Interaction of the Adeno-Associated Virus Rep Protein with a Sequence within the A Palindrome of the Viral Terminal Repeat

DOUGLAS M. McCARTY, JOHN H. RYAN, SERGEI ZOLOTUKHIN, XIAOHUAI ZHOU,
AND NICHOLAS MUZYCZKA*

Department of Microbiology, School of Medicine, University at Stony Brook, Stony Brook, New York 11794

Received 15 February 1994/Accepted 26 April 1994

We have characterized a Rep binding sequence which is within the A stem region of the adeno-associated virus terminal repeat (TR) and compared its affinity with that of the complete hairpinned TR for pure Rep68. Both the A stem and the complete TR substrates produced a complex pattern of protein-DNA complexes in which at least six different bound species could be distinguished. Competition experiments suggested that the dissociation constant for the A stem sequence is approximately 125-fold higher than that for the complete TR. The competition experiments also suggested that the average number of Rep molecules per TR substrate molecule under conditions of saturating substrate is 3.7:1, while for the A stem substrate, the ratio is 10:1. In spite of the apparent difference in protein-to-DNA ratio in the complexes, no major difference was seen in the mobility or the pattern of the protein-DNA complexes with the two kinds of substrates, suggesting that the difference in protein-to-DNA ratio was due to the lower stability of the A stem complex rather than the actual number of Rep molecules per DNA molecule. At least some of the difference in stability of the two kinds of complexes was due to the fact that the dissociation rate of the A stem substrate from the protein-DNA complexes was approximately fourfold faster than that of the complete TR. The dissociation rate curves for both substrates, however, were complex, suggesting that substrate was being released from at least two different kinds of protein-DNA complexes at different rates. In addition, we have analyzed binding to several substitution mutants within the A stem of the TR. A five-base mutant near the terminal resolution site (*trs* site) had little effect on binding. Two other mutants produced seven- or five-base substitutions within the 25-bp sequence of the A stem that had been identified in the accompanying report (D. M. McCarty, D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka, *J. Virol.* 68:4988–4997, 1994) as essential for binding. Each of these mutants eliminated some but not all of the repeating GAGC motifs in the 25-bp A stem region. Both of these mutants completely abolished binding to the A stem substrate but only partially reduced binding in the context of the complete hairpinned TR. Furthermore, neither mutant altered the pattern of Rep-DNA complexes produced. The analysis of these mutants supported the idea that at least three recognition elements which are involved in *trs* endonuclease activity are present in the TR: the A stem binding sequence, the secondary structure of the B and C palindromes, and the *trs* site. The results also suggested that the 25-bp A stem binding sequence probably did not consist of multiple, independent Rep binding motifs but represented one binding motif. Finally, our results suggested that a number of different Rep-DNA complexes can be stably formed and that some of these contain more than one DNA molecule.

The *rep* gene of adeno-associated virus (AAV) codes for four overlapping polypeptides (13, 19). The two larger proteins, Rep68 and Rep78, are essential for both the replication of AAV DNA and the positive and negative regulation of the AAV promoters (2, 3, 5, 7, 8, 14, 16, 18, 21, 22). Analysis of *rep* mutants has shown that Rep transactivation requires a functional ATPase domain and the same two DNA binding domains which are essential for binding to the hairpinned terminal repeat (TR) (11, 15, 24). We have also noted that in the absence of the TRs, there is a complex requirement for multiple AAV promoters in *cis* for maximum transactivation (10). There is no requirement for multiple promoters when the TRs are present in *cis*. Both of these observations suggest a role for DNA binding in Rep-mediated transactivation.

We and others had succeeded in demonstrating the binding of the two larger *rep*-encoded polypeptides, Rep78 and Rep68,

only to the complete AAV TR in the hairpin configuration (1, 6, 17). The TR, as depicted in Fig. 1, is composed of three palindromic sequences, A, B, and C, which fold into a hairpinned T-shaped structure in replicating AAV DNA (4, 9, 20). Upon binding to the hairpinned TR, the Rep protein cuts the bottom strand of the hairpin at the terminal resolution site (*trs* site) and becomes covalently linked to the 5' end of the cut strand (7, 16, 18). The 3' OH primer generated by the nick is used to repair the terminal sequence in a process called terminal resolution (18).

