Stable Secondary Structure near the Nicking Site for Adeno-Associated Virus Type 2 Rep Proteins on Human Chromosome 19

Ming Y. Jang,¹ OrLando H. Yarborough III,^{1†} Gary B. Conyers,¹
Peter McPhie,² and Roland A. Owens^{1*}

Laboratory of Molecular and Cellular Biology¹ and Laboratory of Biochemistry and Genetics,²
National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of
Health, Department of Health and Human Services, Bethesda, Maryland

Received 5 August 2004/Accepted 22 October 2004

Adeno-associated virus serotype 2 (AAV-2) can preferentially integrate its DNA into a 4-kb region of human chromosome 19, designated *AAVSI*. The nicking activity of AAV-2's Rep68 or Rep78 proteins is essential for preferential integration. These proteins nick at the viral origin of DNA replication and at a similar site within *AAVSI*. The current nicking model suggests that the strand containing the nicking site is separated from its complementary strand prior to nicking. In AAV serotypes 1 through 6, the nicking site is flanked by a sequence that is predicted to form a stem-loop with standard Watson-Crick base pairing. The region flanking the nicking site in *AAVSI* (5'-GGCGGCGGT/TGGGGCTCG-3' [the slash indicates the nicking site]) lacks extensive potential for Watson-Crick base pairing. We therefore performed an empirical search for a stable secondary structure. By comparing the migration of radiolabeled oligonucleotides containing wild-type or mutated sequences from the *AAVSI* nicking site to appropriate standards, on native and denaturing polyacrylamide gels, we have found evidence that this region forms a stable secondary structure. Further confirmation was provided by circular dichroism analyses. We identified six bases that appear to be important in forming this putative secondary structure. Mutation of five of these bases, within the context of a double-stranded nicking substrate, reduces the ability of the substrate to be nicked by Rep78 *in vitro*. Four of these five bases are outside the previously recognized GTTGG nicking site motif and include parts of the CTC motif that has been demonstrated to be important for integration targeting.

Adeno-associated virus serotype 2 (AAV-2) is a human parvovirus. AAV-2 normally requires a helper virus, such as an adenovirus or herpesvirus for productive infection (7). In the absence of helper virus, AAV-2 can establish a latent infection by integrating its genome into the host DNA. AAV-2 preferentially integrates its DNA into a 4-kb region of human chromosome 19 (19q13-qter), designated *AAVSI* (27–30). Subsequent infection by helper virus leads to rescue and productive infection. Preferential integration into *AAVSI* requires the Rep68 or Rep78 protein (Rep68/78), encoded by AAV-2, as well as specific DNA sequences within the virus genome and *AAVSI* (3, 11, 35, 42, 47, 55, 57, 59). Binding sites for Rep68/78 within both the viral DNA and *AAVSI* appear to be required (35, 42, 55, 59). These sites, called Rep recognition sequences (RRSs), are comprised of imperfect repeats of the sequence 5'-GCTC-3' or its complement, 5'-GAGC-3' (11, 20, 59). Rep68/78 can form a bridge between RRS-containing DNAs that may facilitate preferential integration (34, 59).

Rep68 and Rep78 also have a nucleoside triphosphate-dependent, strand-specific, site-specific endonuclease (nicking)

activity (22, 23). A Rep68/78 nicking site is called a terminal resolution site (*trs*) because of the role of such nicking sites in AAV-2 replication (22, 50, 52). Nicking requires both a specific sequence flanking the *trs* and a nearby RRS (Fig. 1) (5, 6, 22, 32, 51, 64). RRS/*trs* combinations have been identified in the inverted terminal repeat (ITR) origins of DNA replication for all AAV serotypes that have been sequenced (10, 22, 65), as well as *AAVSI* (11, 57, 59).

Multiple lines of evidence suggest that both binding and nicking of *AAVSI* by Rep68/78 are required for preferential integration (32, 35, 56, 57, 59, 61, 64). This has led to the current hypothesis that *AAVSI* is the preferred integration locus for AAV-2 because it contains the best Rep68/78 nicking site within the human genome. Testing this hypothesis requires detailed knowledge of the nicking mechanism and the specific sequences recognized by Rep68/78.

Although the RRSs have been relatively well characterized with regard to the locations of specific contacts by Rep68/78 (13, 20, 31, 39, 41, 44, 59), analysis of the *trs* region has been more problematic. This is because nicking site recognition appears to require that the DNA strands be separated in the region of the *trs* (6, 14, 49, 51). Rep68/78 have two nucleoside triphosphate-dependent DNA helicase activities. One helicase activity is nonspecific with regard to DNA sequence but requires a 3' single-stranded tail (22, 23, 63, 68). The second activity can unwind a blunt-ended fragment but requires an RRS (68). The current model for nicking suggests that Rep68/78 first binds at the RRS and begins to separate bidi-

* Corresponding author. Mailing address: Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bldg. 8, Rm. 310, National Institutes of Health, Department of Health and Human Services, 8 Center Dr. MSC 0840, Bethesda, MD 20892-0840. Phone: (301) 496-3359. Fax: (301) 402-0053. E-mail: address: ro6n@nih.gov.

† Present address: Genetics Graduate Program, Yale University School of Medicine, New Haven, CT 06520-8005.

FIG. 1. Common sequence motifs between AAVS1 (40, 57, 59) and the AAV-2-ITR (54). Only the *trs*-containing strands are shown. Homologous sequences are underlined. Each *trs* is indicated by a slash. Each RRS is indicated by a bracket. The 18-mers characterized in this study (S1-18 and ITR-18) are also bracketed.

rectionally the two strands (6, 14, 51, 68). Once the region of the *trs* is separated from its complementary strand, the *trs*-containing strand is hypothesized to fold into a stable secondary structure (6). Rep68 or Rep78 can then nick at a site that it recognizes by a combination of its base sequence and its secondary structural context (6, 49). A putative stem-loop structure has been identified in the region of the *trs* for AAV serotypes 1 through 6 (6). Figure 2 shows that a 5- to 7-bp stem could be formed by standard Watson-Crick base pairing (i.e., A-T and G-C) (6, 58). Individual mutation of several of the bases within the putative stem of the AAV-2 *trs* region resulted in a reduction in the ability of the site to be nicked (6).

Comparison of the AAVS1 *trs* region with that of AAV-2 shows several regions of base identity (Fig. 2); the 5'-GT TG G-3' core motif immediately flanking the *trs* (the gap represents the primary cut site), 2 of the 6 bases 5' of the core motif, and 3 of the 11 bases 3' of the core motif. Most of the bases believed to be involved in the AAV-2 *trs* stem structure are not conserved in AAVS1 (6, 57). Furthermore, no stem-loop sim-

FIG. 2. Conserved sequences flanking the known and suspected Rep68/78 nicking sites of AAV serotypes 1 to 6 (6) and AAVS1 (57). Underlined bases have the potential to form a stem by Watson-Crick base pairing. Bases conserved between all sequences (S1/AAV1-6) and between all sequences except AAV-5 (S1/AAV1-4/6) are indicated. The double-pointed arrows indicate bases conserved between AAVS1 and the AAV-2-ITR. The 18-base segment of AAVS1 that is the focus of the present study (S1-18) is bracketed. Bases C3, C6, T9, G12, C15, and G18 of S1-18 are numbered.

