JOURNAL OF VIROLOGY, Mar. 1996, p. 1542–1553 Vol. 70, No. 3
0022-538X/96/$04.00+0

Sequence Requirements for Binding of Rep68 to the Adeno-Associated Virus Terminal Repeats

JOHN H. RYAN, SERGEI ZOLOTUKHIN, AND NICHOLAS MUZYCZKA*

Department of Molecular Genetics and Microbiology, College of Medicine, University of Florida, Gainesville, Florida 32610

Received 26 October 1995/Accepted 29 November 1995

We have used reciprocal competition binding experiments with mutant substrates and chemical modification interference assays to precisely define the sequences within the adeno-associated virus (AAV) terminal repeat (TR) that are involved in site-specific binding to the AAV Rep protein. Mutagenesis experiments were done with a 43-bp oligonucleotide which contained the Rep binding element (RBE) within the A stem of the TR. Experiments in which two adjacent base pairs of the RBE were substituted simultaneously with nucleotides that produced transversions identified a 22-bp sequence (CAGTGAGCGAGCGAGCGCGCAG) in which substitutions measurably affected the binding affinity. Although the 22-bp RBE contains the GAGC motifs that have been found in all known Rep binding sites, our results suggest that the GAGC motifs alone are not the only sequences specifically recognized by Rep. The effects of substitutions within the 22-bp sequence were relatively symmetrical, with nucleotides at the periphery of the RBE having the least effect on binding affinity and those in the middle having the greatest effect. Dinucleotide mutations within 18 (GTGAGCGAGCGAGC GCGC) of the 22 bp were found to decrease the binding affinity by at least threefold. Dinucleotide mutations within a 10-bp core sequence (GCGAGCGAGC) were found to decrease binding affinity by more than 10-fold. Single-base substitutions within the 10-bp core sequence lowered the binding affinity by variable amounts (up to fivefold). The results of the mutagenesis analysis suggested that the A-stem RBE contains only a single Rep binding site rather than two or more independent sites. To confirm the results of the mutant analysis and to determine the relative contribution of each base to binding, chemical modification experiments using dimethyl sulfate and hydrazine were performed on both the linear A-stem sequence and the entire AAV TR in both the flip and flop hairpinned configurations. Interference assays on the linear A stem identified the 18-bp sequence described above as essential for binding. G, C, and T residues on both strands contributed to binding, and the interference pattern correlated well with the results of the mutagenesis experiments. Interference assays with complete hairpinned TR substrates also identified the 18-bp sequence as important for binding. However, the interference patterns on the two strands within the RBE and the relative contributions of the individual bases to binding were clearly different between the hairpinned substrates and the linear A-stem binding element. Interference assays also allowed us to search for residues within the small internal palindromes of the TR (B and C) that contribute to binding. The largest effect was seen by modification of two T residues within the sequence CTTTG. This sequence was present in the same position relative to the terminal resolution site (*trs*) in both the flip and flop orientations of the TR. In addition, the interference pattern suggested that the remaining bases within the CTTTG motif as well as other bases within the B and C palindromes make contacts with the Rep protein, albeit with lower affinities. Regardless of whether the TR was in the flip or flop orientation, most of the contact points were clustered in the small internal palindrome furthest away from the *trs*. We also determined the relative binding affinity of linear substrates containing a complete RBE with hairpinned substrates and found that linear substrates bound Rep less efficiently. Our results were consistent with our previous model that there are three distinct elements within the hairpinned AAV TR that contribute to binding affinity or to efficient nicking at the *trs*: the A-stem RBE, the secondary structure element which consists of the B and C palindromes, and the *trs*. The identification of the CTTTG motif within the B and C palindromes suggested that the interaction of Rep with the secondary structure element is both sequence and structure dependent. In addition, the interaction of the Rep protein with the invariant CTTTG motif suggested a mechanism by which the Rep protein discriminates between hairpinned and linear AAV termini during the terminal resolution process so that it preferentially nicks the hairpinned substrate.

The relatively small genome of adeno-associated virus (AAV) codes for four nonstructural polypeptides referred to by their apparent molecular sizes of 78, 68, 52, and 40 kDa (27, 33). These four related overlapping proteins are encoded by the viral *rep* gene and contain multiple, sometimes redundant functions necessary for the propagation of the virus. The two smaller proteins, Rep52 and Rep40, are thought to be involved in the accumulation of single-stranded AAV DNA and virus packaging (8) but are not required for the accumulation of replicative intermediates (8, 25, 30). In addition, Rep52 has some ability to repress heterologous promoters (16). The two larger Rep proteins, Rep78 and Rep68, are essential in *trans* for AAV DNA replication (12), for transactivation of the AAV promoters (17, 20, 34), and for the repression of viral and heterologous promoters (3, 11, 15, 16, 28, 34).

Our previous biochemical studies of Rep68 and Rep78 have shown that these two proteins are ATP-dependent site-specific and strand-specific endonucleases (14) that preferentially bind

* Corresponding author. Mailing address: Department of Molecular Genetics and Microbiology, College of Medicine, University of Florida, P.O. Box 100266 JHMHSC, Gainesville, FL 32610. Phone: (904) 392-8541. Fax: (904) 392-3133. Electronic mail address: muzyczka@medmicro.med.ufl.edu.

and cut the terminal resolution site (*trs*) in hairpinned AAV terminal repeats (TRs) during the process of terminal resolution (14, 31, 32). During this process, the Rep protein is covalently attached to the 5′ end of the cut site via a tyrosine residue (30). The enzyme also has an intrinsic DNA helicase activity which may participate in unwinding the TR during terminal resolution or in initiating the synthesis of new progeny strands by strand displacement synthesis (14). In vitro replication studies have shown that either Rep68 or Rep78 is capable of supporting AAV DNA replication (25). Finally, a recent report by Wonderling et al. (36) demonstrates that the Rep protein is capable of unwinding RNA-DNA hybrids.

To accomplish its role in DNA replication, the Rep enzyme must be capable of discriminating between hairpinned and linear AAV termini and preferentially cutting the hairpinned substrate. It must also be capable of processing linear dimer intermediates. Furthermore, the enzyme must be oriented on the TR in such a way that the correct strand is cut at the *trs*. To determine how this occurs, we have been mapping the essential recognition elements within the TR for binding and *trs* endonuclease activity. Our analysis suggested that there were at least three elements of the AAV TR that were important for Rep function at the TR (Fig. 1a): the sequence at the *trs*, the secondary structure element composed of the B and C palindromes, and a linear Rep-binding element (RBE) proximal to the B and C palindromes within the A stem of the TR (22, 31). Using homogeneously pure Rep68 and partially purified Rep78, we identified the linear binding element within the A stem as an approximately 25-bp sequence which could bind Rep protein in the absence of the B and C palindromes and the *trs* (22) (Fig. 1a). Degenerate RBEs were also found in pBR322 and in the AAV p5 and p19 promoters (22).

Comparison of these sequences suggested that a repeating GAGC motif contained within the 25-bp A-stem sequence was important for recognition (22). This possibility was supported by the fact that mutations within the GAGC motifs eliminated binding to the linear A-stem sequence and reduced binding to the complete hairpinned TR (23). In addition, the fact that mutagenesis of some of the GAGC motifs did not change the pattern or number of bound Rep species suggested that the A-stem sequence contained only one Rep binding site. In contrast, mutations within the *trs* did not appear to significantly affect Rep binding to the complete hairpinned TR or to substrates that were missing the B and C palindromes (Fig. 1a) (23, 31). The importance of the GAGC motifs also was supported by the fact that Owens and colleagues had earlier mapped four G residues (Fig. 1a) within the GAGC repeat as important for Rep binding to the hairpin TR by methylation interference assays (26). Nevertheless, the p5 promoter binding site contained only one perfect GAGC motif, suggesting that additional sequences in the A-stem binding element might be important for binding (22). Moreover, binding to a fragment truncated at the *Dde*I site in the A stem (Fig. 1a), which contained most of the 25-bp binding element and all of the GAGC motifs as well as the *trs*, was up to 125-fold less efficient in binding to Rep than the complete hairpin TR (23). This finding confirmed earlier comparisons between the complete duplex, linear TR and the hairpinned TR by us and others (1, 13) and suggested that a portion of the B-C secondary structure element contributed to the binding affinity for the AAV TR. We suggested that this difference in binding affinity was at least in part responsible for the difference in site-specific nicking activity at the *trs* that we had seen between hairpinned substrates containing the B and C palindromes and linear substrates containing only the A-stem sequences and the *trs* (22, 31).

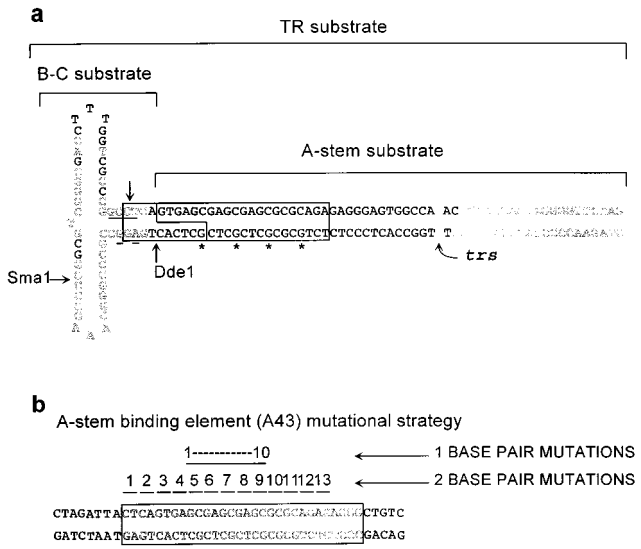

FIG. 1. Mutational strategy for defining the RBE. (a) Sequence of the AAV terminal repeat in the flip hairpin configuration. Boldface letters indicate bases that were protected from DNase I (13) or copper phenanthralene (14a). In the region of the *Dde*I site (vertical arrows), the underlined bases were not protected from DNase I (13). Asterisks indicate the G residues identified by Owens et al. (26) and Weitzman et al. (35) by methylation interference assay as necessary for Rep binding. The A-stem and B-C substrates were used previously by us (22, 23) to define sequences required for Rep binding. The bent line indicates the overlap region in these two substrates that was used to ligate them to make the complete hairpin TR substrate used in our previous studies. The boxed sequence indicates the 25-bp sequence suggested by us to contain most if not all of the linear A-stem RBE (23, 31). It includes the GAGC/GCTG repeats suggested by us (23) and by Chiorini et al. (9, 10) and Weitzman et al. (35) as necessary for Rep binding. The sequence within the box plus the additional 3 bp of A-stem sequence to the left of the box constitutes the 28-bp sequence suggested by Chiorini et al. (10) to be the minimal sequence that contains all of the bases required for specific base contacts by Rep protein. For maximal binding affinity, Chiorini et al. (10) found that the oligonucleotide containing the 28-bp sequence had to be extended with heterologous sequence on the right to a total length of 46 bp. (b) Sequence of the 43-bp synthetic oligonucleotide substrate (A43) used in this study for competition gel shift assays to define the sequence of the A-stem RBE essential for binding. Boxed region indicates the portion of the oligonucleotide that contains a sequence identical to the wild-type A-stem sequence. Flanking boldface nucleotides outside the box are heterologous sequences. A series of oligonucleotide A43 substrates that contained either 2-bp transversions (mutants 1 to 13) or single-base transversions (mutants 1 to 10) was synthesized.

