration events is plotted as a function of n (n is the number of dinucleotide repeats). For comparison, the frequency of random integration labeling is also plotted. The four types of dinucleotide repeats are (AT)n (A), (AC or GT)n (B), (AG or CT)n (C), and (CG)n (D).
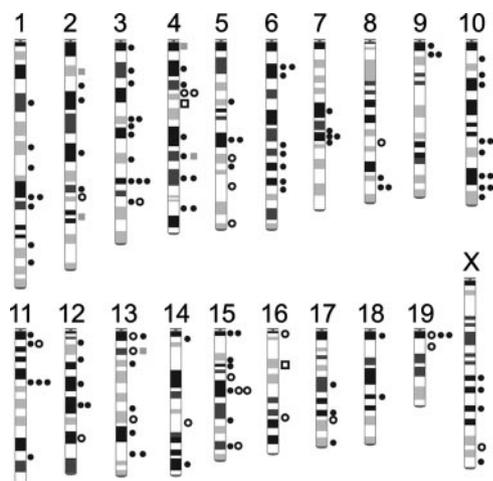
FIG. 8. Distribution of break-prone palindromes in the mouse genome. A total of 134 rAAV-labeled palindromes with an arm length of ≥20 bp are mapped on a normal mouse karyotype. Symbols indicate center-deleted integration in (white circles) or near (black circles) palindromes and center-retained integration in (white squares) or near (gray squares) palindromes, respectively. There was no rAAV-labeled break-prone palindrome on the Y chromosome in the present study.

(AG or CT)n (Fig. 7B and C). (CG)n was rarely found in experimental and simulation data sets. It should be noted that when we counted all (AC or GT)n ($n \geq 6$) repeats in the integration-labeled 1-kb sequence windows, we observed a significant difference in the prevalence of (AC or GT)n ($n \geq 6$) between 134 rAAV integrations labeling palindromes with an arm length of ≥20 bp and 1,000 random integrations. We

found 71 (AC or GT)n repeats ($n \geq 6$) in the 134 1-kb sequence windows containing rAAV-labeled palindromes. This frequency (71/134) was threefold higher than the number expected from random integration (181/1,000), although the (AC or GT)n labeling frequency was not significantly different between the experimental and simulation groups (Fig. 7B). This observation was in accord with our finding that 18 of the 71 (AC or TG)n repeats ($n \geq 6$) were derived from (AC)n (GT)n palindromes or (AC or GT)n dinucleotide repeats occasionally commingled as inverted repeats in the left and right arms of (AT)n palindromes. Thus, only palindromic (AT)n and paired (AC)n and (GT)n are the significant type of dinucleotide repeats that create a platform for preferential rAAV integration.

Only a subset of DNA palindromes with an arm length of ≥20 bp in the mouse genome form a platform for rAAV integration. The 134 rAAV-labeled DNA palindromes with arm lengths of ≥20 bp were found throughout the genome (Fig. 8), with approximately half of them residing in genic regions (Table 2). It was of interest to address whether all or only a portion of DNA palindromes with an arm length of ≥20 bp were susceptible to breakage, serving as a platform for rAAV integration. To investigate this, we took advantage of the observation that certain DNA palindromes were recurrently labeled by rAAV, and such palindromes were predominantly (AT)n ($n \geq 20$) palindromes (Table 3). In the mouse genome database, we identified 10,659 (AT)n ($n \geq 20$) palindromes. There were 119 (AT)n ($n \geq 20$) palindromes labeled by rAAV, and of these, 9 palindromes were labeled twice by independent integration events (Table 3). According to a random model, the probability that a palindrome has n times labeling in the sample size of 119 follows the equation $P(n) = {}_{119}C_n(1/$