TABLE 1. Oligonucleotides used in gel mobility assays

Designation	Sequence
Mark-25	GACTGACTGACTGACTGACTG
Mark-18	GACTGACTGACTGACTGT
Mark-12	GACTGACTGACT
ITR-18	TGATGGAGT TGGCCACTC
S1-11.....	GCGGT TGGGGC
S1-12.....	GGCGGT TGGGGC
S1-13.....	CGGCGGT TGGGGC
S1-14.....	CGGCGGT TGGGGCT
S1-15.....	GCGGCGGT TGGGGCT
S1-16.....	GCGGCGGT TGGGGCTC
S1-17.....	GGCGGCGGT TGGGGCTC
S1-18.....	GGCGGCGGT TGGGGCTCG
S1-18-Pseudo.....	CCGCCGCCA ACCCCGAGC
S1-18-RevComp.....	CGAGCCCCA ACCGCCGCC

ilar to those proposed for the AAV sequences could be formed in AAVS1 by using Watson-Crick base pairing (58). In spite of this, our previous work showed only about a twofold difference in the ability of Rep78 to nick the linear form of the AAV-2 ITR *trs*/RRS versus the AAVS1 *trs*/RRS (64). In the present study, we provide evidence for what appears to be a stable secondary structure based, at least in part, on non-Watson-Crick base pairing, in the AAVS1 *trs* region. Elements of this secondary structure may also contribute to the ability of AAVS1 to be nicked by Rep68/78.

MATERIALS AND METHODS

Oligonucleotides. Please note that previously published corrections (35, 57, 59) to the initially published sequence of this portion of AAVS1 (28) (GenBank accession no. S51329) have been confirmed by the sequencing of human chromosome 19 (GenBank accession no. AC010327). We use the corrected sequence in this work.

All oligonucleotides were synthesized by Invitrogen Life Technologies (Carlsbad, Calif.). Standard purity (desalted) oligonucleotides were used for annealing and insertion into plasmids. Unless otherwise stated, oligonucleotides used for DNA migration analyses were trityl column purified by the company. Oligonucleotides used for circular dichroism (CD) analyses and size standards (Mark-12, Mark-18, and Mark-25) were purified by polyacrylamide gel electrophoresis by the manufacturer.

DNA migration analysis. The oligonucleotides listed in Tables 1 and 2 (500 pmol each) were 5' end labeled by using T4 polynucleotide kinase (New England Biolabs, Beverly, Mass.) and [γ - 32 P]ATP (Perkin-Elmer Life Sciences, Boston, Mass.). After labeling, the samples shown in Fig. 3A, 4, 9A, and 10 were extracted with phenol-chloroform-isoamyl alcohol (25:24:1) and ethanol precipitated, and the labeled oligonucleotides were resuspended in 75 μ l of T₁₀E₁ buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA). For the samples shown in Fig. 3B, the labeling reaction was heated to 65°C for 30 min to inactivate the kinase and unincorporated label was removed by a Microspin G-25 column (Amersham Pharmacia Biotech, Inc., Piscataway, N.J.). T₁₀E₁ buffer was then used to adjust the concentration of each oligonucleotide to ca. 1 to 10 μ M).

We used a model V-16 vertical gel electrophoresis system (Life Technologies, Inc., Gaithersburg, Md.). Standards were routinely run in duplicate or triplicate. This was to provide a clear horizontal axis for comparison of migration on non-denaturing gels and to compensate for "smiling" in the denaturing gels. Smiling was generally not a problem with the non-denaturing gels, as judged by little or no variation between the three standards.

For non-denaturing gels, aliquots of labeled, single-stranded oligonucleotides were combined with non-denaturing loading mix (final concentrations, 5% glycerol [vol/vol], 0.04% bromophenol blue, and 0.04% xylene cyanol FF).

Three to six microliters of each mixture was then electrophoresed (6.25 V/cm) on a 0.8-mm-thick 8% polyacrylamide gel containing 1 $\times$ TBE buffer (45) at room temperature. The gel was then dried and autoradiographed by using X-Omat AR or Biomax MR X-ray film (Kodak, Rochester, N.Y.).

For denaturing gels, 2- μ l aliquots of labeled, single-stranded oligonucleotides were combined with 2.7 μ l of formamide loading mix (95% formamide, 20 mM

TABLE 2. Mutations in nicking substrates and 18-mers

Designation ^a	Sequence ^b	Mobility ^c change	Nicking (% wild type) ^d
S1-WT ^e	GGCGGCGGT TGGGGCTCG	0	100
S1-G2A-C3T-G4A	GataGCGGT TGGGGCTCG	Slower	18*
S1-G2A-G5A	GaCGaCGGT TGGGGCTCG	0	NT
S1-G1A	aGCGGCGGT TGGGGCTCG	0	102
S1-G2A	GaCGGCGGT TGGGGCTCG	0	66*
S1-C3T	GGtGGCGGT TGGGGCTCG	Slower	44*
S1-G4A	GGCaGCGGT TGGGGCTCG	0	109
S1-G5A	GGCGaCGGT TGGGGCTCG	0	99
S1-C6T	GGCGGtGGT TGGGGCTCG	Slower	92
S1-G7A	GGCGGCaGT TGGGGCTCG	0	70
S1-G8A	GGCGGCGaT TGGGGCTCG	0	90*
S1-T9C	GGCGGCGGc TGGGGCTCG	0	12*
S1-T10C	GGCGGCGGT cGGGGCTCG	Faster	21*
S1-G11A	GGCGGCGGT TaGGGCTCG	0	81
S1-G12A	GGCGGCGGT TGaGGCTCG	0	110
S1-G13A	GGCGGCGGT TGGaGCTCG	0	102
S1-G14A	GGCGGCGGT TGGGaCTCG	Slower	30*
S1-C15T	GGCGGCGGT TGGGGtTCG	Slower	23*
S1-T16C	GGCGGCGGT TGGGGCcCG	Faster	66*
S1-C17T	GGCGGCGGT TGGGGCTcG	0	108
S1-G18A	GGCGGCGGT TGGGGCTCa	0	22*

^a Underlined bases are conserved between AAVS1 and AAV-2-ITR.

^b Mutated bases are in lowercase. For nicking substrates, complementary mutations were made in the other strand.

^c The mutant 18-mer oligonucleotide shown migrates faster than, slower than, or about the same as (0) wild-type S1-18 as indicated.

^d NT, not tested. *, Significantly less than or greater than S1-WT ($P < 0.05$).

^e Wild-type sequence.

EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). These mixtures were subsequently heated in a boiling water bath for 5 min and 3 μ l of each was electrophoresed on a preheated, 0.8-mm-thick 8% polyacrylamide gel containing 6 M urea (SequaGel 8; National Diagnostics, Atlanta, Ga.) at 17.5 V/cm. The gel was then dried and autoradiographed by using X-Omat AR X-ray film.

CD analyses. Solutions of oligonucleotides ($A_{260} = 1$, Approximately 5 to 8 μ M) were made up in 50 mM sodium phosphate buffer (pH 7.2). To examine the effect of potassium, 1 M KCl was added to a final concentration of 47.6 mM (this lowered the sodium phosphate concentration to 47.6 mM). After KCl addition, the samples were incubated overnight at 4°C prior to analysis. Exact concentrations were measured by absorbance, with millimolar extinction coefficients provided by the manufacturer (based on the following values: A = 15.3, C = 7.4, G = 11.8, and T = 9.3). CD spectra were measured in a Jasco J-715 spectropolarimeter of solutions in a 1-cm path length quartz cuvette in a cell holder connected to a Neslab RTE-111 circulating water bath. Spectra were scanned four times, from 325 to 210 nm, and averaged (speed, 50 nm/min; time constant, 1 s). Melting curves were measured between 0 and 90°C, at peak wavelengths (260 or 280 nm), increasing the temperature by 1°C/min. After the water bath had reached the required temperature, the solution was equilibrated for 10 min, and the ellipticity was then measured over 3 min (10 values). Solutions were cooled back to 25°C at 1°C/min, and spectra were measured after 10 min of equilibration. After baseline correction, the measured ellipticities (in millidegrees) were converted into mean residue ellipticities (MRE) by using the formula:

$$[\theta] = \frac{\text{mdeg} \cdot 10 \cdot \text{MRW}}{l \cdot c \cdot 100} \quad (1)$$

where MRW is the mean residue weight, l is the path length in centimeters, and c the concentration in milligrams per milliliter.