Weitzman et al. (35), independently, also identified the linear RBE in their studies of the human chromosome 19 target sequence for AAV DNA integration and suggested that the GAGC (or GCTC) repeat was necessary for binding. Mutations within the G residues mapped by Owens et al. (26) eliminated binding to the chromosome 19 target site (35). However, the GAGC repeats themselves were not sufficient for binding, suggesting that additional flanking sequences were necessary. Chiorini et al. (9, 10) reported similar binding activity for a mutant Rep68 fusion protein that contained the maltose binding domain. Again, 18-bp oligonucleotides that contained the tetrameric GAGC repeats bound Rep poorly. This group concluded that a 28-bp sequence within the A stem of the TR that included the imperfect GAGC tetrameric repeat was essential for binding but that high-affinity binding also required flanking DNA of random sequence to stabilize binding by the Rep protein (10) (Fig. 1a). In addition, Chiorini et al. (10) confirmed our reports (22, 31) that the Rep protein cut the *trs* at significantly higher frequencies in the context of the hairpinned TR than in a linear substrate and that sequences in the vicinity of the *trs* do not affect binding. However, they saw no difference in binding between the complete hairpinned TR

and linear substrates that contained only the A-stem sequences and the *trs*. Chiorini et al. (10) concluded that the B-C portion of the hairpin made no contribution to binding affinity but stimulated the *trs* endonuclease activity. They suggested that earlier reports of lower-affinity binding by A-stem substrates were due either to the absence of a complete binding site or to the lack of sufficient nonspecific flanking DNA (10).

The A-stem RBE is clearly central to the function of the large Rep proteins. Its presence in the AAV TR (9, 10, 22, 23, 31), the p5 and p19 promoters (22), heterologous promoters (2), and the chromosome 19 integration site (35) suggests that the RBE is involved in viral DNA replication, transcription, and proviral integration. In this study, we use two different approaches, competition binding experiments with mutated A-stem substrates and chemical modification interference assays, to define in detail the sequence within the A stem that is necessary for sequence-specific binding by Rep68. In addition, we use chemical modification interference assays to identify sequences within the B and C palindromes that are necessary for Rep binding. The most important sequence appears to be a CTTTG motif that occurs in the same position with respect to the *trs* in both the flip and flop orientations of the TR. We suggest that this motif and possibly other contact points found within the B and C palindromes probably explain the difference in binding affinity and most if not all of the difference in nicking activity that we see between the hairpinned and linear TR substrates. In addition, the A-stem RBE and the CTTTG motif together could explain how the Rep protein is oriented on the TR with respect to the *trs*. Finally, our results are consistent with our previous model of a tripartite origin for AAV DNA replication that consists of the A-stem RBE, the B and C palindromes, and the *trs*.

MATERIALS AND METHODS

Rep protein purification. Rep68 was extracted from recombinant baculovirus-infected SF9 cells and purified by chromatography as described previously (25). Alternatively, Rep68 was purified by affinity chromatography as described previously for Rep78 (22). The protein concentration was determined to be 0.3 or 0.05 mg/ml, respectively, with the Bradford reagent (Bio-Rad), using gamma globulin as the standard. Both types of preparations were homogeneously pure, as judged by silver staining after sodium dodecyl sulfate-acrylamide gel electrophoresis. Binding activity was found to be stable when the enzyme was stored at −80°C. To enhance the binding activity, the enzyme was treated, upon thawing, with Tween 20 (polyoxyethylenesorbitan monolaurate). This was done by mixing the enzyme with an equal volume of 1% Tween 20–0.03 M NaCl–20% glycerol–50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid)-NaOH (pH 7.9)–2.5 mM dithiothreitol–1.5 mM $MgCl_2$ and incubating the mixture for 2 h on ice prior to the binding assay.

DNA substrates. (i) A43 substrate. Synthetic oligonucleotides were purified and annealed as described previously (23). The substrates built from the A43 oligonucleotides are shown in Fig. 1b. For each substrate, the concentrations of the gel-purified A-stem strands were determined by A_{260}, and then equal amounts of the two strands were annealed. An aliquot of each double-stranded annealed product was labeled at its 5′ ends with [γ-^{32}P]ATP and T4 polynucleotide kinase and electrophoresed on a polyacrylamide gel to determine the extent of annealing. The remainder of the annealed product was then used as the unlabeled competitor. The labeled wild-type substrate was made by phosphorylating the 5′ ends of a known quantity of the double-stranded wild-type A43 substrate with [γ-^{32}P]ATP and T4 polynucleotide kinase.

(ii) A-D substrate. The A-D substrate for the chemical modification interference assay was synthesized by PCR, using the *Xba*I-*Bgl*II fragment from plasmid pTR_{BS}-3D (Fig. 2a). This fragment consists of an AAV TR in the flop orientation flanked by two D-sequence elements and was designated the 2D linear substrate. The TR sequences in the 2D linear substrate are identical to the TR sequences that are present at the internal position of an AAV dimer replicative-form DNA molecule. Two primers were used for PCR synthesis of the A-D substrate: 5′-ATATCTTTGCCCGGGCG-3′ and 5′-ATGGCCCACAACCAAG ATCT-3′. The first primer was partially complementary to the C palindrome of the TR; the second one was complementary to the plasmid sequence next to the D repeat (Fig. 2a). The PCR was run for 25 cycles of denaturing at 94°C for 1 min, annealing at 59°C for 1 min, and extension at 72°C for 1 min, using *Taq* polymerase. The 98-bp PCR product included the C′/C palindrome and all of the A and D sequences of the AAV TR. The bottom strand of the PCR-generated

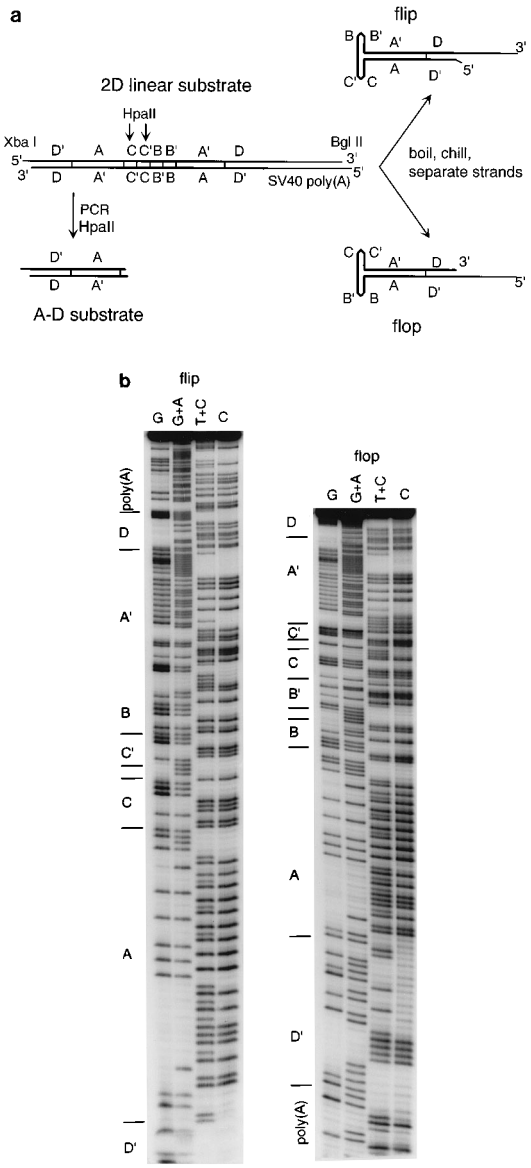

FIG. 2. Hairpinned and linear substrates for competitive Rep binding assays. (a) Diagram for the synthesis of the linear substrate that contains the A and D sequences as well as 5 to 7 bp of the C palindrome (A-D substrate) and for the synthesis of the hairpinned flip and flop TR substrates. In both cases, the starting substrate is the *Xba*I-*Bgl*II fragment from plasmid pTR_{BS}-3D (2D linear substrate). See Materials and Methods for details. (b) The flip and flop substrates isolated by acrylamide gel electrophoresis were subjected to DNA sequence analysis by chemical modification to confirm that each substrate was exclusively flip or flop. This was done to exclude the possibility that either substrate was contaminated with the starting linear *Xba*I-*Bgl*II fragment from plasmid pTR_{BS}-3D. See Materials and Methods for details.

substrate was labeled after digestion of the PCR product with *Hpa*II (a site within the C′/C palindrome) by filling in the 3′ end with the Klenow fragment of DNA polymerase I, using [α-^{32}P]dCTP. The top strand was labeled with dGTP, [α-^{32}P]dATP, and Klenow enzyme after digestion with *Bgl*II. The final substrate included seven residues of the C/C′ palindrome attached to a complete A-D stem and was called the A-D substrate (Fig. 2a). The A-D substrate for the competition binding assay was made in the same way except that the PCR product was cut with *Sma*I instead of *Hpa*II. This version of the A-D substrate contained only five residues of the C/C′ palindrome.