TABLE 3. Summary of palindromes recurrently labeled by rAAV integration

Features of palindromes						Features of rAAV integration breakpoints					
Chromosome no.	Band	Coordinate	Arm length (bp)	Spacer (bp)	Mismatch in the arm (bp)	AT/TA repeat no.	Sample	Mouse genome (coordinate)	rAAV nucleotide position[a]	Insertion size (bp)[b]	Orientation of rAAV[c]	In or near palindrome[d]
1	E2.3	117802950	34	0	0	34	AAV8/Sc/H/hi	117802899	2951	0	Plus	Near
							AAV8/Sc/H/hi	117803000	2891	0	Minus	Near
3	H1	139340183	45	0	1	27	AAV8/B6/H/hi	139340211	3011	0	Minus	In
							AAV8/B6/H/hi	139340237	3024	0	Minus	Near
4	B1	47412002	72	0	3	20	AAV8/B6/H/hi	47412045	3034	0	Minus	In
							AAV8/B6/H/hi	47412059	3034	0	Minus	In
4	C6	97561203	8	0	0	2	AAV8/Sc/H/hi	97561059	2939	0	Plus	In
							AAV8/Sc/H/hi	97561027	2970	0	Plus	In
10	D1	98632057	28	0	0	28	AAV8/B6/Lv/lo	98632020	2999	0	Plus	Near
							AAV8/Sc/Lv/lo	98632105	2965	0	Minus	Near
11	B1.1	44726916	33	0	0	32	AAV8/Sc/H/hi	44726958	2947	0	Minus	Near
							AAV8/Sc/M/hi	44726988	2971	0	Minus	Near
14[e]	A2	14983628	25	0	0	25	AAV8/Sc/M/hi	14983597	2654	0	Minus	Near
							AAV8/B6/Lv/lo	14983597	2654	0	Minus	Near
15	D1	63699599	47	0	4	32	AAV8/B6/H/hi	63699609	3015	2	Minus	In
							AAV8/B6/M/hi	63699610	3020	1	Minus	In
15	F1	99702254	47	0	0	46	AAV8/B6/H/hi	99702203	2968	5	Plus	Near
							AAV8/B6/H/hi	99702257	2989	0	Minus	In
19	A	11606158	28	0	0	27	AAV8/Sc/M/hi	11606437	2831	0	Minus	Near
							AAV8/Sc/M/hi	11606302	2842	0	Minus	Near

[a] The total length of the rAAV vector is 3,069 bases. The nucleotide positions are numbered from 1 to 3069 from the left (5′) to the right (3′).
[b] Insertion indicates the length of nucleotide insertions at rAAV-cellular DNA junctions.
[c] Plus and minus indicate that rAAV vectors at the junctions were found in a plus orientation (centromere:5′ side of rAAV:3′ side of rAAV:telomere) and in a minus orientation (centromere:3′ side of rAAV:5′ side of rAAV:telomere), respectively.
[d] In and near indicate that rAAV integration occurred in or near palindromes, respectively.
[e] Independence of the two integration events was confirmed by the different structures of the isolated rAAV integration plasmids. Plasmid AAV8/Sc/M/hi contained a longer vector genome-flanking mouse genomic sequence than that in plasmid AAV8/B6/Lv/lo. A BglII site in the mouse genome was retained in plasmid AAV8/Sc/M/hi, while it was digested and ligated with the BamHI vector end in plasmid AAV8/B6/Lv/lo. Although such an event caused by incomplete digestion with restriction enzymes was rarely found, this finding has provided strong evidence that they were independent integration events.

TABLE 4. Frequency of rAAV labeling of DNA palindromes with an arm length of ≥ 20 bp in the presence or absence of DNA-PKcs activity[a]

| Mouse group by AAV serotype, tissue type, and vector dose | No. of rAAV labeling events in each category | | | χ^2 Value | P value $(d = 1)$ |
	Palindromes (arm length, ≥ 20 bp)	Other sites	Total		
AAV8/Lv/lo					
B6	18	183	201	0.76	0.39
SCID	8	55	63		
AAV2/Lv/lo					
B6	15	61	76	2.06	0.15
SCID	9	70	79		
AAV8/H/hi					
B6	15	68	83	3.19	0.07
SCID	27	64	91		

[a] Abbreviations: Lv, liver; H, heart; B6, DNA-PKcs-proficient C57BL/6J mice; SCID, DNA-PKcs-deficient C57BL/6J mice; hi, high vector dose; lo, low vector dose; d, degree of freedom.

$10,659)_n(10,658/10,659)^{119-n}$ [where P is probability and $_nC_r = n!/(n - r)!r!$, where $_nC_r$ is the combination symbol and ! is factorial]; i.e., $P(0) = 9.89 \times 10^{-1}$, $P(1) = 1.10 \times 10^{-2}$, and $P(2) = 6.11 \times 10^{-5}$. Recurrent rAAV labeling of the same palindrome in this sample size is not accommodated in a random model ($P < 0.0001$; χ^2 test). Thus, DNA palindromes susceptible to breakage constitute a special subset of all (AT)n ($n \geq 20$) palindromes.