Although the biochemical activities of the Rep protein provided a plausible explanation of its role in terminal resolution, they did not suggest how Rep controlled AAV transcription. Furthermore, it was not clear how the Rep protein was oriented so that the correct strand was cut during terminal resolution. We suggested that there must be a linear sequence within the TR that could be recognized in the absence of secondary structure (17). In the accompanying paper (12), we report direct evidence that the Rep protein can bind to a DNA sequence within the A stem of the AAV TR in the absence of the B and C palindromes (Fig. 1, binding element). A similar

* Corresponding author. Mailing address: Department of Microbiology, University at Stony Brook, School of Medicine, Life Sciences Building 260, Stony Brook, NY 11794-5222. Phone: (516) 632-8817. Fax: (516) 632-8891.

FIG. 1. Schematic representation of the AAV TR and DNA substrates for Rep binding. (a) Features of the AAV TR which are thought to interact with Rep protein. The A, B, and C palindromes and the D repeat are indicated. The boxed regions are areas which, by mutation analysis, are implicated in Rep binding and site-specific endonuclease activity. Filled characters represent bases protected by Rep from DNase I digestion. (b) Sequences of the synthetic oligonucleotides which were annealed to form the A stem substrate for Rep binding. These consisted of wt AAV sequences except for the *Xba*I site at the right-hand end. Prior to use, the 5' overhangs of the A stem substrates were repaired with Klenow enzyme. A third synthetic oligonucleotide, which included the B and C palindromes (B-C substrate), could be ligated to the A stem substrate to form a structure identical to the AAV TR except for the *Xba*I site.

observation has been made recently with a mutant Rep fusion protein containing a maltose binding domain (3a). Additional Rep binding sequences were found in the AAV p5 and p19 promoters as well as in heterologous DNA (12). The finding that Rep68 and Rep78 could bind to a variety of DNA substrates, including the p5 and p19 promoters, suggested that Rep might participate in its multiple functions through differential DNA binding. This would provide a mechanism for the role of Rep in the transactivation and repression of AAV transcription. In addition, binding to the A stem would provide a mechanism for orienting the Rep protein on the terminal hairpin so that the correct strand is cut at the *trs* site.

If, as we expect, direct DNA binding is involved in the multiple functions of the Rep protein, it is likely that the affinity for each type of site that we encounter will be an important aspect of the regulation of the AAV life cycle. In this report, we characterize the relative affinity of Rep for the A stem sequence and the complete hairpinned TR and determine the average number of Rep molecules bound by the TR. Consistent with previous observations, we find also that there is a substantial difference in affinity between the two types of substrates. In addition, we have characterized several mutations in the A stem sequence with respect to their effects on binding. Our results support our earlier suggestion (17) that

binding to the complete TR involves multiple DNA sequence elements.

MATERIALS AND METHODS

Rep protein. Rep68 was expressed in recombinant baculovirus-infected SF9 insect cells and purified to homogeneity as previously described (12, 14). The Mono Q peak fraction was used in all of the DNA binding experiments described in this report. The protein concentration was determined with Bradford reagent (Bio-Rad) to be 0.22 mg/ml, using gamma globulin as the standard. Prior to use, the Rep68 Mono Q peak fraction was treated with Tween 20 (polyoxyethylenesorbitan monolaurate) to reduce aggregation of the protein as previously described (12).

DNA substrates. Synthetic oligonucleotides were purified on sequencing gels, and the concentrations were determined spectrophotometrically at 260 nm. The bottom (pyrimidine-rich) strand of the A stem substrate (Fig. 1) was labeled at its 5' end with [γ - 32 P]ATP and polynucleotide kinase in a 10- μ l reaction mixture that contained 70 mM Tris HCl (pH 7.6), 5 mM dithiothreitol, 0.1 mM spermidine, 10 mM MgCl₂, 200 U of T4 polynucleotide kinase, and 160 pmol of oligonucleotide. The reaction mixture was incubated for 1 h at 37°C, then supplemented with 1 mM ATP and 150 U of kinase, and incubated for an additional 1 h at 37°C. The top (purine-rich) strand of the A stem was kinase treated in the presence of nonradioactive ATP. (In one experiment, shown in Fig. 2, the labeling pattern was reversed; that is, the top strand was labeled at its 5' end rather than the bottom strand.) The two complementary A stem oligonucleotides were mixed (170 pmol of each), heated to 100°C, and then annealed by slow cooling to room temperature for 1 to 2 h in 10 mM Tris-HCl (pH 7.6)–1 mM EDTA–50 mM NaCl. Of the annealed product, 160 pmol was ligated to the B-C oligonucleotide as described below. The remaining 10 pmol was treated with Klenow enzyme to fill in the 5' overhangs, after which it was purified by gel electrophoresis on a 6% nondenaturing acrylamide gel. The filled-in duplex A stem substrates were then used for electrophoretic mobility shift (EMS) experiments as described below.