The melting curves were analyzed by assuming simple two-state equilibria between native (N) and denatured (D) states. At each temperature, T (in degrees Kelvin), the equilibrium constant for unfolding was calculated as follows:

$$K(T) = \frac{[\theta]_N - [\theta](T)}{[\theta](T) - [\theta]_D} \quad (2)$$

where $[\theta]_N$ and $[\theta]_D$ are the extrapolated MREs of the native and denatured states, respectively, and $[\theta](T)$ is the measured MRE. The temperature dependence of $K(T)$ is given by:

$$K(T) = \exp[-\Delta H(1 - T/T_m)/(RT)] \quad (3)$$

in which ΔH is the apparent enthalpy of denaturation, R is the gas constant, and T_m is the melting temperature [$K(T) = 1$]. Curve fitting was performed by using the PC-MLAB program (Civilized Software, Bethesda, Md.).

Nicking assays. The nicking substrates were created as described previously (64). To make the wild-type AAVS1 nicking substrate, oligonucleotides (53-mers) were synthesized that contained the sequence shown in Fig. 1 and its complement. An EcoRI overhang was added to the 5' end of the oligonucleotide shown, and a BamHI overhang was added to the 5' end of the complementary oligonucleotide. After annealing, the oligonucleotide pair was ligated into plasmid pBluescript II SK(+) (Stratagene, La Jolla, Calif.) that had been digested with BamHI and EcoRI. After amplification of the resulting plasmid in *Escherichia coli* strain DH5 α (Invitrogen), the plasmid was digested with SacII and Eco0109I. The resulting fragments' 5' (Eco0109I) overhangs were end labeled by using T4 polynucleotide kinase and [γ -³²P]ATP. The fragments' 5' overhangs were then filled in by using Klenow polymerase (New England Biolabs) and nonradioactive deoxynucleoside triphosphates. The labeled fragments were run on a nonde-

FIG. 3. DNA migration analysis. End-labeled, single-stranded oligonucleotides (see Tables 1 and 2 for sequences) were run on nonde-naturing, 8% polyacrylamide gels. The gels were then dried and autoradiographed. (A) Samples contained ~ 4 pmol each of either an 18-mer designed to have a low potential to form a stable secondary structure (Mark-18) or 18-mers containing sequences from the AAV-2-ITR (ITR-18) or AAVS1 (S1-18). All three oligonucleotides were purified on a denaturing polyacrylamide gel prior to labeling. (B) Samples contained ~ 1 pmol each of either an 18-mer or 12-mer designed to have a low potential to form a stable secondary structure (Mark-18 or Mark-12, respectively) or 18-mers containing a wild-type sequence from AAVS1 (S1-18), a version of the AAVS1 18-mer with two altered bases (S1-18-G2A-G5A), the pseudo-complement (C for G, A for T, etc.) of the AAVS1 18-mer (S1-18-Pseudo), or the reverse complement of S1-18 (S1-18-RevComp). Mark-12, Mark-18, and S1-18 (Gel) were purified on a denaturing polyacrylamide gel prior to labeling. All other oligonucleotides were trityl column purified prior to labeling.

FIG. 4. Stairstep analysis of the *AAVSI*, *trs*-containing strand. End-labeled, single-stranded oligonucleotides (ca. 8 to 10 pmol) containing 11 to 18 bases of the *AAVSI* *trs*-containing strand (S1-11 through S1-18) or else 25-mer or 18-mer oligonucleotides designed to have limited potential to form a stable secondary structure (Mark-25 or Mark-18, respectively) were run on denaturing (A) or nondenaturing (B) 8% polyacrylamide gels. The number in each oligonucleotide name indicates the number of bases. See Table 1 for sequences. Asterisks indicate oligonucleotides that deviate from the stairstep pattern on the nondenaturing gel.

turing, 6% polyacrylamide gel. The gel was then wrapped in plastic and exposed to X-Ray Film (X-Omat AR or Biomax MR; Kodak). The substrate band was cut out, and the DNA was extracted by the crush-and-soak method (38).

This procedure results in a double-stranded substrate of ca. 140 bp in length that is only labeled on the 5' end of the *trs*-containing strand. This process also produces a substrate large enough for optimal binding by Rep68/78 (12, 31, 66).

Nicking substrates containing the mutated *AAVSI* sequences, except for the T16C mutation, were made as described above, but the starting oligonucleotides contained the base alterations indicated in Table 2. Since the T16C mutation created an additional Eco0109I site, this substrate was made by digesting the plasmid with SacII and XhoI. A special preparation of wild-type *AAVSI* substrate (also with SacII and XhoI) was made for comparison to the T16C substrate.

The *trs* endonuclease (nicking) assays were performed according to the method of Im and Muzyczka (22). Radiolabeled substrate DNA (5, 2.5, or 1.25 fmol per reaction) was incubated with in vitro-synthesized Rep78 in a 20- μ l final reaction volume containing 25 mM HEPES-KOH (pH 7.5), 5 mM MgCl₂, 1 mM dithiothreitol, 0.2 μ g of bovine serum albumin, and 0.4 mM ATP. Rep78 was synthesized by using plasmid pMAT4, which encodes wild-type Rep78 (41) and the TNT coupled T7-rabbit reticulocyte lysate, in vitro transcription-translation system, as directed by the manufacturer (Promega, Madison, Wis.). Two microliters of lysate containing Rep78 was added per reaction mixture for endonuclease assays. Lysate not programmed with plasmid DNA was used as a negative control. The reactions were incubated for 1 h at 37°C and terminated by placing the samples on ice. An aliquot of each reaction mix was combined with formamide loading mix (at a ratio of 1:1.35). This mixture was then boiled to release the nicked fragment. The reaction products were resolved by denaturing polyacrylamide gel electrophoresis as described for the DNA migration analysis. The gel was then dried and autoradiographed. Quantitation was performed by using a Fuji FLA-3000 Fluorescent Image Analyzer and Image Gauge version 3.12 software.

RESULTS

Aberrantly rapid migration of oligonucleotides containing *AAVSI* sequences. To test the hypothesis that a stable secondary structure forms at or near the *AAVSI* nicking site (*trs*), we analyzed the migration of single-stranded oligonucleotides containing various lengths of the *AAVSI* *trs*-containing strand on both native and denaturing polyacrylamide gels. This strategy is similar to that used by Spear et al. (53) to characterize the hairpins formed by the AAV-2 ITRs.

Preliminary experiments showed aberrantly rapid migration, on nondenaturing gels, of a 49-mer containing the *AAVSI* RRS and *trs* (data not shown); a 25-mer that included the *trs* but not the RRS (data not shown); and an 18-mer containing nine bases on either side of the *AAVSI* *trs* (Fig. 3). We therefore concentrated on this 18-base segment, designated S1-18 (Table 1) and established a numbering system that refers to these bases (5'-GGCGGCGGT/TGGGGCTCG-3') as 1 through 18 (Fig. 2). Under this numbering system, Rep68/78 predominantly nicks between bases T9 and T10 (32, 57).