(iii) Flip and flop hairpinned substrates. The *Xba*I-*Bgl*II fragment of pTR_{BS}-3D (Fig. 2a) was labeled by Klenow DNA polymerase in the presence of [α-^{32}P]dATP (to label flip) or [α-^{32}P]dCTP (to label flop). The labeled fragments were then denatured by boiling, quickly chilled, and separated on a 4% nonde-

naturing acrylamide gel in 1× Tris-borate-EDTA at 7.5 V/cm for 8 h. Under these conditions, the three possible products migrated in the following order: flip, flop, renatured double-stranded linear. Following electrophoresis, the flip and flop bands were eluted from the gel.

To ensure that equal molar amounts of unlabeled competitor DNA were used in the competition assay between flip and flop hairpin, 2D linear, A43, and A-D substrates, the concentration of each substrate was determined by A_{260}. The concentrations were confirmed by comparing the intensities of ethidium bromide staining of aliquots of each substrate. Finally, an aliquot of each substrate was 3′ labeled with [α-^{32}P]dCTP and Klenow polymerase to determine whether equimolar amounts of substrates contained equivalent numbers of 3′ ends. The latter test was omitted for the A43 substrate.

Plasmids. Plasmid pTR_{BR} was constructed as follows. The 1,270-bp adenovirus type 5 *Bgl*II fragment was inserted into the unique *Bgl*II site of the previously described vector *dl*3-94 (24). The resulting plasmid was then partially digested with *Bgl*II and completely digested with *Pst*I, and a 1,440-bp fragment which contained the left AAV TR and the adenovirus stuffer fragment was isolated. In a separate series of reactions, a 50-bp synthetic fragment coding for the simian virus 40 (SV40) early polyadenylation signal and containing *Bam*HI- and *Bgl*II-compatible ends was ligated to *dl*3-94 *Bgl*II linear DNA. The ligation mixture was then cut with *Pst*I to isolate the 225-bp fragment which consisted of the left TR from *dl*3-94 joined through a *Bgl*II-*Bam*HI junction to the synthetic poly(A) fragment. The 225-bp fragment and the 1,440-bp fragment were then ligated to each other, and a 1,640-bp fragment was isolated from the ligation mixture after it was digested with *Pst*I. The resulting 1,640-bp fragment was then inserted into the *Pst*I site of pBR322 and called pTR_{BR}.

To construct pTR_{BS}-3D, 293 cells were cotransfected with pTR_{BR} and pIM45 (20) and infected with adenovirus type 5 *ts*149 at a multiplicity of infection of 10. At 36 h posttransfection, Hirt DNA was isolated and digested with *Dpn*I. A DNA adapter that consisted of the AAV TR sequence from the *trs* to the end of the D sequence (nucleotides 125 to 145) and an *Xba*I sticky end was synthesized. The adapter was ligated to the Hirt DNA. The products of the ligation reaction were separated on a 1% agarose gel, and the DNA band corresponding to the monomer duplex form was eluted and purified. The 5′ ends of this fragment were kinase treated, and the DNA was cloned into the *Xba*I site of plasmid pBS(+) (Stratagene). Several independent clones were sequenced; one was found that contained the right AAV TR from pTR_{BR} joined to the synthetic D sequence on one side and the SV40 poly(A) site on the other and contained an intact *trs* at the ligation junction. This plasmid was p418wt. The *Bgl*II-*Pst*I partial digest fragment from pTR_{BR} which contains the left AAV TR and the adenovirus stuffer fragment was then ligated to the *Bgl*II-*Xba*I fragment from p418wt which contains the SV40 poly(A) site and the 2D TR sequence. The ligation products were digested with *Pst*I and *Xba*I and cloned into the polylinker site of plasmid pBS(+) that had been digested with *Pst*I and *Xba*I. The resulting plasmid, pTR_{BS}-3D, contains, from left to right: the left AAV TR, the adenovirus stuffer fragment, the SV40 poly(A) site, and the right AAV TR joined to an additional D sequence. The *Xba*I-*Bgl*II fragment containing the right AAV TR is illustrated in Fig. 2a.

EMS competition assay. The electrophoretic mobility shift (EMS) assay was performed as described earlier (23). Binding conditions were chosen so that 10% or less of the starting substrate was bound. To determine the ratio of dissociation constants (K_ds) of the wild-type and mutant A43 substrates, we used a competition binding analysis at high substrate concentrations (Fig. 3). For example, the EMS assay shown in Fig. 3A illustrates the binding of 0.75 pmol of Rep68 to 0.25 pmol of labeled wild-type A43 substrate in the absence of competitor or in the presence of either unlabeled wild-type A43 substrate or unlabeled mutant A43 substrate. The bound and free DNA fractions in each experiment were measured by excising gel slices of each lane that encompassed all of the protein-DNA complexes and the free species, respectively, and then counting each gel slice with a scintillation counter. The fraction of bound wild-type A43 substrate was then plotted as a function of picomoles of cold competitor (Fig. 3B). The ratio of the K_ds of the wild-type and mutant RBEs was calculated by dividing the amount of homologous wild-type competitor by the amount of heterologous mutant competitor that was required to reduce the fraction of labeled wild-type A43 substrate bound to 50% of the starting value in the absence of any competitor.

Chemical modification interference assay. Chemical modifications of the A-D substrate and the flip and flop hairpinned substrates were performed as described previously (7), with slight modifications. For the G reaction, about 10^7 cpm of labeled A-D substrate was dissolved in 10 μl of Tris-EDTA (TE) buffer and diluted with 200 μl of dimethyl sulfate (DMS) reaction buffer (50 mM sodium cacodylate [pH 8.0]–1 mM EDTA [pH 8.0]). After addition of 1 μl of DMS, the reaction was allowed to proceed at room temperature for 5 min and then quenched by adding 40 μl of 1.5 M sodium acetate (pH 7.0), 1 mM 2-mercaptoethanol, 1 μl of tRNA (10 mg/ml), and 600 μl of 100% ethanol. The DNA was precipitated twice with ethanol, rinsed, dried, and dissolved in TE. For the C+T reaction, the same amount of DNA substrate dissolved in 25 μl of H_2O was mixed with 15 μl of hydrazine, and the reaction was carried out at room temperature. After 30 min of incubation, the reaction was stopped by the addition of 160 μl of hydrazine stop buffer (0.3 M sodium acetate [pH 7.0], 0.1 mM EDTA, 25 μg of tRNA per ml) and precipitated with 700 μl of ethanol. The DNA was purified as described above and dissolved in TE. The DMS- and hydrazine-modified DNA substrates were then used in a standard binding reaction with Rep68 as described previously (23). Time course reactions were done to determine the time required to bind and shift approximately 50% of the starting substrate, and an incubation time of 15 min was chosen. The bound and unbound fractions were separated on a 4% nondenaturing polyacrylamide gel for 1.5 h at 12.5 V/cm. After electrophoresis, the gel was autoradiographed and the bands corresponding to free and bound DNA were eluted by electroelution onto DEAE paper, eluted, extracted with phenol-chloroform, precipitated with ethanol, rinsed, and dried. The DNA was then subjected to piperidine cleavage (19), purified, and dissolved in water. Samples containing equal amounts of radioactivity were then separated on an 8% sequencing gel. A reference sequence ladder was generated by treating an aliquot of the labeled DNA substrate with formic acid and piperidine. The exposed autoradiographs were processed on a densitometer (UltroScan; LKB), and arbitrary intensity values were derived for each band of the sequencing gel.

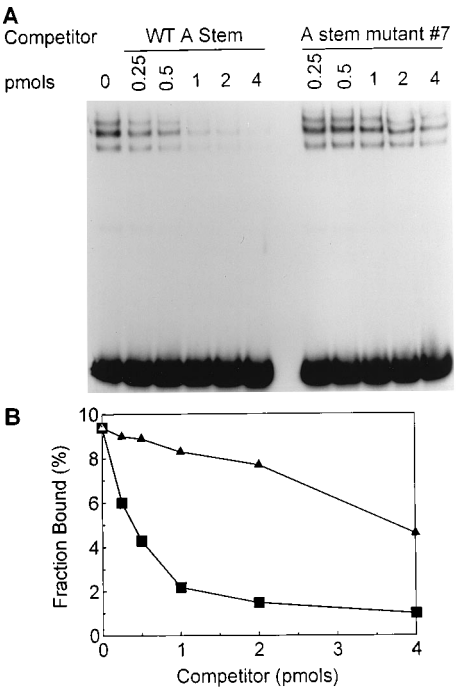

FIG. 3. Competition binding assay used to define the RBE in the A stem. (A) Example of the competition binding assay. For the EMS assay, the standard 10-μl reaction mixture contained 0.75 pmol of Rep68, 0.25 pmol of ^{32}P-labeled A43 wild-type (WT) substrate, and 0 to 4 pmol of unlabeled homologous (WT A stem) or heterologous (A stem mutant #7) competitor DNA. Mutant 7 contained a dinucleotide GC-to-TA transversion at the position shown in Fig. 1b. (B) The fraction of bound labeled substrate was plotted as a function of the amount of competitor DNA added. Squares, homologous wild-type competitor; triangles, mutant 7 competitor.

RESULTS

Defining the A-stem RBE. As mentioned earlier, previous studies have demonstrated the ability of Rep protein to bind a subsequence of the A palindrome in the absence of secondary structure (9, 10, 22, 23). To define the A-stem binding element precisely, we performed competition binding assays between wild-type and mutant A-stem substrates. The wild-type substrate was a synthetic 43-bp double-stranded DNA molecule (Fig. 1b, A43) that contained 30 bp of A stem flanked by 8 and 5 bp of heterologous DNA on the B-C and *trs* proximal sides, respectively. On the B-C proximal side, the 30 bp of A-stem sequence contained all of the sequences shown to be protected from DNase I digestion by Im and Muzyczka (13) as well as an additional 2 bp that had not been protected. In addition, the A-stem substrate was longer on both sides than the maximum region of DNase I protection seen by Chiorini et al. (10) and

contained all but 3 of the 28 bp that these workers suggested contained the RBE. These 3 bp on the B-C proximal side were substituted with heterologous DNA. The wild-type substrate also contained 4 bp on the B-C proximal side within the *Dde*I site that had been shown by Chiorini et al. (10) to be necessary for maximum binding but that had not been present in previous linear A-stem substrates used by us, either in the synthetic A-stem substrate (22, 23) or in the *Dde*I fragment (13, 31) (compare Fig. 1a and b). The wild-type A-stem binding substrate, therefore, contained all of the A sequences that had been implicated in binding Rep (9, 10, 13, 22, 23, 31) and contained sufficient flanking heterologous DNA as suggested by Chiorini et al. (10). We then constructed a series of 13 mutants in which two consecutive base pairs were substituted with nucleotides that produced transversions (Fig. 1b). For example, in mutant 1, the dinucleotide CT was substituted with AG.