DNA-PKcs is not required for preferential rAAV integration at DNA palindromes. Approximately half of the rAAV integration sites in the analysis were collected from DNA-PKcs-deficient SCID mice. Interestingly, as shown in Table 2, there was no significant difference between the spectrum of rAAV integration sites of wild-type and DNA-PKcs-deficient SCID mice. The identity of the cellular factor that resolves DNA palindrome secondary structures, presumably cruciform or hairpin structures, is not yet established. Possible candidates are Mre11 (44, 56) and Artemis (27), both of which are DNA repair enzymes with hairpin-nicking activity. DNA-PKcs is a kinase that activates Artemis endonuclease activity by phosphorylation and/or direct association with autophosphorylated DNA-PKcs (8, 27). Consequently, one would anticipate that a preference for rAAV integration at palindromes ought to be mitigated in DNA-PKcs-deficient SCID mice if this same DNA-PKcs/Artemis pathway is operative. To investigate this, we compared the frequency with which DNA palindromes with an arm length of ≥ 20 bp were labeled in wild-type mice to that in SCID mice. We compared tissues for which over 60 integration sites were identified in both the wild-type mouse and SCID mouse groups, and no significant difference in the frequency of palindrome labeling was found between the two mouse strains in the sample size of our analysis (Table 4). Although further analyses may be required, DNA-PKcs does not appear to be required for preferential rAAV integration at DNA palindromes. Since we often found microhomology up to 8 bp at rAAV integration sites in both DNA-PKcs-proficient and -deficient mice, a DNA-PK-independent microhomology-dependent alternative end-joining pathway(s) (13, 26) might be operative in rAAV integration.

DNA palindromes with an arm length range of at least 28 to 33 bp were relatively stably maintained in bacteria. The stability of DNA palindromes upon cloning in bacteria was assessed for four DNA palindromes recurrently labeled by rAAV integrations (coordinates chr 10: 98632057, chr 11: 44726916, chr 15: 99702254, and chr 19: 11606158; see Table 3) to verify our experimental observations described above. For this, pBluescript KSII($-$) plasmid containing each DNA palindrome was cloned in bacteria, and plasmid DNA was recovered from each colony. The stability of DNA palindromes then was analyzed by restriction enzyme digestion and sequencing.

EcoRI and HindIII double digestion of plasmid DNA containing the DNA palindrome with an arm length of 47 bp (chr 15: 99702254) recovered from bacteria consistently showed a deletion of approximately ~60 bp by 2% agarose gel electrophoresis. Sequencing analysis of a plasmid clone revealed a 64-bp deletion within the 92-bp central AT dinucleotide repeat tract [(AT)n; $n = 46$] in the DNA palindrome. In the other three DNA palindromes with an arm length range of 28 to 33 bp (coordinates chr 10: 98632057, chr 11: 44726916, and chr 19: 11606158), no apparent deletion was detected by EcoRI and HindIII double digestion followed by 2% agarose gel electrophoresis. Sequencing analysis of these three DNA palindromes cloned in bacteria indicated that palindromes had been relatively stably maintained in bacteria, although complete sequencing of entire PCR products cloned in plasmids often was not successful due to the presence of a long stretch of AT dinucleotide repeats, which significantly reduced sequence signals in electropherograms beyond the AT dinucleotide repeats. Fluctuations of the lengths of PCR-amplified AT dinucleotide repeats within the DNA palindromes (-6 bp to $+12$ bp for deletions and additions, respectively) were observed depending on the plasmid clones sequenced.

To further assess the stability of DNA palindromes at coordinates chr 10: 98632057, chr 11: 44726916, and chr 19: 11606158, transformed bacteria from a single colony were replated on LB agar plates, and plasmid DNA recovered from 18 colonies representing the progeny of a single colony was analyzed by EcoRI and HindIII double digestion followed by 2% agarose gel electrophoresis. Three of 18 colonies for palindrome chr 10: 98632057 carried plasmids with an ~20-bp deletion; 2 of 18 colonies for palindrome chr 11: 44726916 carried plasmids with an ~50 bp deletion; and none of 18 colonies for palindrome chr 19: 11606158 had plasmids with a deletion. Although we observed deletions in some clones, the results indicate that plasmid DNA molecules with a deletion should constitute only a minor fraction of total plasmid DNA molecules harboring each of the three DNA palindromes.