The A stem substrate that had not been filled in was ligated to the B-C oligonucleotide. The ligation reaction mixture contained, in 10 μ l, 50 mM Tris HCl (pH 7.8), 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM ATP, 2,000 U of T4 DNA ligase (New England Biolabs), 160 pmol of duplex A stem substrate, and 80 pmol of the kinase-treated B-C oligonucleotide. The ligation reaction mixture was incubated for 6 h at room temperature, and the products of the reaction were separated on a 6% sequencing gel. The 163-base TR substrate (Fig. 1) was extracted from the gel and used for gel shift assays. Because the radioactive labels in the A stem substrate and the TR substrate were of the same specific activity, the use of identical amounts of radioactivity in the gel shift reactions ensured the use of equivalent concentrations of the two substrates.

For the competition experiments, unlabeled A stem competitor DNA was generated by annealing the complementary A stem oligonucleotides and filling in the 5' overhangs as described above. To generate unlabeled TR competitor DNA, the 5'-labeled TR was treated with calf intestinal phosphatase. This procedure removed more than 95% of the label. The contribution of the remaining radioactivity in the competitor was calculated and subtracted from the bound and free substrate fractions as described in Results.

For the measurement of the dissociation rate of the Rep-

DNA complex, unlabeled competitor was obtained from plasmid pTR418wt. This plasmid contained the TR sequences of AAV cloned into plasmid pBluescript so that they could be excised by digestion with the restriction enzymes *Bgl*II and *Xba*I. The TR sequences obtained from the plasmid were purified by electrophoresis on a 5% nondenaturing acrylamide gel, heated to 100°C, and quickly chilled to form the complete hairpin TR competitor. A more complete description of pTR418wt will be provided elsewhere.

EMS assays. The standard Rep binding reaction mixture contained, in a volume of 10 μ l, 10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid)-NaOH (pH 7.9), 8 mM $MgCl_2$, 1 μ g of poly(dI-dC), 40 mM KCl, 8% glycerol, and 0.2 mM dithiothreitol. The amounts of Rep protein and specific DNA are indicated for each experiment. The reaction mixtures were incubated at room temperature for 20 to 30 min and loaded onto 4% polyacrylamide gels (29:1 acrylamide/bisacrylamide weight ratio) containing 0.5 $\times$ TBE buffer (45 mM Tris-borate, 1 mM EDTA). The gels were electrophoresed at room temperature in 0.5 $\times$ TBE buffer for 2 h at 9 V/cm. The amount of bound and unbound substrate was determined by counting the dried gel in an AMBIS gas flow counter or by cutting out the relevant portions of the gel and counting them in a liquid scintillation counter.

RESULTS

Relative affinity of Rep for TR or A stem. We have demonstrated that Rep protein from recombinant baculovirus-infected insect cells can bind to the linear duplex A stem region of the TR in the absence of the secondary structure contributed by the B and C palindromes. To determine the difference in Rep binding affinity between the hairpin TR and the A stem alone, we used synthetic substrates which could be accurately normalized for DNA concentration. The top (purine-rich) strand of the A stem oligonucleotide (Fig. 1) was labeled at the 5' end with ^{32}P and annealed to the bottom strand. This duplex A stem fragment was then purified by electrophoresis on a nondenaturing polyacrylamide gel. Half of the A stem oligonucleotide was treated with Klenow fragment to fill in the six-nucleotide protruding 5' ends and then used as a substrate for testing the affinity of Rep for the A stem sequence. The remaining half of the A stem duplex was ligated to a synthetic oligonucleotide containing the B and C palindromes (Fig. 1, B-C substrate) via the six-nucleotide protruding ends. The 163-nucleotide ligated product was purified on a denaturing polyacrylamide gel and used as a substrate to determine the affinity of Rep for the complete hairpinned TR. The two substrates were identical in specific activity because both were labeled via the same phosphate at the 5' end of the top strand of their A stem oligonucleotides.