Figure 3 shows a comparison of the migration of S1-18 to both an 18-mer of the corresponding region flanking the AAV-2-ITR *trs* (ITR-18) and a control 18-mer (Mark-18) that is predicted to have a limited capacity for the formation of a stable secondary structure (see Table 1 for sequences). ITR-18 ran faster than Mark-18, a finding consistent with its predicted ability to form a stable stem-loop through Watson-Crick base pairing. S1-18 also migrated faster than Mark-18, suggesting that S1-18 can also form a stable, compact secondary structure. In multiple trials ($n = 10$), S1-18 always ran further than Mark-18, with a mean difference in migration of 0.65 cm (between band centers). A paired-sample t test (67) showed that the probability of consistently seeing this great a difference in migration by chance was <0.0005 . We interpret this aberrant migration as an indication of a stable secondary structure being formed by S1-18.

Given the limited potential for S1-18 to form intramolecular Watson-Crick base pairs, we suspected that the putative stable secondary structure requires unusual base pairings (e.g., G-G pairs). To test this hypothesis, we created a pseudocomplementary 18-mer in which each base in S1-18 is replaced with its Watson-Crick complementary base (e.g., G for C), as well as the reverse complement of S1-18 (Table 1). We predicted that the pseudocomplementary 18-mer (S1-18-Pseudo) and the reverse complement 18-mer (S1-18-RevComp) would not be able to form a stable secondary structure, even though they had the capacity to form the same Watson-Crick base pairs as S1-18. Consistent with this hypothesis, S1-18-Pseudo and S1-18-RevComp comigrate with the Mark-18 control on a nondenaturing polyacrylamide gel (Fig. 3B).

Stairstep analysis. We next sought to identify some of the bases that contribute to this putative stable secondary structure. We hypothesized that if we made a series of oligonucleotides containing increasing numbers of bases from *AAVSI* we would get a stairstep pattern of bands on a denaturing gel (e.g., a 12-mer would migrate slower than an 11-mer), but aberrations to this stairstep pattern on a nondenaturing gel would indicate bases that contribute to the unusual migration of the *AAVSI* 18-mer (and by inference, to the stable secondary structure). The oligonucleotides tested are shown in Table 1.

The gap in the sequence of each *AAVSI* oligonucleotide indicates the position of the nicking site.

Figure 4 shows the results of this analysis. We started with an 11-mer (S1-11) roughly centered on the *trs* (contains bases G5 through C15) and added bases on either end until we reached the sequence of the *AAVSI* 18-mer (S1-18). As predicted, we have a staircase pattern on a denaturing gel (Fig. 4A) and S1-18 comigrates with Mark-18.

When these same oligonucleotides are run on a nondenaturing polyacrylamide gel (Fig. 4B), we no longer get a uniform staircase pattern. We identified three bases in a row (G2, C3, and G4 in our numbering system) that when added do not decrease the mobility of the oligonucleotide. These three bases are therefore believed to be contributors to the stable secondary structure.

CD analyses. To obtain independent evidence for the presence of a stable secondary structure within S1-18, we performed CD analyses. The CD spectra of nucleic acids are dependent on base-base interactions, reflecting the effects of both sequence and secondary structures, including single-stranded stacked bases and multistranded helical structures. Analysis of spectra in terms of secondary structure is largely empirical (24), based on comparisons with model compounds. This situation is complicated by the fact that S1-18 is rich in contiguous guanine and cytosine residues, which can form non-Watson-Crick structures (37). However, we can make inferences about the structure of such oligomers from the shapes and the temperature dependencies of their spectra (15).

The CD spectra of oligonucleotides Mark-18, S1-18, and S1-11, measured at three temperatures, are shown in Fig. 5, 6, and 7, respectively. The very distinct spectra, which they show at low temperatures, can be assumed to reflect their different secondary structures, which are stable under these conditions. For solutions in sodium phosphate buffer, the low-temperature spectra were unchanged after 1 month at 4°C (data not shown). Potassium ions are known to contribute to the formation of certain DNA secondary structures. The addition of potassium chloride (to 47.6 mM final concentration) resulted in large but very slow changes in the CD spectra of S1-18 and S1-11 (Fig. 6B and 7B, respectively, and data not shown), which were complete only after overnight incubation, at 4°C. At 0°C, Mark-18, S1-18, and S1-11 show very distinct intense spectra (Fig. 5, 6, and 7, respectively). However, when the solutions were heated to 70°C, they showed diminished spectra, a finding characteristic of denatured DNAs (37). As will be seen in Fig. 8, the transitions between these spectra occur over a wide temperature range. The temperature dependencies shown in Fig. 8 were analyzed by using equations 2 and 3 (see Materials and Methods) to give the curves shown in the figure. The derived best-fit values of the thermodynamic parameters are shown in Table 3. The thermal unfolding curves of Mark-18 and of S1-18 (in the absence of K⁺) were readily reversible (Fig. 5 and 6A, respectively). Cooling of heated solutions of S1-18 (plus K⁺) or S1-11 (with or without K⁺) produced CD spectra very different from the initial ones (Fig. 6B and 7, respectively). After overnight incubation at 4°C the original spectra were recovered (data not shown), indicating that unfolding was fully reversible.

In sodium phosphate buffer, the CD spectrum of S1-18 shows a positive peak at 280 nm and a weak negative peak at

245 nm (Fig. 6A), characteristic of GC-rich DNA, in the B conformation (18). This structure is unfolded at high temperature and is rapidly reformed on cooling (Fig. 6A). Overnight incubation at 4°C, in the presence of K⁺, produces large changes in the CD spectrum, with formation of an intense positive peak at 260 nm (Fig. 6B), which has been associated with the presence of parallel four-stranded structures formed by runs of G residues (37). These are destroyed by heating and reform very slowly on cooling (Fig. 6B and data not shown).

Even in the absence of K⁺, the CD spectrum of S1-11 shows a positive band at 260 nm, together with another positive band above 290 nm (Fig. 7A), which has been assigned to antiparallel quadruplex structures also formed by association of runs of G residues (37). The addition of K⁺ results in slow increases in the intensities of both of these peaks (Fig. 7B). Again, both of these structures are diminished at high temperatures and reappear very slowly after cooling (Fig. 7B and data not shown).

Mark-18, the control molecule, lacks runs of contiguous G residues, as well as the potential for extensive intramolecular or intermolecular Watson-Crick base pairing. Its low temperature CD spectrum shows a positive band at 280 nm and double negative bands at 255 and 240 nm, a finding consistent with single-stranded base stacking (Fig. 5A), which is only slightly perturbed by prolonged incubation with K⁺ (Fig. 5B and data not shown). It is denatured at high temperatures in transitions, which are rapidly reversed on cooling (Fig. 5B).

Mutational analysis. We next sought to determine the impact of mutating bases flanking the *AAVSI trs*. First, the bases that the staircase analysis suggested are contributors to the secondary structure were mutated individually and together, within the context of single-stranded 18-mers (Fig. 9A). We chose to make transition, rather than transversion mutations, because these would have a minimal impact on the molecular weight of the oligonucleotides.

To our surprise, only mutation of C3 to a T (C3T) detectably altered migration on a nondenaturing polyacrylamide gel. The mutated oligonucleotide migrated about halfway between S1-18 and Mark-18. Simultaneous mutation of all three bases had no greater impact on migration than the C3T mutation (Fig. 9A).

In parallel, we simultaneously mutated these same three bases (and their complements), in the context of a double-stranded Rep68/78 nicking substrate that contained the RRS and *trs* regions of *AAVSI*. The G2A-C3T-G4A mutated substrate was only nicked ca. 18% as efficiently as a substrate containing the wild-type *AAVSI* sequence, when incubated with rabbit reticulocyte lysate containing in vitro-synthesized Rep78 (Fig. 9B and Table 2).