To determine the ratio of the K_ds of the wild-type and mutant substrates, each mutant substrate was compared with the wild-type substrate as a competitor for binding to homogeneously pure Rep68. This was done by using EMS assays to measure the amounts of wild-type and mutant competitors that were required to achieve a 50% reduction in the fraction of bound ^{32}P-labeled wild-type substrate under relatively high substrate concentrations. Under these conditions, most if not all of the enzyme is bound to substrate, and the ratio of wild-type to mutant competitor required to achieve some arbitrary level of competition is equal to the ratio of the K_ds. Figure 3 illustrates the competition binding experiment that was done for mutant 7, a GC-to-TA transversion. The ratio of the wild-type K_d to mutant K_d for mutant 7 was calculated from these data to be 0.067. A ratio of 1 would have indicated no difference in the dissociation constants; in this case, the wild-type substrate had a 15-fold-higher affinity for Rep than the mutant. Each mutant was individually compared with the wild type, and the ratios of the K_ds obtained were plotted as a function of the A-stem sequence that was mutated (Fig. 4). The data indicated that the A-stem RBE consists of a 22-bp sequence, CAGTGA GCGAGCGAGCGCGCAG. Each mutation within this 22-bp sequence increased the K_d for Rep by at least twofold. Mutations within an internal 18-bp sequence, GTGAGCGAGC GAGCGCGC, decreased binding affinity by at least threefold. Finally, mutations within a 10-bp core sequence, GCGAGC-GAGC, produced at least a 10-fold change in binding affinity. The symmetry of the effect of these mutations over the 22-bp region suggested that the 22-bp sequence contained a single Rep binding site. If there had been two or more discrete Rep binding sites, we would have expected to find a region within the 22-bp sequence that had relatively little effect on the K_d.

In our previous studies of Rep binding to the A stem, we had used either the A-stem *Dde*I fragment (13, 31) or a synthetic A-stem substrate (22, 23) (Fig. 1a). As suggested by Chiorini et al. (10) and the results in Fig. 4, both of these substrates were missing the 2 bp at the far left end of the 22-bp RBE. The results in Fig. 4 predicted that there would be about a twofold difference in K_d between the A-stem substrate (or *Dde*I fragment) that we had used previously (13, 22, 23, 31) and the A43 substrate. To test this, we compared the A-stem substrate with A43 by competition mobility shift assays. The results confirmed that the A-stem substrate had a twofold-lower affinity for Rep than A43 (data not shown).

To define the contribution of each base within the A-stem core region, we synthesized a series of single-base-pair mutations within the 10-bp core region in which each mutant contained a substitution that produced a transversion (Fig. 1b). The effects of these single-base-pair mutations were measured

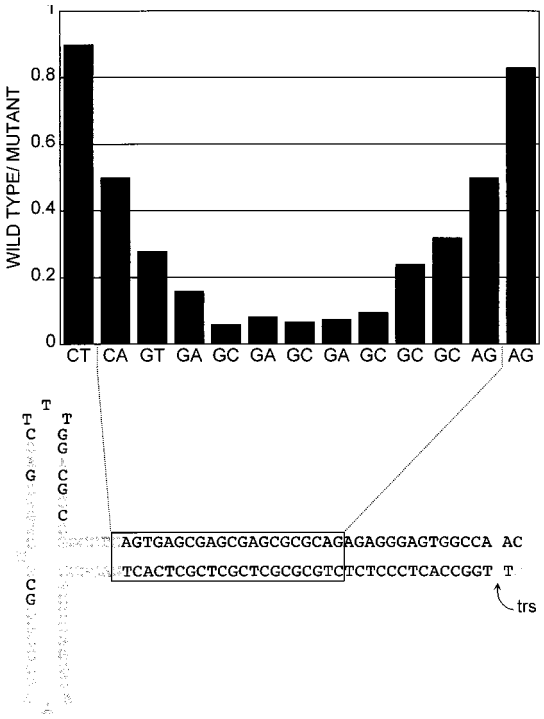

FIG. 4. Ratio of the wild-type K_d to mutant K_d for the dinucleotide transversion mutants within the RBE. For each dinucleotide transversion mutant (mutants 1 to 13 in Fig. 1b), a competition binding assay of the type shown in Fig. 2 was used to determine the ratio of wild-type A43 to mutant A43 substrate required to achieve a 50% reduction in the fraction of bound substrate. A ratio of 1 indicates no change in the Rep binding affinity. The 22-bp sequence that showed measurable changes in binding affinity is shown below the graph as a boxed sequence in the flip orientation of the hairpinned AAV TR. Boldface letters in the sequence are the bases previously shown to be protected from DNase I (13) or copper phenanthralene (14a).

by competition binding gel shift experiments as described above, and the ratio of the wild-type to mutant K_ds was plotted as a function of the mutated base (Fig. 5). Unlike the 2-bp mutations, the effects of the single-base-pair mutations were not symmetrical. The two AT base pairs had no measurable effect on binding, and the remaining base pairs produced changes in the K_d that varied over a sixfold range. The difference in K_d produced by the single-base-transversion mutants

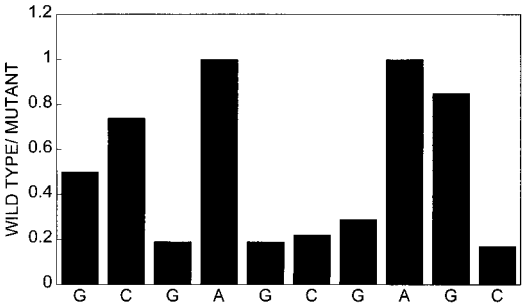

FIG. 5. Ratio of the wild-type K_d to mutant K_d for the single-nucleotide transversion mutants within the 10-bp core region of the RBE. For each single-nucleotide-transversion mutant (mutants 1 to 10 in Fig. 1b), a competition binding assay of the type shown in Fig. 2 was used to determine the ratio of wild-type A43 to mutant A43 substrate required to achieve a 50% reduction in the fraction of bound substrate. A ratio of 1 indicates no change in the Rep binding affinity.

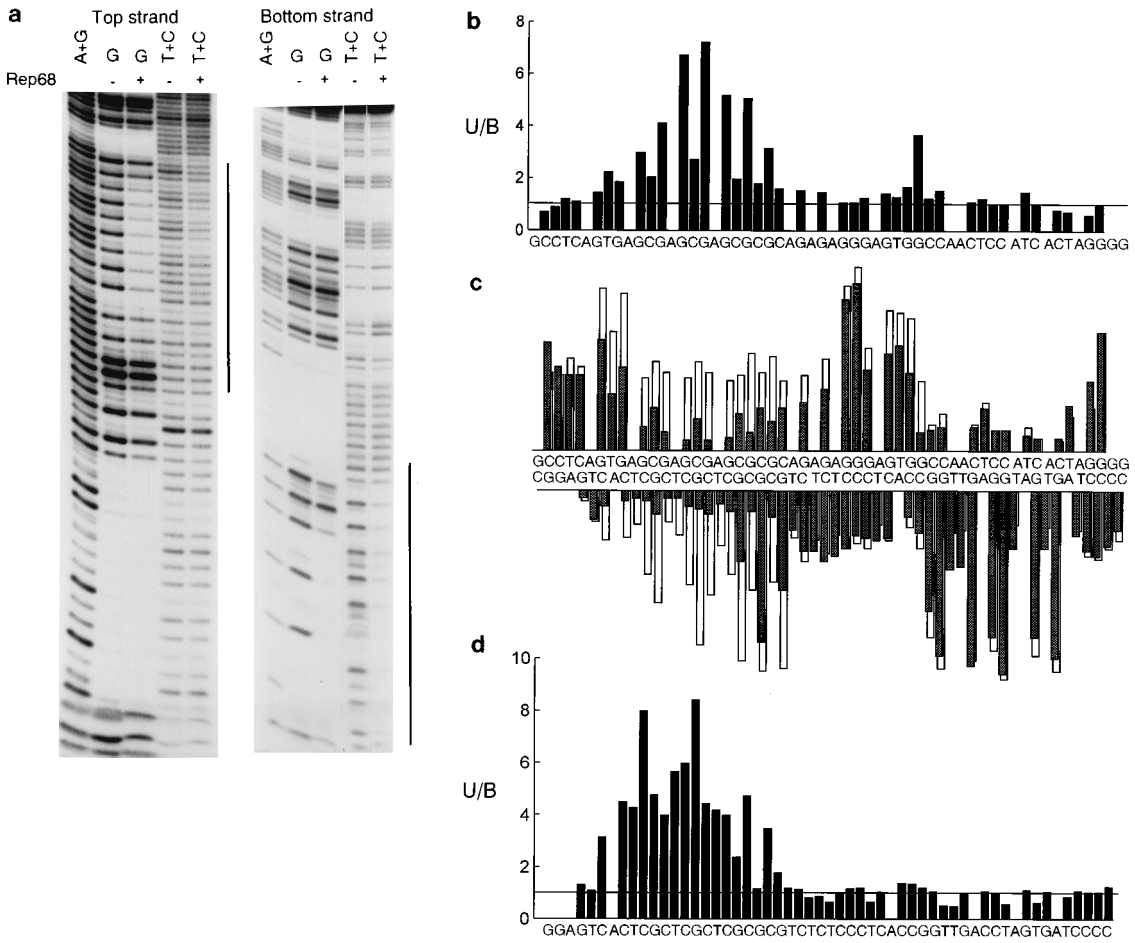

FIG. 6. DMS and hydrazine interference assays using the linear A-D substrate. (a) The A-D substrate (Fig. 2a) was labeled with ^{32}P on the strand containing the *trs* (bottom strand) or its complement (top strand), modified with either DMS (G lanes) or hydrazine (C+T lanes), and bound to Rep. Following cleavage at modified residues, equal amounts of the bound (+) and unbound (−) fractions were compared. Bars at the side of each panel mark the regions where interference was seen. (b to d) Graphical representation of the DMS and hydrazine interference assay using the A-D substrate. In panel c, arbitrary numerical values for each residue were derived by laser densitometry scanning of the autoradiogram and plotted for each residue of the top or bottom strand. Shaded bars represent the amount of bound DNA substrate; clear bars indicate unbound substrate. Panels b and d represent the ratios of unbound to bound substrate (U/B) for modifications in the top and bottom strands of the A-D substrate, respectively. A ratio close to 1 (horizontal line) indicates that modification of the particular base had little or no effect on Rep binding affinity.

did not always accurately predict the effect of the 2-bp substitutions shown in Fig. 4. The predicted values differed from the actual values measured in the dinucleotide substitution experiment over a 1.5- to 6-fold range. We were not certain what these discrepancies meant. At the lower end, these differences may reflect errors in our measurements; at the high end, they may suggest that mutation of multiple base pairs can produce compensatory or synergistic effects on binding affinity within some regions of the core sequence.