The relative affinities of the TR and A stem were first compared by titration of the Rep protein against a constant amount (0.02 pmol) of each DNA substrate in EMS assays (Fig. 2). The ladder of higher-molecular-weight species in the TR substrate in the absence of Rep protein (Fig. 2, lane 7) arises from oligomerization of the DNA, which we believe to be due to free-radical reactions initiated by the decay of ^{32}P . We have observed similar patterns in preparations of TR substrate made by other means. At low concentrations of Rep, an additional, faster-migrating product which is unrelated to the substrate oligomers appears (lanes 10 to 12). In total, there were six resolvable bands generated by Rep bound to the TR substrate. Although we did not count each protein-DNA complex (PDC) individually, laser densitometry scanning of each lane of the autoradiograph indicated that the intensity of

FIG. 2. Rep protein titrations. Twofold dilutions of Rep protein were bound to 0.02 pmol of DNA substrates consisting of the oligonucleotide containing the B and C palindromes (B-C; lanes 1 to 6), the B-C segment ligated to the A stem substrate (TR; lanes 7 to 12), or the Klenow-repaired A stem substrate alone (A stem; lanes 13 to 18). The ladder of higher-molecular-weight species in the TR without Rep (lane 7) results from multimerization of the labeled substrate DNA.

the slowest-migrating complexes decreased with decreasing Rep concentration while that of the fastest-migrating complexes increased (lanes 8 to 12). This finding suggested either that each PDC species contains a single TR bound to a different number of Rep molecules or that the slower-migrating PDCs contain multiple TR molecules held together by a single Rep complex. Interestingly, the A stem and TR substrates were similarly shifted in mobility and exhibited similar patterns of bands at high Rep concentrations. However, the A stem substrate did not appear to generate the fastest-migrating PDC at low Rep concentrations.

The DNA segment comprising the B and C palindromes in the absence of the A stem was also tested for Rep binding (Fig. 2, lanes 1 to 6). The substrate concentration was the same as had been used for the TR and A stem assays, 0.02 pmol per reaction. Binding to the B-C substrate was barely detectable after long exposures of the gel at the highest Rep concentration (not visible in Fig. 2).

The areas of the dried gel containing the total bound or the free DNA from each reaction were counted in a liquid scintillation counter, and the results were plotted (not shown). Ninety-five percent of the TR substrate could be bound by 120 ng (2.0 pmol) of Rep protein (Fig. 2, lane 9). In contrast to the TR, the highest concentration of Rep used in this assay bound only 18% of the A stem substrate, which was present in the same molar concentration as the TR, i.e., a molar ratio of Rep to total DNA of 200:1. Under conditions in which Rep bound 18% of the TR substrate, the ratio of Rep to TR was 7.5:1. Thus, the apparent difference in affinity for the two substrates by this kind of analysis was approximately 27-fold.

Competition experiments. Because of the complexity of the Rep-substrate interaction, measurement of the apparent K_D for TR and A stem substrates was not likely to give us an accurate estimate of the difference in affinity between the two substrates for Rep. However, regardless of the complexity, it was possible to accurately determine the ratio of the K_D s for the TR and A stem by using a reciprocal competition analysis at high substrate concentrations (Fig. 3). For this purpose, radioactive A stem and TR substrates were synthesized as described for the previous experiment except that the ^{32}P label was at the 5' end of the bottom (pyrimidine-rich) strand of the

FIG. 3. (a) Reciprocal competition analysis of TR and A stem substrates. A 0.1-pmol aliquot of labeled TR (lanes 1 to 10) or A stem (lanes 11 to 20) was used as the substrate for binding to 0.15 pmol of Tween 20-treated Rep protein in the presence of 0.1 to 12.5 pmol of unlabeled homologous or heterologous competitor DNA. The additional bands visible in lanes with TR competitor (lanes 3 to 6 and 13 to 16) result from incomplete dephosphorylation of labeled TR which was used to generate the unlabeled TR competitor. (b) The fraction of bound labeled substrate plotted as a function of the amount of competitor DNA for each labeled substrate. Triangles, TR competitor; circles, A stem competitor.

A stem (Fig. 1). The cold A stem competitor was made by annealing the two synthetic oligonucleotides and repairing the 5' overhangs with Klenow enzyme. The unlabeled TR competitor DNA was synthesized by dephosphorylating the 5'-labeled TR substrate with calf intestinal alkaline phosphatase to remove the label.

The unlabeled TR competitor made in this way contained a small fraction (<0.5%) of TR which was resistant to dephosphorylation and therefore appeared as additional labeled bands in Fig. 3a (lanes 13 to 16). The contribution of this source of label was quantitated in lane 16, and the values were used to make the appropriate corrections to the bound and free DNAs excised from regions which contained the unlabeled TR competitor. The bound and free DNA fractions in each experiment were measured by cutting out the region of each lane encompassing all of the PDC species or free species and counting them. The fraction of bound substrate DNA (TR or A stem) was then plotted as function of competitor (Fig. 3b).