We next decided to individually mutate each base of S1-18 (Fig. 10 and Table 2). Nondenaturing gel migration analyses showed a consistent mobility shift with several mutants. The C3T, C6T, G14A, and C15T mutated oligonucleotides ran slower than the wild-type S1-18 but faster than the control, Mark-18, suggesting a partial disruption of the stable secondary structure. The T10C and T16C mutated oligonucleotides ran faster than the wild-type S1-18, suggesting the possible creation of new base-base interactions that might further compact the secondary structure. Given the high proportion of G residues in the 18-mers, it is likely that new G-C bonds are created by the T10C and T16C mutations. Consistent with this

FIG. 5. CD spectra of oligonucleotide Mark-18 solutions in 50 mM sodium phosphate (pH 7.2) (A) and in 47.6 mM sodium phosphate (pH 7.2) plus 47.6 mM potassium chloride (B) measured in a Jasco J-715 spectropolarimeter (four scans, speed = 50 nm/min, time constant = 1 min). The data are expressed as mean residue ellipticities (degrees · square centimeter/decimole) as a function of wavelength (in nanometers): $\square$, 0°C; +, 70°C; $\diamond$, cooled back to 25°C.

hypothesis, the T16C mutant even ran faster than the other 18-mers on a denaturing gel (Fig. 10D).

We then performed nicking analyses on double-stranded substrates containing these same mutations (Table 2). Assays

were performed at least four times. Over half of the mutations (G1A, G4A, G5A, C6T, G7A, G8A, G11A, G12A, G13A, and C17T) resulted in a mean difference in nicking, compared to wild-type, of $\leq 30\%$. The probability of such a difference oc-

FIG. 6. CD spectra of oligonucleotide S1-18 solutions in 50 mM sodium phosphate (pH 7.2) (A) and in 47.6 mM sodium phosphate (pH 7.2) plus 47.6 mM potassium chloride (B) measured in a Jasco J-715 spectropolarimeter (four scans, speed = 50 nm/min, time constant = 1 min). The data are expressed as mean residue ellipticities (degrees · square centimeter/decimole) as a function of wavelength (in nanometers): □, 0°C; +, 70°C; ◇, cooled back to 25°C.

curing by chance was almost always >5% (67). These differences must therefore be considered to be within the range of experimental error. A second group of mutations (G2A, C3T, T9C, T10C, G14A, C15T, T16C, and G18A) resulted in a mean

reduction in nicking, compared to wild-type, of >30%. The probability of such a difference occurring by chance was <5% for this set of mutations (67). None of the mutations eliminated nicking by Rep78.

FIG. 7. CD spectra of oligonucleotide S1-11 solutions in 50 mM sodium phosphate (pH 7.2) (A) and in 47.6 mM sodium phosphate (pH 7.2) plus 47.6 mM potassium chloride (B) measured in a Jasco J-715 spectropolarimeter (four scans, speed = 50 nm/min, time constant = 1 min). The data are expressed as mean residue ellipticities (degrees $\cdot$ cm²/decimole) as a function of wavelength (in nanometers): $\square$, 0°C; +, 70°C; $\diamond$, cooled back to 25°C.

DISCUSSION

We have demonstrated the presence of a stable secondary structure within single-stranded oligonucleotides containing the Rep68/78 nicking site of *AAVS1*. This secondary structure has been demonstrated by multiple methods. First, an oligo-

nucleotide containing 18 bases of *AAVS1* (S1-18), centered on the nicking site, migrates faster on nondenaturing polyacrylamide gels than a control 18-mer. Second, the results of CD analyses are consistent with the presence of a stable secondary structure. Third, a 12-mer containing *AAVS1 trs* flanking se-

FIG. 8. Temperature dependencies of the peak CD intensities of oligonucleotides in 50 mM sodium phosphate (pH 7.2) (○) or in 47.6 mM sodium phosphate (pH 7.2) plus 47.6 mM potassium chloride (□). (A) Mark-18; (B) S1-18; (C) S1-11. The lines show the curves fitted by using equations 2 and 3 as described in Materials and Methods and in Table 3.

quence (S1-12) failed to migrate more slowly on a nondenaturing polyacrylamide gel than did an 11-mer (S1-11) that, with the exception of one base deleted at the end, is identical to S1-12.

This structure appears to involve unusual base interactions, since there is very limited possibility for Watson-Crick base pairing. In addition, an 18-mer oligonucleotide made by substitution of the bases in S1-18 with the complementary bases

(i.e., G to C, A to T, etc.) and the reverse complement of S1-18 (S1-18-Pseudo and S1-18-RevComp, respectively) migrated with the Mark-18 control (Fig. 3B), suggesting that S1-18-Pseudo and S1-18-RevComp lacked stable secondary structures. One would predict that these pseudocomplementary and reverse complementary DNAs would be able to form equally compact secondary structures to S1-18 if the secondary structure was primarily stabilized by Watson-Crick base pairs.

G-rich sequences, such as oligonucleotides S1-11 and S1-18 (each contains a run of four contiguous G residues), have the potential to form either unimolecular hairpins or multistranded structures, either of which can involve non-Watson-Crick, G-G, base pairing (4, 8, 19, 21, 26, 36, 37, 60). Potassium ions can stabilize certain multistranded structures, such as G quadruplexes (4). The spectra in Fig. 7 show that, even in sodium phosphate buffer, S1-11 forms a mixture of parallel and antiparallel G quadruplexes, which reappear slowly after thermal denaturation. Addition of KCl leads to slow increases in the concentrations of both. S1-18 contains all of the sequence of S1-11, but even after 1 month in solution in sodium phosphate buffer, its CD spectrum shows no indication of the formation of multistranded G quadruplexes (data not shown). Rather its spectrum (Fig. 6) is that of GC-rich B form DNA. This same structure reforms rapidly after thermal denaturation, suggesting that it is intramolecular in nature. In the presence of KCl, the structure is slowly changed into a parallel G quadruplex. These observations suggest that the extra bases in S1-18 stabilize an intramolecular hairpin with B-DNA structure, which inhibits the formation of multistranded G quadruplexes, in the absence of potassium ions. Addition of potassium lowers the free energy of the multistranded complexes and initiates a slow conformational change. In the presence of potassium, these additional sequence elements in S1-18 seem to favor the formation of a parallel G quadruplex.

Under physiological conditions, this intramolecular hairpin can be expected to form rapidly in single-stranded DNA. Even in the presence of intracellular potassium, formation of multistranded structures will be so slow as to have no relevance for this reaction. Consistent with this intramolecular hairpin hypothesis, the absolute and relative migration S1-18 on a non-denaturing gel (in the absence of K^+) was not affected by DNA concentration or boiling and quick chilling the oligonucleotide (data not shown). Further analyses are required to determine

TABLE 3. Best-fit values derived from the melting curves shown in Fig. 8

Oligonucleotide	$[\theta]_N$	$[\theta]_D$	Mean $\pm$ SD	
			T_m ($^{\circ}$ C)	ΔH (kcal/mol)
MARK-18				
No K^+	13,812	6,659	38.4 ± 0.7	17.6 ± 1.1
+ K^+	12,545	6,846	40.6 ± 0.6	20.5 ± 1.4
S1-11				
No K^+	10,958	-694	49.6 ± 3.3	10.3 ± 1.8
+ K^+	15,049	5,862	30.8 ± 0.6	21.5 ± 1.6
S1-18				
No K^+	6,706	941	66.4 ± 2.6	21.9 ± 3.7
+ K^+	21,204	3,679	44.1 ± 0.2	40.6 ± 1.7