Effects of chemical modifications on Rep binding to a linear A-stem substrate. The competition analysis described above used substrates containing transversions on both strands of the RBE sequence. Therefore, it could not distinguish which residue on which strand of the DNA helix was actually a contact point for the bound Rep molecule. To overcome this limitation, we did interference binding assays using chemically modified substrates. Two types of chemical modifications were used, methylation of guanine residues at the N-7 position with DMS and modification of thymine and cytosine residues with hydrazine. The substrate used for this experiment contained all of the A and D sequences and 7 bp of the C palindrome. (Fig. 2a, A-D substrate; see Materials and Methods.) The chemical modifications were done under conditions that produced on average one modified base per substrate molecule. Bound and unbound substrate molecules were separated by EMS assay (not shown) and treated with piperidine to cleave at the modified bases. The cleavage sites within the bound and unbound substrates were then compared on a DNA sequencing gel (Fig. 6a). To get a quantitative measure of the effect of each base on binding, the autoradiogram of Fig. 6 was subjected to optical densitometry and the relative intensity of each band was plotted in arbitrary units for both the bound and unbound substrates (Fig. 6c). The ratio of unbound to bound band intensities for each base was then calculated and plotted as a function of the DNA sequence for both the *trs*-containing strand (Fig. 6d) and the complementary strand (Fig. 6b). An unbound-to-bound ratio of 1 indicated that a particular base modification had no effect on Rep binding. Ratios greater than 1 suggested that the particular base in question was overrepresented in the unbound fraction and, therefore, made a specific contact with the Rep protein. To get an approximate estimate of the variation in this kind of analysis, we examined the unbound-to-bound ratio in the D sequence, which all previous data had shown was not involved in Rep binding (10, 13,

23, 31). In the case of Fig. 6, we concluded that a ratio of 1.5 or greater was likely to indicate that a particular base was involved in Rep binding. Using this standard, we concluded that the 18-base sequence GCGCGCTCGCTCGCTCAC was recognized by the Rep protein on the *trs*-containing strand and the 17-base sequence TGAGCGAGCGAGCGCGC was bound on the complementary strand. These 17- and 18-bp sequences are essentially identical to the 18-bp sequence identified by the dinucleotide mutagenesis analysis described above. We anticipated that the interference assays would be less sensitive than the dinucleotide mutagenesis analysis because they were measuring on average the effect of single-base modifications on only one strand rather than the effect of four base changes distributed on both strands. Thus, there was an excellent correspondence between the results of the interference assays and the mutagenesis experiments (compare Fig. 4 and 6b to d; see also Fig. 9).

The DMS treatment used in these experiments was not sufficient to produce significant methylation of A residues. Thus, we could not determine whether A residues in either strand were involved in Rep binding. However, with the possible exception of one G residue on the *trs*-containing strand, all of the remaining bases (G, C, and T) within the 18-bp region clearly contributed to Rep binding. The four G residues on the *trs*-containing strand that were identified earlier by methylation interference in a study by Owens et al. (26) (Fig. 1a) are within the 18-bp region identified in this study. Finally, we note that a single G residue (underlined) on the *trs*-minus strand (within the sequence G<u>G</u>CCAA) near the *trs* position also appeared to be involved in binding.

Effects of chemical modifications on Rep binding to the complete hairpinned TR. We and others had shown that *trs* endonuclease activity was higher on complete hairpinned substrates than on molecules that were missing the B and C palindromes (10, 23, 31). This finding implied that some contact occurred between the bound Rep protein and a portion of the B and C palindromes. Furthermore, our DNase I protection assays (13) (Fig. 1a) and the methylation interference experiments of Ashktorab and Srivastava (1) suggested that most of the contacts occurred within the small internal palindrome that was furthest away from the *trs* regardless of whether this was the B or C palindrome. Since these two palindromes are flipped during AAV DNA replication to produce two different sequence configurations, the Rep protein was believed to recognize only the secondary structure within the B-C region and not a particular DNA sequence (4, 18, 22). To see if we could detect specific bases within the B-C region that make contact with the Rep protein and to see if the contacts in the A-stem binding element are different in the presence of the B and C hairpins, we performed chemical modification interference experiments with both the flip and flop orientations of the hairpinned TR. The two orientations of the hairpin were isolated from a plasmid which contains the complete TR sequence plus an additional D sequence (Fig. 2a). Upon boiling and chilling, the hairpinned flip and flop orientations were separated from each other and from nonhairpinned starting material by gel electrophoresis (not shown). Both the flip and flop hairpinned substrates were then sequenced to confirm that each substrate was not contaminated with the starting linear duplex DNA and contained only a hairpinned species. In addition, the sequence confirmed that each substrate contained the correct orientation (Fig. 2b). Each substrate was then treated with DMS and hydrazine as described above to produce on average a single modification per molecule and then bound to Rep protein. Bound and unbound species were isolated and compared to identify possible Rep contact points (Fig. 7).

As expected, modifications in the A-stem RBE interfered with Rep binding to both the flip and flop substrates (Fig. 7b). Essentially the same 18- to 20-base region that was seen with the linear A-stem substrates described above (A43 and A-D substrates) was involved in binding Rep in both the flip and flop orientations (Fig. 7b; summarized in Fig. 9). However, the shape of the interference pattern appeared to be different between the linear A-stem RBE and the same sequence in the context of the hairpinned substrate (compare Fig. 7b and 6b to d). In addition, the contribution of the *trs*-containing strand to binding appeared to be lower in the hairpinned substrates than in the linear A-D substrate. Finally, comparison of the flip and flop orientations (Fig. 7b) suggested that there might be some differences in the contributions of specific bases in the two orientations, particularly bases in the *trs*-minus strand.

Surprisingly, examination of the sequences in the B and C palindromes also revealed specific bases that were involved in Rep binding. This finding suggested that the secondary structure of the B-C region was not the only feature recognized by Rep. Particularly prominent was the contribution of the CT<u>T</u><u>T</u>G motif that is present at the tip of the palindrome furthest away from the *trs* site. This motif and particularly the two underlined T residues clearly affected Rep binding in both the flip and flop substrates. The CTTTG motif was interesting in that it is one of only two sequence motifs in the B and C palindromes that are identical in sequence and in position with respect to the *trs* site in both the flip and flop orientations. In addition, other bases within the B and C palindromes also appeared to interfere with binding, albeit to a lesser extent. These bases were not the same in the two orientations. However, in both kinds of substrates, all of the bases interacting with Rep were clustered in or near the small internal palindrome that was furthest away from the *trs*, that is, the B palindrome in the flip orientation and the C palindrome in the flop orientation (Fig. 7c and 9).

Differential binding affinities of the hairpinned TR and linear A-stem substrates for Rep. We have pointed out before (23) that during AAV DNA replication, the Rep protein should theoretically encounter three kinds of substrates that contain an RBE. The first is a hairpinned end in the flip or flop configuration. This kind of substrate must be resolved to an open duplex end for net DNA synthesis to occur. The second kind of substrate is an end that has already been resolved. In principle, there should be a mechanism for discriminating against this substrate because nicking linear ends would waste time and energy during DNA replication. In fact, we and others have shown that linear substrates that are not capable of forming the secondary structure element are nicked inefficiently in vitro at the *trs* (10, 22, 31) and replicate poorly in vivo (6, 18, 29). Finally, Rep is also likely to encounter a third substrate which consists of a linear TR flanked by two D sequences, which we refer to as a 2D linear substrate. 2D linear substrates are found in the middle of dimer replicative intermediates which are commonly formed during AAV DNA synthesis. An example of this kind of substrate is the linear *Xba*I-*Bgl*II fragment from $pTR_{BR}3D$ (Fig. 2a). To see if there was a difference in binding affinity for these three kinds of substrates, we compared the A43, A-D, flop, flip, and 2D linear substrates by competition gel shift assay for the ability to compete for binding with flop hairpin DNA (Fig. 8). Essentially no difference was seen in binding affinities of flop and flip hairpins (data not shown). A-D and A43 DNA, both of which contain a complete RBE but are incapable of forming a secondary structure element, competed to approximately the same extent with homologous flop DNA. The four- to fivefold difference between A-D and A43 DNA may be due to the fact that the A-D