DISCUSSION

In the present study, we demonstrated that DNA palindromes with a modest arm length are a significant target for rAAV vector genome integrations in mice. Our observation that DNA palindromes with a modest arm length (total length of ≥ 40 bp) could have an impact in vivo and serve as a platform for preferential rAAV integration provides significant insights into the mechanism of rAAV integration and the sta-
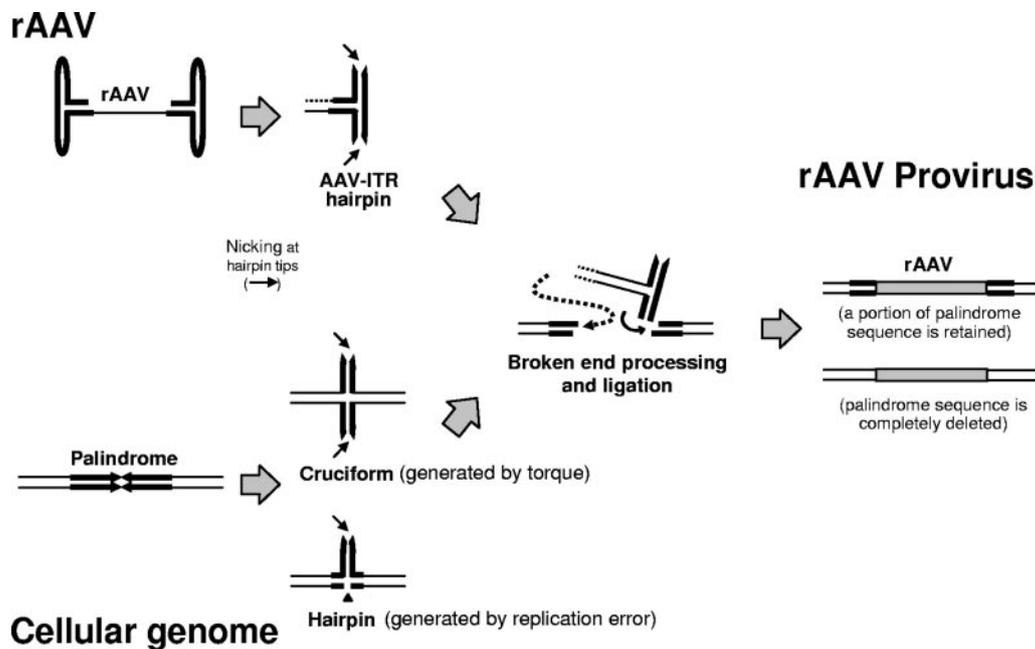
FIG. 9. Proposed model for rAAV integration at palindromes in either a cruciform or hairpin structure. In this model, hairpin loops of palindromic AAV-ITR and DNA palindromes on the genome are nicked, trimmed, and joined by cellular DNA repair mechanisms. Various degrees of host chromosomal DNA deletions occur at rAAV integration, resulting in exclusively center-deleted integrations.

bility of this DNA motif in the genome of quiescent somatic cells in animals. Miller et al. have recently demonstrated that rAAV vectors integrate at DNA double-strand breaks created by ISceI digestion in cultured cells. Based on their experimental observations, they proposed that rAAV vectors do not cause chromosomal breaks and integrate at preexisting chromosomal breakage sites (30). Although whether rAAV vectors integrate only at preexisting DNA breaks or whether rAAV vectors create DNA breaks to make a platform for integration may require further studies, our observations are at least consistent with the notion that rAAV integrates at fragile sites in the host chromosomal DNA.

It is important to understand how rAAV vectors integrate at DNA palindromes. A stream of evidence in the studies of rare cases of DNA palindrome-associated de novo human chromosome translocations (18), central rearrangement of exogenously introduced genomic DNA palindromes in mice (1, 4, 22, 23), and in vitro studies of AT-rich palindromes isolated from translocation sites (19), as well as studies in yeast (25, 33, 42), have indicated that palindromes form a hairpin or cruciform structure, and then palindrome resolution and recombination follow, resulting in genomic alteration (reviewed in references 18 and 23). In the present study, the 134 rAAV integration events in the vicinity of DNA palindromes (arm length, ≥20 bp), including 26 rAAV integrations that occurred within DNA palindromes, exhibited exclusively the center-deleted integration pattern (Fig. 5C). Although an argument may exist that such a significant bias has resulted from deletions of DNA palindromes in bacteria, most DNA palindromes in the study were those in a modest size range that are generally not grossly unstable upon cloning in *E. coli* (21). In fact, as long as we analyzed the stability of four DNA palindromes in DH10B *E. coli* (palindrome coordinates chr 10: 98632057, chr 11:

44726916, chr 15: 99702254, and chr 19: 11606158 in Table 3), palindromes with an arm length range of at least 28 to 33 bp could be relatively stably maintained in this strain of bacteria. Only one palindrome with a longer arm length of 47 bp (total length of 94 bp; coordinate chr 15: 99702254) consistently exhibited an ~60-bp deletion within the central AT dinucleotide repeats. It should be noted that even when we focus on only the 93 rAAV-labeled palindromes with arm lengths of ≤33 bp, only 5 of them showed the center-retained integration pattern. All of the above results indicate that DNA palindromes were the primary targets for rAAV integration, and various degrees of chromosomal DNA deletions occurred at integration, resulting in exclusively center-deleted integrations. Based on these observations, we propose a model for how rAAV might integrate at DNA palindromes (Fig. 9). In this model, DNA palindromes form double-stranded cruciforms by torsional stress or single-stranded hairpins by strand slippage or misalignment. Double-stranded or single-stranded breaks then are created by not-yet-defined mechanisms, forming a platform for rAAV vector integration.

It is also important to address whether rAAV genomes are fortuitously captured by palindromes undergoing a normal palindrome break-repair process (4) or whether rAAV transduction itself either promotes palindrome breakage or alters the repair process of palindromes. At present, neither scenario can be completely dismissed out of hand. If the proposal by Miller et al. that rAAV vectors do not create DNA breaks upon infection and integrate only at preexisting breaks (30) is correct, many DNA palindromes of modest length in the genome should have been naturally broken in quiescent somatic cells under normal cellular metabolic activities in animals. Such broken palindromes could be rejoined as a normal part of palindromic DNA metabolism (23), and rAAV may become

incorporated at emerging palindrome breakage sites. An alternative possibility is that rAAV DNA triggers a cellular DNA damage response (11), either disrupting the normal repair process and causing palindromes to become rAAV labeled or creating new breaks at DNA palindromes. Along these lines, the introduction of rAAV genomes could perturb signaling and provoke improper repair, putting quiescent rAAV-transduced cells at risk for the accumulation of a variety of DNA aberrations. These sites of misrepair need not be limited to palindromes and need not all acquire rAAV adducts. This possibility might explain an observed increase in the incidence of liver cancer in at least some rAAV-treated animals in which insertional mutagenesis was not likely the case (5).

It was a surprise that preferential rAAV integration at DNA palindromes was not revealed in our previous study, in which we investigated rAAV integrations in mouse liver using a hereditary tyrosinemia type I mouse model and in vivo selection (Fig. 3K and L) (40). This trend also was not observed in another large-scale rAAV integration site study by Miller et al., in which rAAV integration events were collected from dividing human cells infected with rAAV in vitro under no selective pressure (32). Since Miller et al. did not specifically focus on DNA palindromes in their analysis, we reanalyzed 815 rAAV integration sites reported by them (405 and 410 of the left and right sides of rAAV-host cellular DNA junctions in human cells, respectively, from data made available online; see Materials and Methods) in the same way as we did for the present study and found no evidence of preferential integration at DNA palindromes. At this point, it remains unclear why there were significant differences in the experimental observations among these three studies. There are a number of parameters that we need to consider to interpret the differences (i.e., mouse cells versus human cells, quiescent cells versus proliferating cells, in vivo assay versus in vitro assay, selection versus no selection, differences in the vector construct and plasmid rescue procedures, and so on). Although these possibilities should be investigated further in future experiments, a tempting and straightforward inference is that entering the cell cycle will erase palindrome labeling. One key difference among the three studies is the proliferation status of the sampled cells. It has been reported that DNA breaks can persist as unrepaired DNA lesions in nondividing cells both in vitro (47) and in animal tissues in vivo (50). The palindrome-labeling events may thus create a DNA lesion that is eradicated through repair or apoptosis if the cell starts to proliferate but may remain unrepaired if the cell remains quiescent. Once a cell with unrepaired DNA lesions with rAAV adducts enters the cell cycle, the unbroken sister chromatid will become available in late S and G_2 phases, in which the cell may use homologous recombination to eliminate any rAAV adducts. If the cell enters the cell cycle and fails to repair the DNA lesions with rAAV adducts, it would die by apoptosis, thus eliminating the cell itself. Alternatively, the quiescent state of cells studied might create an environment in which DNA palindromes become more prone to breakage.