FIG. 9. Mutational analysis of bases G2 through G4. (A) End-labeled 18-mer oligonucleotides (~ 2 pmol) Mark-18, S1-18, or mutated versions of S1-18, containing the base alterations indicated (see Tables 1 and 2 for sequences) were run on a nondenaturing, 8% polyacrylamide gel. (B) Radiolabeled nicking substrates were prepared as described in Materials and Methods. Substrates contained double-stranded versions of either the wild-type *AAVSI* sequence shown in Fig. 1 (S1-WT) or a mutated *AAVSI* sequence containing changes in the 3 bp corresponding to bases G2, C3, and G4 (G2A-C3T-G4A) of oligonucleotide S1-18 (See Table 2 for sequence). The substrates (5, 2.5, or 1.25 fmol) were incubated with either 2 μ l of unprogrammed rabbit reticulocyte lysate (U) or 2 μ l of rabbit reticulocyte lysate containing in vitro-synthesized Rep78 (R) in a 20- μ l reaction volume. Samples from each reaction were run on a denaturing 8% polyacrylamide gel. The positions of the substrates (S) and products (P) are indicated.

the exact nature of the *AAVSI* secondary structure, as well as the structure of the Rep68/78-*trs* complex. However, we speculate that the *AAVSI* hairpin involves G-G base pairing. In addition, we interpret the observed increased mobility of the S1-T10C mutated 18-mer (relative to the wild-type S1-18-mer) as indicating that base T10 is unpaired in the wild-type hairpin (Fig. 10B). Mutation of this base to a C may allow it to pair with one of the 11 G residues within the 18-mer to form a more compact hairpin. T10 is one of two conserved Ts immediately flanking the Rep68/78 nicking site (Fig. 2) and the T10C mutation reduces Rep78 nicking by 79% (Table 2). It is interesting that the proposed model for the *trs* hairpin within the AAV-2 ITR has the equivalent flanking T residue unpaired (6). Perhaps an unpaired T is important for nicking site recognition.

We suspect that this stable secondary structure was responsible for the initial errors in sequencing this portion of *AAVSI*

(28, 57, 59). Given the stable nature of this secondary structure (Table 3), we recommend against using annealed complementary *AAVSI* oligonucleotides as double-stranded nicking substrates. We routinely ligate our test DNA into a plasmid. In this way, we lower the risk that the *AAVSI trs* strand would adopt this unusual secondary structure prior to unwinding of the sequence by the Rep68/78 proteins.

Our mutations fell into four categories (Table 2): (i) those that altered gel migration and decreased nicking by $>30\%$ (G2A-C3T-G4A, C3T, T10C, G14A, C15T, and T16C), (ii) one that altered gel migration but not nicking (C6T), (iii) those that reduced nicking by $>30\%$ but did not affect gel migration (G2A, T9C, and G18A), and (iv) those that did not have a large impact on either nicking or gel migration (G1A, G4A, G5A, G7A, G8A, G11A, G12A, G13A, and C17T). We interpret these data as indicating a complex secondary structure surrounding the *AAVSI trs*, in which some, but not all components of the secondary structure enhance nicking site recognition. Our mutational analyses also suggest that components of the stable secondary structure lie on both sides of the nicking site, as is suspected with the AAV-2-ITR (6).

Mutation of either base C15 or T16 alters gel migration and reduces nicking. These bases are part of the CTC motif (C15, T16, and C17; Fig. 1) identified by Meneses et al. as being important for AAV-2 integration targeting to *AAVSI* (40).

We suspect that bases whose mutation affects nicking, but not gel migration, are recognized by Rep78 through base-specific interactions rather than via secondary structure. Consistent with this hypothesis, bases G2 and T9 are two of only four bases within the *AAVSI* 18-mer that are conserved in AAV serotypes 1 through 6 (Fig. 2). It should be noted, however, that the AAV-5 *trs* has been shown not to be cleaved by AAV-2 Rep68/78 (9). Six bases within S1-18 are conserved in AAV serotypes 1, 2, 3, 4, and 6 (Fig. 2). Of these conserved bases (G2, G5, G8, T9, T10, and G11), mutation of G2, T9, or T10 resulted in a $>30\%$ reduction in nicking by Rep78. This is also consistent with our hypothesis that a combination of base-specific and secondary structure elements is required for nicking. We also mutated bases G2 and G5 simultaneously and found no change in gel migration (Fig. 3B). However, the stairstep analysis (Fig. 4), suggested that G2 also contributes to the secondary structure. It is possible that the stairstep analysis can detect secondary structure components that mutation within the context of an 18-mer might miss.

Our data roughly agree with the results of a previous mutational analysis of the *AAVSI* core *trs* flanking sequence (G7 through G12) by Lamartina et al. (32). They detected a drop in nicking by Rep68 with mutation of either T within the GGT TGG motif, but only mutation of T9 resulted in a $>25\%$ drop in nicking. Our data also show that mutation of T9 has the greater impact (Table 3). Lamartina et al. found that simultaneous mutation of both T9 and T10 eliminated detectable nicking. They also simultaneously mutated G7, G8, G11, and G12, and it only resulted in a 25% drop in nicking efficiency as long as T9 and T10 were left intact (32). This result, at the time, was difficult to reconcile with the fact that a sequence from a *BRC1* candidate gene (GenBank accession no. U11292) that contained an RRS, two T's (eight bases from the RRS), and a CTC motif between the Ts and the RRS could not be detectably nicked by Rep68 (1, 32, 61). Taken together, these

FIG. 10. Mutational analysis of *AAVS1* 18-mer migration. End-labeled oligonucleotides (ca. 1 to 4 pmol) were run on either nondenaturing (A and B) or denaturing (C and D) 8% polyacrylamide gels and autoradiographed. See Tables 1 and 2 for sequences. The oligonucleotides marked Mark-18, Mark-12, and S1-18 (Gel) were purified on a denaturing polyacrylamide gel prior to labeling. All others were trityl column purified prior to labeling.

data suggested that a region outside the GGTTGG motif of *AAVS1* is important for nicking. Given the lack of conservation of many bases outside the GGTTGG motif that are important for nicking of *AAVS1* (namely, C3, G14, C15, T16, and G18)

between *AAVS1* and the AAV-2 ITR, we propose that a secondary structure is being recognized.

The results of experiments in which the GGTTGG motif of *AAVS1* was moved further away from its RRS must be reeval-

uated in light of this need for a context (32, 40). Such experiments show that nicking by Rep68/78 and AAV-2 integration are greatly reduced when the core *trs* motif is moved further away from the RRS (32, 40). These results had been considered to reflect the limits of Rep68/78's ability to cleave DNA at a distance from its primary binding site (32, 40). Although this is probably at least partly true (5), the removal of the core motif from its secondary structure context may also contribute to the loss of nicking and integration.

Brister and Muzyczka (6), in their mutational analysis of the AAV-2 ITR *trs* region, also identified multiple classes of mutations. Although most of their mutations predicted to disrupt their putative stem-loop resulted in a 19 to 91% reduction in nicking by Rep68, one such mutation actually resulted in a 66% increase in nicking. These results are consistent with our hypothesis that some, but not all components of the secondary structure are important for nicking.

Our work defines further the number of bases recognized at the *AAVSI* by Rep68/78. The size of the nicking site recognition motif is an important issue for determining if the DNA sequence at *AAVSI* is sufficient to define a unique locus within the human genome. A 33-bp segment of *AAVSI* that contains the RRS, CTC, and core *trs* motifs has been shown to be sufficient to target integration of the AAV-2 genome into an episome in human cells, although not as efficiently as a longer sequence (35). Mutation of the *AAVSI* RRS or *trs* core sequence (GGTTGG) abolishes targeted integration (35). Mutation of the CTC motif reduces targeted integration (40). Although the archetypal RRS contains four repeats of GCTC or its complement, GAGC, sequences with only one perfect GCTC (depending on context) can be bound by Rep68/78 (31, 39, 61, 62). Even the presence of such a cryptic RRS can allow the unwinding of a blunt-ended, double-stranded DNA by Rep68 (48). High-throughput screening of randomly generated, double-stranded oligonucleotides suggests that even some sequences with no perfect GCTC boxes can bind Rep68/78 (13). Given the degenerate nature of RRSs, it has been estimated that as many as 2×10^5 RRSs may exist within the human genome (66). It is therefore unlikely that the three previously recognized components of the *AAVSI* core targeting sequence are sufficient to define a unique locus within the human genome, especially given that variation seems to be allowed in the spacing between the three elements (Fig. 1).