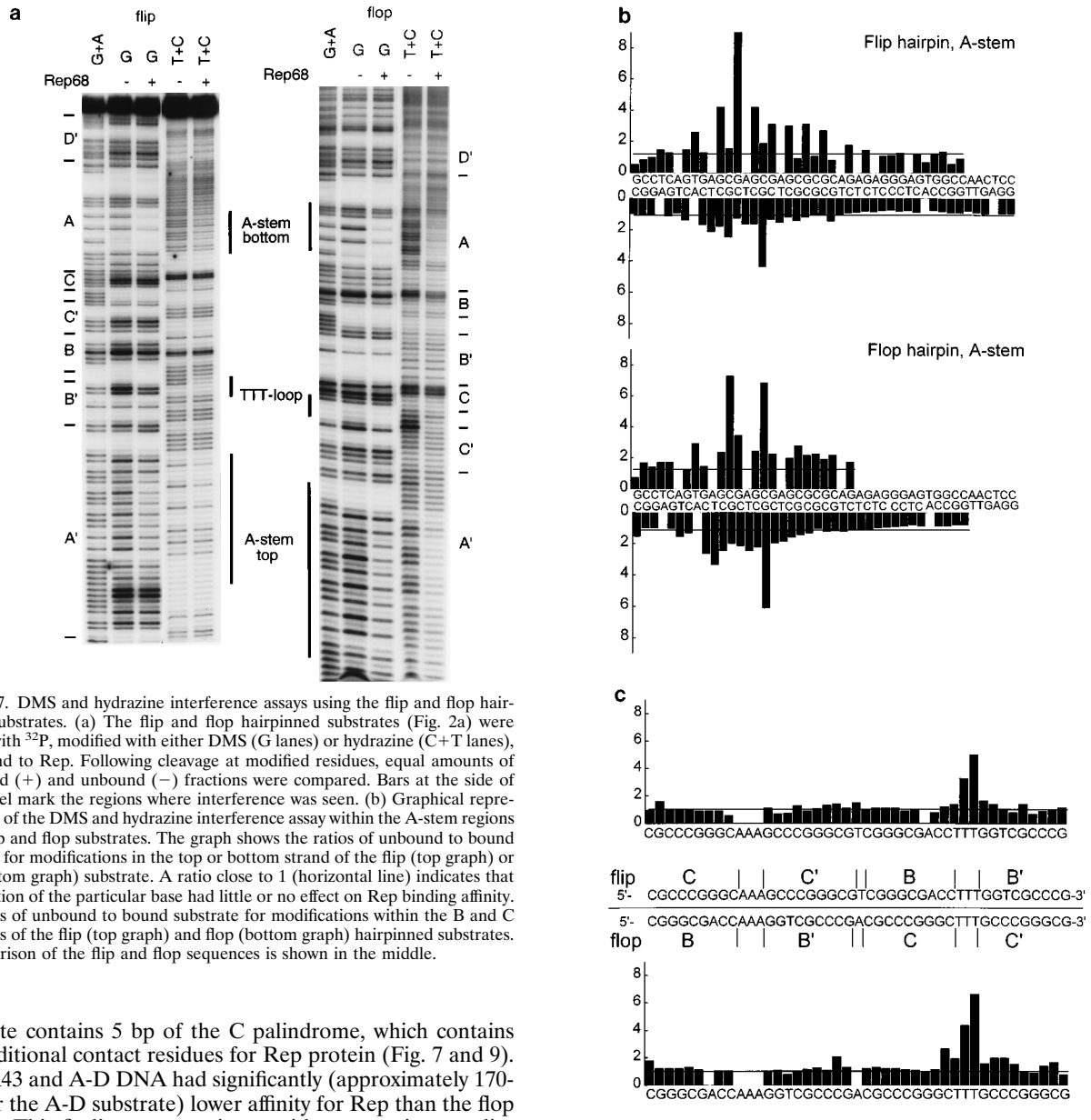

FIG. 7. DMS and hydrazine interference assays using the flip and flop hairpinned substrates. (a) The flip and flop hairpinned substrates (Fig. 2a) were labeled with ^{32}P, modified with either DMS (G lanes) or hydrazine (C+T lanes), and bound to Rep. Following cleavage at modified residues, equal amounts of the bound (+) and unbound (−) fractions were compared. Bars at the side of each panel mark the regions where interference was seen. (b) Graphical representation of the DMS and hydrazine interference assay within the A-stem regions of the flip and flop substrates. The graph shows the ratios of unbound to bound substrate for modifications in the top or bottom strand of the flip (top graph) or flop (bottom graph) substrate. A ratio close to 1 (horizontal line) indicates that modification of the particular base had little or no effect on Rep binding affinity. (c) Ratios of unbound to bound substrate for modifications within the B and C sequences of the flip (top graph) and flop (bottom graph) hairpinned substrates. A comparison of the flip and flop sequences is shown in the middle.

substrate contains 5 bp of the C palindrome, which contains two additional contact residues for Rep protein (Fig. 7 and 9). Both A43 and A-D DNA had significantly (approximately 170-fold for the A-D substrate) lower affinity for Rep than the flop hairpin. This finding was consistent with our previous studies using substrates similar to A-D DNA (13, 23, 31). Finally, the 2D linear substrate also had a lower affinity for the Rep protein than the flop hairpin, but the difference was considerably less, approximately 10-fold when a correction was made for the fact that the 2D substrate has two RBEs.

DISCUSSION

Mapping the A-stem RBE. We have used two approaches to map in detail the bases within the A stem that affect Rep binding affinity, competition gel shift assays and chemical modification interference assays. These two approaches complement each other. The reciprocal competition analysis provides information about whether a particular set of base pairs contribute to binding affinity. The chemical modification interference assay can identify critical base residues on either DNA strand and their relative effects on binding. Both methods are base specific; that is, nonspecific contacts with the sugar or phosphate backbone of the substrate are not likely to be affected. It is worth noting that chemical modifications of nucleotide residues produce base moieties that are of a different type than transversion mutations. However, the results obtained from these two methods closely match in regard to the bases that are shown to be involved in binding by the Rep protein. Analysis of the A43 2-bp-substitution mutants identified an unusually long 22-bp recognition sequence (plus or minus one base at either end) that was necessary for optimum binding (Fig. 9a). Chemical modification experiments of the A-D substrate, which were expected to be inherently less sensitive because only one base is modified, identified an 18-bp sequence that is a subset of the larger 22-bp sequence (Fig. 9a).

Analysis of the 2-bp-transversion mutants and chemical modification interference assays both suggested that the RBE was a single recognition site. Had there been two independent sites within the 22-bp sequence, we would have expected to see regions that had relatively little effect on binding flanked by regions that had a larger effect. Instead, the effects of the 2-bp

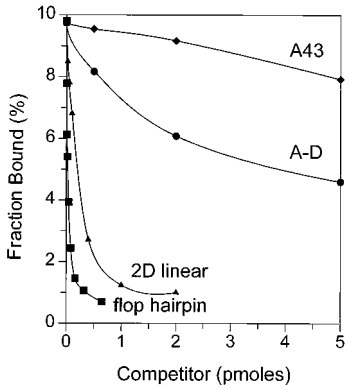

FIG. 8. Competition binding assay comparing linear and hairpinned substrates. For the EMS assay, the standard 10-μl reaction mixture contained 0.075 pmol of Rep68, 0.05 pmol of ^{32}P-labeled flop hairpinned substrate, and 0 to 14 pmol of unlabeled homologous flop substrate (squares) or heterologous competitor DNA (A-D substrate [circles], 2D linear substrate [triangles], or A43 substrate [diamonds]). The fraction of bound labeled substrate was plotted as a function of the amount of competitor DNA added. Competition with unlabeled flip hairpin substrate produced a competition curve identical to the flop curve (not shown). Only a portion of the competition curve is shown; competition with the A43 substrate reached 62% at 14 pmol of competitor. From this, it was extrapolated that 50% competition would require 21 pmol of A43 competitor DNA.

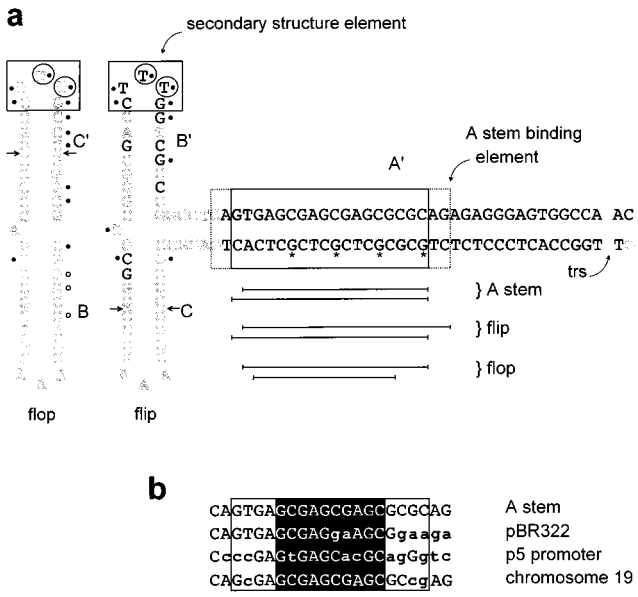

FIG. 9. Summary of information about sequences within the TR that affect Rep binding affinity. (a) The complete hairpinned AAV TR is shown in the flip orientation; only the B and C sequences of the flop orientation are shown. Nucleotides protected from DNase I (13) cleavage are shown in boldface. The solid line box within the A stem indicates the 18-bp sequence that was shown to affect binding affinity by mutagenesis and by DMS and hydrazine interference in this study. The dotted-line extensions indicate two additional bases at either end whose mutagenesis affected the binding affinity by approximately twofold; the 22-bp sequence bounded by the dotted lines is likely to be the minimum RBE, plus or minus one base at each end. The left end of the solid box is the *Dde*I cleavage site. The solid lines below the box show the regions of DMS and hydrazine interference for the top and bottom strands of the A-stem binding element for the A-D substrate (A stem) and the flip and flop substrates. Asterisks indicate the G residues previously identified by Owens et al. (26) and Weitzman et al. (35) by DMS interference assays with the AAV TR and the chromosome 19 RBE, respectively. The 22-bp RBE defined in this study contains the GAGC or GCTC repeats identified earlier in the studies by Weitzman et al. (35), McCarty et al. (22, 23), and Chiorini et al. (9, 10). It is also the sequence predicted earlier by Snyder et al. (31) to contain the RBE on the basis of different arguments and is consistent with the 28-bp sequence suggested by Chiorini et al. (10), which contained 1 and 5 additional bp on the right and left, respectively. Within the secondary structure element, the box indicates the CTTTG element that is identical with respect to both position and sequence in both orientations. Bases which contribute to Rep binding affinity by interference assays are indicated by solid dots. The strongest contribution to binding appeared to come from the two circled T residues. The three open dots in the flop orientation indicate residues that were found to contribute little if anything to binding in the study by Ashktorab and Srivastava (1). Consistent with the previous suggestions by Im and Muzyczka (13) and Ashktorab and Srivastava (1), the residues within the B and C palindromes that are involved in Rep binding are clustered within the palindrome that is further from the terminal resolution site (B in the flip substrate and C in the flop substrate). Arrows indicate the positions of *Sma*I cleavage in the two orientations. Consistent with the results of this study, cleavage of the flip hairpinned substrate with *Sma*I does not affect binding or *trs* endonuclease activity (13, 31). Also consistent with this study, mutations within the *trs* region outside the RBE do not appear to affect Rep binding affinity (10, 23, 31). (b) Comparison of the 22-bp RBE within the A stem of the TR with the other known RBEs in pBR322 and the AAV p5 promoter (22) and in human chromosome 19 (35). The boxed region indicates the 18-bp sequence consistent with the interference data in this study; the shaded sequence indicates the 10-bp core region which has the greatest effect on binding affinity by mutagenesis and interference analysis.