In summary, our demonstration that DNA palindromes are a significant target for rAAV vector integration provides significant new insights into the mechanisms of rAAV integration in vivo and the biological impact of DNA palindromes on the mouse genome. Further investigation might productively focus

on how and whether rAAV evokes DNA repair machinery directed toward little-understood processes involved in the metabolism of DNA hairpin structures in both rAAV genomes (i.e., inverted terminal repeats) and cellular genomes. The demonstration that rAAV preferentially integrates into DNA palindromes and presumably can label a break-prone subset of DNA palindromes is a beginning. Further studies on the interactions between rAAV vector genome, host cellular DNA, and cellular DNA repair machinery will not only provide clues on how to improve the current rAAV systems for human gene therapy but will also suggest new opportunities to explore the intrinsic sources of genome instability, palindrome metabolism, carcinogenesis, and aging.

REFERENCES

1. **Akgün, E., J. Zahn, S. Baumes, G. Brown, F. Liang, P. J. Romanienko, S. Lewis, and M. Jasin.** 1997. Palindrome resolution and recombination in the mammalian germ line. Mol. Cell. Biol. **17:**5559–5570.
2. **Butler, D. K., L. E. Yasuda, and M. C. Yao.** 1996. Induction of large DNA palindrome formation in yeast: implications for gene amplification and genome stability in eukaryotes. Cell **87:**1115–1122.
3. **Collick, A., J. Drew, J. Penberth, P. Bois, J. Luckett, F. Scaerou, A. Jeffreys, and W. Reik.** 1996. Instability of long inverted repeats within mouse transgenes. EMBO J. **15:**1163–1171.
4. **Cunningham, L. A., A. G. Cote, C. Cam-Ozdemir, and S. M. Lewis.** 2003. Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism. Mol. Cell. Biol. **23:**8740–8750.
5. **Donsante, A., C. Vogler, N. Muzyczka, J. M. Crawford, J. Barker, T. Flotte, M. Campbell-Thompson, T. Daly, and M. S. Sands.** 2001. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther. **8:**1343–1346.
6. **Edelmann, L., E. Spiteri, K. Koren, V. Pulijaal, M. G. Bialer, A. Shanske, R. Goldberg, and B. E. Morrow.** 2001. AT-rich palindromes mediate the constitutional t(11;22) translocation. Am. J. Hum. Genet. **68:**1–13.
7. **Ford, M., and M. Fried.** 1986. Large inverted duplications are associated with gene amplification. Cell **45:**425–430.
8. **Goodarzi, A. A., Y. Yu, E. Riballo, P. Douglas, S. A. Walker, R. Ye, C. Harer, C. Marchetti, N. Morrice, P. A. Jeggo, and S. P. Lees-Miller.** 2006. DNA-PK autophosphorylation facilitates Artemis endonuclease activity. EMBO J. **25:**3880–3889.
9. **Gotter, A. L., T. H. Shaikh, M. L. Budarf, C. H. Rhodes, and B. S. Emanuel.** 2004. A palindrome-mediated mechanism distinguishes translocations involving LCR-B of chromosome 22q11.2. Hum. Mol. Genet. **13:**103–115.
10. **Inagaki, K., S. Fuess, T. A. Storm, G. A. Gibson, C. F. McTiernan, M. A. Kay, and H. Nakai.** 2006. Robust systemic transduction with AAV9 vectors in mice: efficient global cardiac gene transfer superior to that of AAV8. Mol. Ther. **14:**45–53.
11. **Jurvansuu, J., K. Raj, A. Stasiak, and P. Beard.** 2005. Viral transport of DNA damage that mimics a stalled replication fork. J. Virol. **79:**569–580.
12. **Kato, T., H. Inagaki, K. Yamada, H. Kogo, T. Ohye, H. Kowa, K. Nagaoka, M. Taniguchi, B. S. Emanuel, and H. Kurahashi.** 2006. Genetic variation affects de novo translocation frequency. Science **311:**971.
13. **Katsura, Y., S. Sasaki, M. Sato, K. Yamaoka, K. Suzukawa, T. Nagasawa, J. Yokota, and T. Kohno.** 2007. Involvement of Ku80 in microhomology-mediated end joining for DNA double-strand breaks in vivo. DNA Repair (Amsterdam) **6:**639–648.
14. **Katz, R. A., K. Gravuer, and A. M. Skalka.** 1998. A preferred target DNA structure for retroviral integrase in vitro. J. Biol. Chem. **273:**24190–24195.
15. **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.

16. **Kurahashi, H., and B. S. Emanuel.** 2001. Long AT-rich palindromes and the constitutional t(11;22) breakpoint. Hum. Mol. Genet. **10:**2605–2617.