Lamartina et al. reported the existence of a DNase I-hypersensitive site near the *AAVSI* *trs* homologue and suggested that the chromatin configuration in this region may contribute to integration (33). The same authors also demonstrated that supercoiled circular DNA is nicked by Rep68 more efficiently than linear DNA containing the same sequence (32), suggesting that the topological state of the DNA may be important for integration. However, the ability to move human *AAVSI* sequences to either a human episome or the mouse or rat genome and still get targeted integration of AAV-2 suggests that either the integration signals within the DNA sequence are sufficient for targeting or the DNA sequence drives the structure of the chromatin environment (2, 16, 17, 33, 35, 40, 43, 66).

It is also important to note that the 33-bp minimal target of Linden et al. was truncated at base G4 of our *AAVSI* 18-mer and is approximately twofold less efficient at targeting AAV-2 integration than a 62-bp segment that includes our entire 18-

base sequence, along with the CTC motif, the RRS, and additional upstream sequences (35). By an accident of cloning, base G2 was preserved, but bases G1 and C3 were altered in their 33-base insert (35). These data are consistent with our observation that mutation of C3 resulted in a 56% decrease in nicking (Table 2). Their 33-mer did, however, contain the entire region of our 12-mer oligonucleotide (S1-12) that also showed abnormal migration and thus is assumed to be able to form some elements of the secondary structure around the *trs* (Fig. 4). An additional requirement for a secondary structure, or at least an extended *trs* context, may be sufficient to make the *AAVSI* core integration target unique in the human genome.

Approximately 70 to 90% of AAV-2 integration events occur at the *AAVSI* locus (25, 30, 46). It is not clear whether the integration events at other loci are distributed randomly. Given the hypothesis that a good Rep68/78 nicking site is a preferred AAV-2 integration locus, it is important to define the variety of sequences that can be nicked by Rep68/78.

ACKNOWLEDGMENTS

We thank Karen Usdin and J. Rodney Brister for critical reading of the manuscript, as well as other useful advice. We also thank Victor Zhurkin, Kagnew Gebreyesus, and Ad Bax for useful advice.

R.A.O. is a coinventor on a patent related to recombinant AAV technology. To the extent that the work in the present study increases the value of this patent, he has a conflict of interest.

REFERENCES

1. Albertsen, H. M., S. A. Smith, S. Mazoyer, E. Fujimoto, J. Stevens, B. Williams, P. Rodriguez, C. S. Cropp, P. Slijepcevic, M. Carlson, M. Robertson, P. Bradley, E. Lawrence, T. Harrington, Z. M. Sheng, R. Hoopes, N. Sternberg, A. Brothman, R. Callahan, B. A. J. Ponder, and R. White. 1994. A physical map and candidate genes in the *BRCA1* region on chromosome 17q12-21. *Nat. Genet.* 7:472-479.
2. Bakowska, J. C., M. V. Di Maria, S. M. Camp, Y. Wang, P. D. Allen, and X. O. Breakfield. 2003. Targeted transgene integration into transgenic mouse fibroblasts carrying the full-length human *AAVSI* locus mediated by HSV/AAV rep⁺ hybrid amplicon vector. *Gene Ther.* 10:1691-1702.
3. Balagué, C., M. Kalla, and W. W. Zhang. 1997. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J. Virol.* 71:3299-3306.
4. Balagurumoorthy, P., and S. K. Brahmachari. 1994. Structure and stability of human telomeric sequence. *J. Biol. Chem.* 269:21858-21869.
5. Brister, J. R., and N. Muzyczka. 2000. Mechanism of Rep-mediated adeno-associated virus origin nicking. *J. Virol.* 74:7762-7771.
6. Brister, J. R., and N. Muzyczka. 1999. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. *J. Virol.* 73:9325-9336.
7. Carter, B. J. 1990. Adeno-associated virus helper functions, p. 255-282. In P. Tijssen (ed.), *Handbook of parvoviruses*, vol. 1. CRC Press, Inc., Boca Raton, Fla.
8. Chen, X., S. V. Mariappan, P. Catasti, R. Ratliff, R. K. Moyzis, A. Laayoun, S. S. Smith, E. M. Bradbury, and G. Gupta. 1995. Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* 92:5199-5203.
9. Chiorini, J. A., S. Afione, and R. M. Kotin. 1999. Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. *J. Virol.* 73:4293-4298.
10. Chiorini, J. A., F. Kim, L. Yang, and R. M. Kotin. 1999. Cloning and characterization of adeno-associated virus type 5. *J. Virol.* 73:1309-1319.
11. Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urclay, B. Safer, and R. M. Kotin. 1994. Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in *Escherichia coli*. *J. Virol.* 68:797-804.
12. Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. M. Kyöstiö, R. M. Kotin, and B. Safer. 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. *J. Virol.* 68:7448-7457.
13. Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin. 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. *J. Virol.* 69:7334-7338.
14. Davis, M. D., J. Wu, and R. A. Owens. 2000. Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates. *J. Virol.* 74:2936-2942.