mutations across the RBE were symmetrical across the sequence, with mutations at the center having greater effects on binding affinity than mutations at the ends of the RBE (Fig. 4). This was seen as well in the chemical modification interference assays using the A-D substrate. The ratio of unbound to bound substrate found in these experiments (Fig. 6b to d) was directly proportional to the relative contribution of each residue to binding affinity by Rep protein. The differences between G modifications and C and T modifications did not necessarily show the relative importance of purines versus pyrimidines but rather could reflect the different character of the modifications caused by DMS or hydrazine. However, when G residues alone were examined, it seemed clear that G residues in the middle of the RBE had a greater effect on binding affinity than those at the periphery. For the most part, the same was true when C residues were compared with other C residues, and the pattern of interference within the string of G's, as well as C's, closely matched the results for the 2-bp-transversion mutants (Fig. 4 and 6b to d).

The 22-bp RBE sequence mapped in this study is consistent with most of the previous work from our laboratory and others (10, 22, 23, 26, 31, 35). The 22-bp sequence is virtually identical in sequence and position to the site predicted by us (31) on the basis of mutations within other regions of the A stem as well as our previous DNase I protection studies (Fig. 9a) (13). In the DNase I protection studies (13), 21 of the 22 bp of the RBE were clearly protected; the remaining base pair, a CG that was closest to the B and C palindromes, might also have been protected but could not be clearly scored. The 22-bp RBE also contains the sites of the 5- and 7-bp substitution mutants that reduced binding affinity in our previous studies (23), and it is consistent with the 28-bp sequence suggested by the work of Chiorini et al. (10). Owens et al. (26) had identified four G residues by methylation interference assays on the *trs*-containing strand (Fig. 9a) that appeared to be involved in Rep binding, and Weitzman et al. (35) identified the same four G residues in their analysis of the chromosome 19 target site for AAV integration. Our results confirm their observations. Ashktorab and Srivastava (1) had identified a set of three G residues by methylation interference within the B palindrome of the flop hairpin that did not appear to be involved in Rep binding, and this finding was confirmed by our results (Fig. 9a). Chiorini et al. (10) demonstrated that fragments truncated at the *Dde*I site, which would have 2 bp missing from the left end (B-C proximal) of the RBE, had reduced binding affinity. These 2 bp were missing in the *Dde*I fragment and the filled-in

synthetic A-stem substrate that we had used previously for binding studies (Fig. 1) (13, 22, 23, 31). Consistent with the observation of Chiorini et al. (10), we found that when these 2 bp were mutagenized, there was approximately a twofold drop in binding affinity (Fig. 4). Direct comparison by competition gel shift assay of our previously used A-stem substrate with the A43 substrate used in this study confirmed that there was a twofold difference in the binding affinities of these two substrates (data not shown). Finally, the RBE sequence shows strong similarity to three other sequences that have been shown to bind to Rep protein. The 22-bp sequence has a 19-of-22-base identity to the chromosome 19 integration site (35) and is identical at 15 and 11 positions to the pBR322 site and the p5 promoter site, respectively (22) (Fig. 9b).

The 22-bp RBE sequence contains a set of four imperfect GAGC motifs which we and others (1, 10, 22, 23, 35) had speculated might be the important feature of this element for Rep binding. Our analysis suggests that the sequence recognized by Rep may be more complex and may be affected by other nearby sequences. First, the central 10 bp of the RBE, which appeared to have the most pronounced effect on binding when we analyzed the 2-bp-transversion mutants, consisted of a portion of the GAGC motifs, GCGAGCGAGC (Fig. 4 and 9b). Furthermore, it was clear that bases outside the GAGC motifs have an effect on binding. This would explain in part why an oligonucleotide that consists only of the GAGC motifs is not sufficient to bind Rep protein (35) or binds poorly (10). Chiorini et al. (10) have suggested that additional nonspecific contacts are necessary in addition to the GAGC motifs for Rep binding. Our results did not exclude this possibility but did demonstrate that additional specific base contacts that do not involve the GAGC repeats are necessary for optimum Rep binding. In addition, the results of the chemical interference assay using the A-D substrate suggested that the center of the interaction curve of the bottom strand appeared to be shifted leftward toward the BC palindromes, which could reflect differential binding of Rep to the two DNA strands (Fig. 6 to d). Differential binding to the two strands appeared to be even more pronounced in the hairpinned substrates (Fig. 7b). There were also several indications that there was some flexibility in sequence recognition by the Rep protein. For example, examination of the chemical interference pattern on the *trs*-containing strand of the A-D substrate (Fig. 6d) indicated that the two T residues in the 10-bp core region of the RBE contributed significantly to binding affinity. Modification of either T produced approximately a fourfold difference in the ratio of unbound to bound substrate. In contrast, when the AT base pairs containing these two T residues were individually substituted with CG residues, no effect was seen on binding affinity (Fig. 5). This result suggested that the AT-to-CG transversion mutations produced compensatory interactions on one or both strands. Finally, the A-D substrate and the hairpinned substrates exhibited different patterns of interference (compare Fig. 6b to d and 7b), and there were a number of differences as well between the two hairpinned substrates, flip and flop (Fig. 7b). These differences may be due to the specific and slightly different base contacts made by Rep protein within the B and C palindromes of the flip and flop hairpins (Fig. 7c and 9a) and to the fact that none of these contact sites are available in the A-D substrate.

Specific base contacts in the B and C palindromes. From previous work, it could be argued that contacts between the Rep protein and the B and C palindrome sequences were not likely to be base specific. The argument was that since Rep is capable of binding and resolving two orientations of the B and C palindromes, flip and flop, and the sequences of these two orientations are not the same (4), it is more likely that Rep protein recognizes the secondary structure of the B and C palindromes rather than a specific base sequence. Berns and his colleagues (18) were the first to propose this idea and test it directly. When they substituted either an 8-bp or a 12-bp symmetrical linker for a 9-bp *Sma*I fragment from the C palindrome (Fig. 9a), the ability of the mutant to replicate in vivo was nearly normal (5, 6, 18). A deletion of the *Sma*I fragment or a nonsymmetrical substitution was defective for DNA replication (6, 29). They concluded that the secondary structure of the C palindrome was more important than the actual sequence.

In contrast, the results reported here suggest that the sequences of the internal palindromes, B and C, may have a role in Rep binding and DNA replication. Our chemical modification experiments indicated that strong base contacts occurred between Rep and two T residues (underlined) in a CTTTG motif that is present in the same position of the hairpin structure with respect to the *trs* and the RBE in both the flip and flop orientations (Fig. 7c and 9a). These two T residues affected the unbound-to-bound ratio by four- to sixfold, an effect that was comparable to that seen with bases in the core of the RBE. Modification of the remaining bases of the CTTTG motif also appeared to affect binding affinity, particularly in the flop orientation (Fig. 7c). In addition, several other bases within the B and C palindromes of both the flip and flop orientations appeared to be making base contacts with Rep protein. These additional contacts were not in the same position or sequence with respect to the *trs* and RBE. We note also that most of the contact sites revealed by chemical modification were in the internal palindrome (B or C) that was further away from the *trs*, regardless of whether the orientation was flip or flop. This observation confirmed previous data from methylation interference and DNase I protection studies (1, 13). It was also consistent with our previous finding that the lower end of the C palindrome could be removed by *Sma*I digestion from the flip hairpin substrate without affecting Rep binding or nicking in vitro (Fig. 9a) (13, 31).

Taken together, our results imply that binding by Rep protein to the hairpinned TR is inherently asymmetrical. First, the RBE is a nonpalindromic sequence; thus, Rep is likely to bind to it in a particular orientation. Second, Rep makes contacts predominately with only one of the two internal palindromes, which again implies an asymmetric alignment of Rep with the hairpinned TR. These two elements are, therefore, likely to ensure that Rep is correctly aligned on the TR to nick only the correct strand at the *trs*.

The observation of specific base contacts within the B and C palindromes is not incompatible with the results of the substitution mutants studied by Berns and colleagues (5, 6, 18). All of the substitution mutants used by Berns and colleagues retained at least one of the T residues that is present in the wild-type sequence. Furthermore, the mutations reported by this group would have affected terminal resolution of only the flip, not the flop, hairpin. Thus, a significant difference in replication capacity might not have been apparent. In addition, Berns and his colleagues (5, 6) found that the wild-type sequence was preferred over the substitution mutants under some conditions, suggesting that sequence substitutions within the C palindrome did affect the efficiency of DNA replication. Finally, we and others have shown that regardless of whether the wild-type CTTTG motif is present, there is still a significant difference in the abilities of hairpinned and linear TRs to be nicked in vitro (10, 22, 31). Thus, the fact that the secondary structure of the B and C palindromes places the contact sites in

this region in a particular spatial orientation for Rep interaction clearly plays a role in Rep function at the *trs*.