17. **Kurahashi, H., and B. S. Emanuel.** 2001. Unexpectedly high rate of de novo constitutional t(11;22) translocations in sperm from normal males. Nat. Genet. **29:**139–140.

18. **Kurahashi, H., H. Inagaki, T. Ohye, H. Kogo, T. Kato, and B. S. Emanuel.** 2006. Palindrome-mediated chromosomal translocations in humans. DNA Repair (Amsterdam) **5:**1136–1145.

19. **Kurahashi, H., H. Inagaki, K. Yamada, T. Ohye, M. Taniguchi, B. S. Emanuel, and T. Toda.** 2004. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. J. Biol. Chem. **279:**35377–35383.

20. **Kurahashi, H., T. Shaikh, M. Takata, T. Toda, and B. S. Emanuel.** 2003. The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats. Am. J. Hum. Genet. **72:**733–738.

21. **Leach, D. R.** 1994. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. Bioessays **16:**893–900.

22. **Lewis, S., E. Akgun, and M. Jasin.** 1999. Palindromic DNA and genome stability. Further studies. Ann. N. Y. Acad. Sci. **870:**45–57.

23. **Lewis, S. M., and A. G. Cote.** 2006. Palindromes and genomic stress fractures: bracing and repairing the damage. DNA Repair (Amsterdam) **5:**1146–1160.

24. **Lisniæ, B., I. K. Svetec, H. Saric, I. Nikolic, and Z. Zgaga.** 2005. Palindrome content of the yeast Saccharomyces cerevisiae genome. Curr. Genet. **47:**289–297.

25. **Lobachev, K. S., D. A. Gordenin, and M. A. Resnick.** 2002. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell **108:**183–193.

26. **Ma, J. L., E. M. Kim, J. E. Haber, and S. E. Lee.** 2003. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol. Cell. Biol. **23:**8820–8828.

27. **Ma, Y., U. Pannicke, K. Schwarz, and M. R. Lieber.** 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell **108:**781–794.

28. **Manno, C. S., V. R. Arruda, G. F. Pierce, B. Glader, M. Ragni, J. Rasko, M. C. Ozelo, K. Hoots, P. Blatt, B. Konkle, M. Dake, R. Kaye, M. Razavi, A. Zajko, J. Zehnder, H. Nakai, A. Chew, D. Leonard, J. F. Wright, R. R. Lessard, J. M. Sommer, M. Tigges, D. Sabatino, A. Luk, H. Jiang, F. Mingozzi, L. Couto, H. C. Ertl, K. A. High, and M. A. Kay.** 2006. Successful transduction of liver in hemophilia by AAV-factor IX and limitations imposed by the host immune response. Nat. Med. **12:**342–347.

29. **McCarty, D. M., S. M. Young, Jr., and R. J. Samulski.** 2004. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu. Rev. Genet. **38:**819–845.

30. **Miller, D. G., L. M. Petek, and D. W. Russell.** 2004. Adeno-associated virus vectors integrate at chromosome breakage sites. Nat. Genet. **36:**767–773.

31. **Miller, D. G., E. A. Rutledge, and D. W. Russell.** 2002. Chromosomal effects of adeno-associated virus vector integration. Nat. Genet. **30:**147–148.

32. **Miller, D. G., G. D. Trobridge, L. M. Petek, M. A. Jacobs, R. Kaul, and D. W. Russell.** 2005. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J. Virol. **79:**11434–11442.

33. **Nag, D. K., and A. Kurst.** 1997. A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast Saccharomyces cerevisiae. Genetics **146:**835–847.

34. **Nakai, H., S. Fuess, T. A. Storm, S. Muramatsu, Y. Nara, and M. A. Kay.** 2005. Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. J. Virol. **79:**214–224.

35. **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.

36. **Nakai, H., E. Montini, S. Fuess, T. A. Storm, M. Grompe, and M. A. Kay.** 2003. AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat. Genet. **34:**297–302.

37. **Nakai, H., E. Montini, S. Fuess, T. A. Storm, L. Meuse, M. Finegold, M. Grompe, and M. A. Kay.** 2003. Helper-independent and AAV-ITR-independent chromosomal integration of double-stranded linear DNA vectors in mice. Mol. Ther. **7:**101–111.

38. **Nakai, H., T. A. Storm, and M. A. Kay.** 2000. Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular re-

combination are responsible for stable transduction of liver in vivo. J. Virol. **74:**9451–9463.