15. Felsenfeld, G., and H. T. Miles. 1967. The physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.* **36**:407–448.
16. Giraud, C., E. Winocour, and K. I. Berns. 1995. Recombinant junctions formed by site-specific integration of adeno-associated virus into an episome. *J. Virol.* **69**:6917–6924.
17. Giraud, C., E. Winocour, and K. I. Berns. 1994. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc. Natl. Acad. Sci. USA* **91**:10039–10043.
18. Gratzer, W. B., L. R. Hill, and R. J. Owen. 1970. Circular dichroism of DNA. *Eur. J. Biochem.* **15**:209–214.
19. Hardin, C. C., T. Watson, M. Corregan, and C. Bailey. 1992. Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG₃GCG). *Biochemistry* **31**:833–841.
20. Hickman, A. B., D. R. Ronning, Z. N. Perez, R. M. Kotin, and F. Dyda. 2004. The nuclease domain of adeno-associated virus Rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell* **13**:403–414.
21. Howell, R. M., K. J. Woodford, M. N. Weitzmann, and K. Usdin. 1996. The chicken beta-globin gene promoter forms a novel “cinched” tetrahelical structure. *J. Biol. Chem.* **271**:5208–5214.
22. Im, D. S., and N. Muzyczka. 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell* **61**:447–457.
23. Im, D. S., and N. Muzyczka. 1992. Partial purification of adeno-associated virus Rep78, Rep52, and Rep40 and their biochemical characterization. *J. Virol.* **66**:1119–1128.
24. Johnson, W. C. 2000. CD of nucleic acids, p. 703–718. *In* N. Berova, K. Nakanishi, and R. W. Woody (ed.), *Circular dichroism principles and applications*, 2nd ed. Wiley-VCH, New York, N.Y.
25. Kearns, W. G., S. A. Afione, S. B. Fulmer, M. G. Pang, D. Erikson, M. Egan, M. J. Landrum, T. R. Flotte, and G. R. Cutting. 1996. Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther.* **3**:748–755.
26. Kettani, A., S. Bouaziz, A. Gorin, H. Zhao, R. A. Jones, and D. J. Patel. 1998. Solution structure of a Na cation stabilized DNA quadruplex containing G · G · G · G and G · C · G · C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* **282**:619–636.
27. Kotin, R. M., and K. I. Berns. 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. *Virology* **170**:460–467.
28. Kotin, R. M., R. M. Linden, and K. I. Berns. 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11**:5071–5078.
29. Kotin, R. M., J. C. Menninger, D. C. Ward, and K. I. Berns. 1991. Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. *Genomics* **10**:831–834.
30. Kotin, R. M., M. Siniscalco, R. J. Samulski, X. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns. 1990. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. USA* **87**:2211–2215.
31. Kyöstö, S. R. M., R. S. Wonderling, and R. A. Owens. 1995. Negative regulation of the adeno-associated virus (AAV) P₅ promoter involves both the P₅ Rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. *J. Virol.* **69**:6787–6796.
32. Lamartina, S., G. Ciliberto, and C. Toniatti. 2000. Selective cleavage of AAVS1 substrates by the adeno-associated virus type 2 Rep68 protein is dependent on topological and sequence constraints. *J. Virol.* **74**:8831–8842.
33. Lamartina, S., E. Sporeno, E. Fattori, and C. Toniatti. 2000. Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment. *J. Virol.* **74**:7671–7677.
34. Li, Z., J. R. Brister, D. S. Im, and N. Muzyczka. 2003. Characterization of the adeno-associated virus Rep protein complex formed on the viral origin of DNA replication. *Virology* **313**:364–376.
35. Linden, R. M., E. Winocour, and K. I. Berns. 1996. The recombination signals for adeno-associated virus site-specific integration. *Proc. Natl. Acad. Sci. USA* **93**:7966–7972.
36. Lu, M., Q. Guo, and N. R. Kallenbach. 1993. Thermodynamics of G-tetraplex formation by telomeric DNAs. *Biochemistry* **32**:598–601.
37. Maurizot, J. C. 2000. Circular dichroism of nucleic acids; nonclassical conformations and modified oligonucleotides, p. 719–739. *In* N. Berova, K. Nakanishi, and R. W. Woody (ed.), *Circular dichroism principles and applications*, 2nd ed. Wiley-VCH, New York, N.Y.
38. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
39. McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka. 1994. Identification of linear AAV DNA sequences that specifically bind the adeno-associated virus Rep protein. *J. Virol.* **68**:4988–4997.
40. Meneses, P., K. I. Berns, and E. Winocour. 2000. DNA sequence motifs which direct adeno-associated virus site-specific integration in a model system. *J. Virol.* **74**:6213–6216.
41. Owens, R. A., M. D. Weitzman, S. R. M. Kyöstö, and B. J. Carter. 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. *J. Virol.* **67**:997–1005.
42. Philpott, N. J., J. Gomos, K. I. Berns, and E. Falck-Pedersen. 2002. A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at chromosome 19. *Proc. Natl. Acad. Sci. USA* **99**:12381–12385.
43. Rizzuto, G., B. Gorgoni, M. Cappelletti, D. Lazzaro, I. Gloaguen, V. Poli, A. Sgura, D. Cimini, G. Ciliberto, R. Cortese, E. Fattori, and N. La Monica. 1999. Development of animal models for adeno-associated virus site-specific integration. *J. Virol.* **73**:2517–2526.
44. Ryan, J. H., S. Zolotukhin, and N. Muzyczka. 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. *J. Virol.* **70**:1542–1553.
45. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed., vol. 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
46. Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter. 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10**:3941–3950.
47. Shelling, A. N., and M. G. Smith. 1994. Targeted integration of transfected and infected adeno-associated virus vectors containing the neomycin resistance gene. *Gene Ther.* **1**:165–169.
48. Smith, D. H., P. Ward, and R. M. Linden. 1999. Comparative characterization of Rep proteins from the helper-dependent adeno-associated virus type 2 and the autonomous goose parvovirus. *J. Virol.* **73**:2930–2937.
49. Smith, R. H., and R. M. Kotin. 2000. An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV *ori* DNA. *J. Virol.* **74**:3122–3129.
50. Snyder, R. O., D. S. Im, and N. Muzyczka. 1990. Evidence for covalent attachment of the adeno-associated virus (AAV) Rep protein to the ends of the AAV genome. *J. Virol.* **64**:6204–6213.
51. Snyder, R. O., D. S. Im, T. Ni, X. Xiao, R. J. Samulski, and N. Muzyczka. 1993. Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. *J. Virol.* **67**:6096–6104.
52. Snyder, R. O., R. J. Samulski, and N. Muzyczka. 1990. In vitro resolution of covalently joined AAV chromosome ends. *Cell* **60**:105–113.
53. Spear, I. S., K. H. Fife, W. W. Hauswirth, C. J. Jones, and K. I. Berns. 1977. Evidence for two nucleotide sequence orientations within the terminal repetition of adeno-associated virus DNA. *J. Virol.* **24**:627–634.
54. Srivastava, A., E. W. Lusby, and K. I. Berns. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* **45**:555–564.
55. Steigerwald, R., C. Rabe, V. Schmitz, I. G. Schmidt-Wolf, M. Alt, and W. H. Caselmann. 2003. Requirements for adeno-associated virus-derived non-viral vectors to achieve stable and site-specific integration of plasmid DNA in liver carcinoma cells. *Digestion* **68**:13–23.
56. Urabe, M., Y. Hasumi, A. Kume, R. T. Surosky, G. J. Kurtzman, K. Tobita, and K. Ozawa. 1999. Charged-to-alanine scanning mutagenesis of the N-terminal half of adeno-associated virus type 2 Rep78 protein. *J. Virol.* **73**:2682–2693.
57. Urcelay, E., P. Ward, S. M. Wiener, B. Safer, and R. M. Kotin. 1995. Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. *J. Virol.* **69**:2038–2046.
58. Watson, J. D., and F. H. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**:737–738.
59. Weitzman, M. D., S. R. M. Kyöstö, R. M. Kotin, and R. A. Owens. 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. USA* **91**:5808–5812.
60. Wen, J. D., and D. M. Gray. 2002. The Ff gene 5 single-stranded DNA-binding protein binds to the transiently folded form of an intramolecular G-quadruplex. *Biochemistry* **41**:11438–11448.
61. Wonderling, R. S., and R. A. Owens. 1997. Binding sites for adeno-associated virus Rep proteins within the human genome. *J. Virol.* **71**:2528–2534.
62. Wonderling, R. S., and R. A. Owens. 1996. The Rep68 protein of adeno-associated virus type 2 stimulates expression of the platelet-derived growth factor B c-sis proto-oncogene. *J. Virol.* **70**:4783–4786. (Author's correction, 70:9084.)
63. Wu, J., M. D. Davis, and R. A. Owens. 1999. Factors affecting the terminal resolution site endonuclease, helicase, and ATPase activities of adeno-associated virus type 2 Rep proteins. *J. Virol.* **73**:8235–8244.
64. Wu, J., M. D. Davis, and R. A. Owens. 2001. A Rep recognition sequence is necessary but not sufficient for nicking of DNA by adeno-associated virus type-2 Rep proteins. *Arch. Biochem. Biophys.* **389**:271–277.
65. Xiao, W., N. Chirmule, S. C. Berta, B. McCullough, G. Gao, and J. M. Wilson. 1999. Gene therapy vectors based on adeno-associated virus type 1. *J. Virol.* **73**:3994–4003.
66. Young, S. M., D. M. McCarty, N. Degtyareva, and R. J. Samulski. 2000. Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination. *J. Virol.* **74**:3953–3966.
67. Zar, J. H. 1974. Paired-sample hypotheses, p. 121–129. *Biostatistical analysis*. Prentice-Hall, Inc., Englewood Cliffs, N.J.
68. Zhou, X., I. Zolotukhin, D. S. Im, and N. Muzyczka. 1999. Biochemical characterization of adeno-associated virus Rep68 DNA helicase and ATPase activities. *J. Virol.* **73**:1580–1590.