Contact points in the *trs* region. When we examined the DMS interference pattern obtained with A-D substrate, we found approximately a fourfold interference at a single G residue on the strand opposite to the *trs*, 3 bp away from the cut site (Fig. 6b to d). The significance of this observation was not clear. DMS interference assays on the hairpinned substrates did not show contacts at this position (Fig. 7b, flip). In addition, we and others have reported that mutants in this region that are missing this G residue have approximately the same binding affinity for Rep (10, 23). Nevertheless, these binding experiments might well have missed a fourfold effect. Moreover, it seems clear that mutations in the *trs* region can abolish nicking activity, suggesting that a specific sequence is recognized by Rep in the *trs* region prior to nicking (31). The G residue that emerged from the A-D substrate analysis may be one of these specific *trs* contact sites, but a more systematic analysis of mutations in the *trs* region will be needed to determine the significance of this G residue.

Differential binding affinities of Rep to hairpinned and linear substrates. We reported previously that there was a significant difference in binding affinity between linear substrates that contained the RBE and hairpinned substrates that contained the B and C palindromes in addition to the RBE (13, 23, 31). We estimated the difference in K_d to be greater than 125-fold. Recently, Chiorini et al. (10) have suggested that there is no difference in binding affinity between linear and hairpinned substrates and suggested that the difference in binding affinity that we observed might be due to the fact that we used a linear substrate (the *Dde*I fragment or the A-stem substrate) that was missing a portion of the RBE or possibly was not long enough to provide nonspecific contacts needed for optimal Rep binding. As already mentioned, we found in this study that both the *Dde*I fragment and the A-stem substrate used in our previous studies were indeed missing two residues of the RBE. However, our results suggested that the difference in binding affinity due to these two base pairs was only twofold. We also compared the affinity of Rep for linear and hairpinned substrates directly (Fig. 8) and found that the difference in binding affinity between the hairpinned flip substrate and the A-D substrate containing a complete RBE was approximately 170-fold. This result was consistent with the results of the chemical modification interference experiments which demonstrated that there were specific contacts between Rep and bases within the B and C palindromes. These contact sites were not present in the A-D substrate and probably accounted for at least some of the difference in binding affinity. Finally, we have presented evidence before that substitution mutants within the RBE that consisted of 5 or 7 bp had a greater effect on binding to the linear A-stem substrate than to hairpinned substrate (23). This also suggested a difference in binding affinity between linear and hairpinned substrates. We concluded that there is a substantial difference in binding affinity between linear substrates containing the RBE and substrates which contain the RBE and the B and C palindromes in a hairpinned configuration.

We are not certain why there is a discrepancy between our results with linear and hairpinned substrates and those of Chiorini et al. (9, 10, 13, 23, 31). Possibly it is due to the fact that many of the studies by Chiorini et al. (10) were done with a mutant Rep68 protein that is fused at its N terminus to a portion of the maltose-binding protein of *Escherichia coli* (9). We and others have shown that sequences within the N terminus of Rep are essential for binding to the TR (21, 26, 37). Thus, the 37-kDa maltose-binding domain of the chimeric Rep protein might change the binding characteristics of Rep. Regardless of the reason for the discrepancy, it is worth noting that there is general agreement that Rep protein nicks linear substrates at a much lower frequency than hairpinned substrates (10, 23, 31). Thus, Rep has a mechanism for discriminating between these two kinds of terminal sequences and cuts one preferentially. Our results suggest that much of the discrimination occurs at the level of binding affinity for the two kinds of substrates.

We also compared the binding affinities of flip hairpinned substrates and the 2D linear substrate. The 2D linear substrate contains two RBE sites, two *trs* regions, and a single copy of the B and C palindromes. Essentially, the 2D linear substrate contains the kind of TR structure that is present within dimer molecules produced during AAV DNA replication. We expected that binding to the 2D linear substrate would occur at about the same level as binding to the A-D substrate. Surprisingly, the 2D linear substrate bound Rep much more efficiently than A-D substrate even when a correction was made for the fact that 2D linear DNA contains two RBE sites per molecule. The difference in binding affinity between a hairpinned TR and a 2D linear substrate was approximately 10-fold. It is worth noting that monomer and dimer duplex replicative forms accumulate in vivo at approximately a 5:1 ratio during AAV DNA replication. Thus, the difference in binding affinity between 2D linear DNA and hairpin DNA could account for the relative accumulation of monomer and dimer DNA during AAV DNA replication.

ACKNOWLEDGMENTS

This work was supported by grants PO1 CA2814607, RO1 GM3572302, and HL/DK 50257 from the National Institutes of Health to N.M. and by Public Health Service training grant (T32 AI25530) to J.H.R.

REFERENCES

1. **Ashktorab, H., and A. Srivastava.** 1989. Identification of nuclear proteins that specifically interact with adeno-associated virus type 2 inverted terminal repeat hairpin DNA. J. Virol. **63:**3034–3039.
2. **Batchu, R. B., R. M. Kotin, and P. L. Hermonat.** 1994. The regulatory rep protein of adeno-associated virus binds to sequences within the c-H-ras promoter. Cancer Lett. **86:**23–31.
3. **Beaton, A., P. Palumbo, and K. I. Berns.** 1989. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in *trans* by the Rep protein. J. Virol. **63:**4450–4454.
4. **Berns, K. I., W. W. Hauswirth, K. H. Fife, and E. Lusby.** 1979. Adeno-associated virus DNA replication. Cold Spring Harbor Symp. Quant. Biol. **43**(Pt. 2)**:**781–787.
5. **Bohenzky, R. A., and K. I. Berns.** 1989. Interactions between the termini of adeno-associated virus DNA. J. Mol. Biol. **206:**91–100.
6. **Bohenzky, R. A., R. B. LeFebvre, and K. I. Berns.** 1988. Sequence and symmetry requirements within the internal palindromic sequences of the adeno-associated virus terminal repeat. Virology **166:**316–327.
7. **Brunelle, A., and R. F. Schleif.** 1987. Missing contact probing of DNA-protein interactions. Proc. Natl. Acad. Sci. USA **84:**6673–6676.
8. **Chejanovsky, N., and B. J. Carter.** 1989. Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication. Virology **173:**120–128.
9. **Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urcelay, 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.
10. **Chiorini, J. A, S. M. Wiener, R. A. Owens, S. R. M. Kyostio, 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.
11. **Hermonat, P. L.** 1994. Down-regulation of the human c-fos and c-myc proto-oncogene promoters by adeno-associated virus Rep78. Cancer Lett. **81:**129–136.
12. **Hermonat, P. L., M. A. Labow, R. Wright, K. I. Berns, and N. Muzyczka.** 1984. Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. J. Virol. **51:**329–339.
13. **Im, D. S., and N. Muzyczka.** 1989. Factors that bind to adeno-associated virus terminal repeats. J. Virol. **63:**3095–3104.

14. **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.
14a. **Im, D. S., and N. Muzyczka.** Unpublished observation.
15. **Kyostio, S. R., R. A. Owens, M. D. Weitzman, B. A. Antoni, N. Chejanovsky, and B. J. Carter.** 1994. Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels. J. Virol. **68:**2947–2957.
16. **Labow, M. A., L. H. Graf, Jr., and K. I. Berns.** 1987. Adeno-associated virus gene expression inhibits cellular transformation by heterologous genes. Mol. Cell. Biol. **7:**1320–1325.
17. **Labow, M. A., P. L. Hermonat, and K. I. Berns.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J. Virol. **60:** 251–258.
18. **LeFebvre, R. B., S. Riva, and K. I. Berns.** 1984. Conformation takes precedence over sequence in adeno-associated virus DNA replication. Mol. Cell. Biol. **4:**1416–1419.
19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
20. **McCarty, D. M., M. Christensen, and N. Muzyczka.** 1991. Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein. J. Virol. **65:**2936–2945.
21. **McCarty, D. M., T. H. Ni, and N. Muzyczka.** 1992. Analysis of mutations in adeno-associated virus Rep protein in vivo and in vitro. J. Virol. **66:**4050–4057.
22. **McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka.** 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J. Virol. **68:**4988–4997.
23. **McCarty, D. M., J. H. Ryan, S. Zolotukhin, X. Zhou, and N. Muzyczka.** 1994. Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat. J. Virol. **68:**4998–5006.
24. **McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J. Virol. **62:**1963–1973.
25. **Ni, T. H., X. Zhou, D. M. McCarty, I. Zolotukhin, and N. Muzyczka.** 1994. In vitro replication of adeno-associated virus DNA. J. Virol. **68:**1128–1138.
26. **Owens, R. A., M. D. Weitzman, S. R. Kyostio, and B. J. Carter.** 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J. Virol. **67:**997–1005.
27. **Redemann, B. E., E. Mendelson, and B. J. Carter.** 1989. Adeno-associated virus Rep protein synthesis during productive infection. J. Virol. **63:**873–882.
28. **Rittner, K., R. Heilbronn, J. A. Kleinschmidt, and G. Sczakiel.** 1992. Adeno-associated virus type 2-mediated inhibition of human immunodeficiency virus type 1 (HIV-1) replication: involvement of p78rep/p68rep and the HIV-1 long terminal repeat. J. Gen. Virol. **73:**2977–2981.
29. **Samulski, R. J., A. Srivastava, K. I. Berns, and N. Muzyczka.** 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. Cell **33:**135–143.
30. **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.
31. **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.
32. **Snyder, R. O., R. J. Samulski, and N. Muzyczka.** 1990. In vitro resolution of covalently joined AAV chromosome ends. Cell **60:**105–113.
33. **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.
34. **Tratschin, J. D., J. Tal, and B. J. Carter.** 1986. Negative and positive regulation in *trans* of gene expression from adeno-associated virus vectors in mammalian cells by a viral *rep* gene product. Mol. Cell. Biol. **6:**2884–2894.
35. **Weitzman, M. D., S. R. Kyostio, R. M. Kotin, and R. A. Owens.** 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc. Natl. Acad. Sci. USA **91:**5808–5812.
36. **Wonderling, R. S., S. R. Kyostio, and R. A. Owens.** 1995. A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. J. Virol. **69:**3542–3548.
37. **Yang, Q., and J. P. Trempe.** 1993. Analysis of the terminal repeat binding abilities of mutant adeno-associated virus replication proteins. J. Virol. **67:** 4442–4447.