39. **Nakai, H., C. E. Thomas, T. A. Storm, S. Fuess, S. Powell, J. F. Wright, and M. A. Kay.** 2002. A limited number of transducible hepatocytes restricts a wide-range linear vector dose response in recombinant adeno-associated virus-mediated liver transduction. J. Virol. **76:**11343–11349.

40. **Nakai, H., X. Wu, S. Fuess, T. A. Storm, D. Munroe, E. Montini, S. M. Burgess, M. Grompe, and M. A. Kay.** 2005. Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. J. Virol. **79:**3606–3614.

41. **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.

42. **Narayanan, V., P. A. Mieczkowski, H. M. Kim, T. D. Petes, and K. S. Lobachev.** 2006. The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell **125:**1283–1296.

43. **Neiman, P. E., R. Kimmel, A. Icreverzi, K. Elsaesser, S. J. Bowers, J. Burnside, and J. Delrow.** 2006. Genomic instability during Myc-induced lymphomagenesis in the bursa of Fabricius. Oncogene **25:**6325–6335.

44. **Paull, T. T., and M. Gellert.** 1999. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev. **13:**1276–1288.

45. **Posey, J. E., M. J. Pytlos, R. R. Sinden, and D. B. Roth.** 2006. Target DNA structure plays a critical role in RAG transposition. PLoS Biol. **4:**e350.

46. **Rattray, A. J., B. K. Shafer, B. Neelam, and J. N. Strathern.** 2005. A mechanism of palindromic gene amplification in Saccharomyces cerevisiae. Genes Dev. **19:**1390–1399.

47. **Rothkamm, K., and M. Lobrich.** 2003. Evidence for a lack of DNA double-strand break repair in human cells exposed to very low X-ray doses. Proc. Natl. Acad. Sci. USA **100:**5057–5062.

48. **Schmidt, M., P. Zickler, G. Hoffmann, S. Haas, M. Wissler, A. Muessig, J. F. Tisdale, K. Kuramoto, R. G. Andrews, T. Wu, H. P. Kiem, C. E. Dunbar, and C. von Kalle.** 2002. Polyclonal long-term repopulating stem cell clones in a primate model. Blood **100:**2737–2743.

49. **Schröder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:**521–529.

50. **Sedelnikova, O. A., I. Horikawa, D. B. Zimonjic, N. C. Popescu, W. M. Bonner, and J. C. Barrett.** 2004. Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks. Nat. Cell Biol. **6:**168–170.

51. **Song, S., Y. Lu, Y. K. Choi, Y. Han, Q. Tang, G. Zhao, K. I. Berns, and T. R. Flotte.** 2004. DNA-dependent PK inhibits adeno-associated virus DNA integration. Proc. Natl. Acad. Sci. USA **101:**2112–2116.

52. **Tanaka, H., D. A. Bergstrom, M. C. Yao, and S. J. Tapscott.** 2006. Large DNA palindromes as a common form of structural chromosome aberrations in human cancers. Hum. Cell **19:**17–23.

53. **Tanaka, H., D. A. Bergstrom, M. C. Yao, and S. J. Tapscott.** 2005. Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification. Nat. Genet. **37:**320–327.

54. **Tanaka, H., S. J. Tapscott, B. J. Trask, and M. C. Yao.** 2002. Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. Proc. Natl. Acad. Sci. USA **99:**8772–8777.

55. **Tapia-Páez, I., M. Kost-Alimova, P. Hu, B. A. Roe, E. Blennow, L. Fedorova, S. Imreh, and J. P. Dumanski.** 2001. The position of t(11;22)(q23;q11) constitutional translocation breakpoint is conserved among its carriers. Hum. Genet. **109:**167–177.

56. **Trujillo, K. M., and P. Sung.** 2001. DNA structure-specific nuclease activities in the Saccharomyces cerevisiae Rad50*Mre11 complex. J. Biol. Chem. **276:**35458–35464.

57. **Wang, G., and K. M. Vasquez.** 2006. Non-B DNA structure-induced genetic instability. Mutat. Res. **598:**103–119.

58. **Wang, Y., and F. C. Leung.** 2006. Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. FEBS Lett. **580:**1277–1284.

59. **Wu, X., Y. Li, B. Crise, and S. M. Burgess.** 2003. Transcription start regions in the human genome are favored targets for MLV integration. Science **300:**1749–1751.

60. **Yasuda, L. F., and M. C. Yao.** 1991. Short inverted repeats at a free end signal large palindromic DNA formation in Tetrahymena. Cell **67:**505–516.