The EMS assay in Fig. 3a, lanes 1 to 10, was a competition for the binding of 15 ng (0.25 pmol) of Rep between 0.1 pmol of labeled TR and either unlabeled TR (lanes 3 to 6) or unlabeled A stem (lanes 7 to 10). Despite the use of a 10-fold-higher DNA-to-protein ratio over previous assays, a relatively large fraction (20%) of the labeled substrate was bound. Hence, the addition of equimolar unlabeled homo-

FIG. 4. Scatchard plot of Rep-TR and Rep-A stem binding, using the competition data in Fig. 3 (lanes 2 to 6 for the TR and lanes 12 and 17 to 20 for the A stem).

gous competitor reduced the bound fraction by only 28% rather than the expected 50%. Under these conditions, the difference in affinity for the two different substrates would have been underestimated because a rather high fraction of the starting substrate was bound. Nevertheless, it was clear that the difference in K_D was substantial because the addition of even the highest concentrations of A stem competitor did not decrease the fraction of TR that was bound. This experiment essentially confirmed earlier observations that the nonhair-pinned terminal sequence was a poor competitor for the hairpinned TR (1, 6, 17) and suggested that the difference in K_D for the two substrates was at least 125-fold. We could, however, demonstrate A stem competition of TR binding if the competition experiment was modified by using a lower initial concentration of TR so that approximately 70% of the starting substrate was bound (data not shown).

In the reciprocal experiment (Fig. 3a, lanes 11 to 20), the substrate was labeled A stem and the competitor was either unlabeled TR (lanes 13 to 16) or unlabeled A stem (lanes 17 to 20). The reaction in lane 13 showed that binding to the A stem was eliminated when equimolar TR was present. This was consistent with the observation that the A stem did not compete for TR binding at the concentrations used in Fig. 3, lanes 7 to 10. The effect of the TR competitor on A stem binding illustrated the point that the TR bound the same pool of protein as the A stem. Thus, it was unlikely that there were different kinds of Rep protein, one that bound the TR and another that bound the A stem sequence. The homologous competition with A stem (lanes 17 to 20) showed that the competitor was binding Rep as expected. As before with the TR substrate, the concentration of labeled A stem was not saturating for Rep, and equimolar homologous competitor reduced the fraction of labeled substrate bound by 30% rather than 50%. Because 0.1 pmol of the TR competitor decreased binding to the A stem substrate to the same level (approximately zero) as 12.5 pmol of the homologous competitor, we could estimate a 125-fold difference in affinity between the two DNAs. This represented a minimum difference because the first datum point of the TR competition (0.1 pmol) was essentially at zero bound substrate.

A Scatchard analysis (Fig. 4) of the homologous competition experiments was used to derive constants of the apparent K_D s of the Rep-TR and Rep-A stem complexes. The negative inverses of the slopes of the two lines yielded apparent K_D s of 6.3×10^{-9} and 2.0×10^{-8} for the TR and A stem substrates, respectively. We note, however, that these estimates were based on the assumption that the interaction between Rep and the substrate was simple, consisting of a single molecule of Rep bound to a single site on each substrate molecule. Our pattern

of multiple PDCs suggested that this was clearly not the case, and therefore these estimates should be interpreted with caution.

The x intercepts of the Scatchard plots (Fig. 4) were a function of the number of Rep molecules bound per DNA molecule, assuming that there was one DNA binding site per Rep molecule. Interestingly, the TR substrate bound an average of 3.7 Rep molecules, while the A stem substrate bound 10 Rep molecules. We have noted these values as averages because the electrophoretic patterns of PDCs were complex and probably represented substrate with variable numbers of bound Rep molecules.

There were several possible explanations for the 2.7-fold difference in enzyme-to-substrate ratio at saturation between A stem and TR. One possibility was that the TR and the A stem substrates were forming different kinds of complexes. Another was that the Mono Q fraction of Rep68 that was used contained a fraction of molecules which was capable of binding the TR but not the A stem. Neither of these possibilities seemed likely because the two kinds of substrates produced virtually identical patterns of PDCs and because each substrate could compete with the other under appropriate assay conditions. An alternative explanation was that the A stem-Rep complex was less stable than the TR complex because either the association or dissociation rates for the two substrates were different.

Dissociation rates. The dissociation rates for the two substrates could be determined by challenging the preformed PDCs with an excess of unlabeled substrate and measuring the rate of PDC decay. In this experiment, we used TR competitor which was isolated from plasmid pTR418wt as described in Materials and Methods. This plasmid was used because it provided a large amount of competitor fragment whose precise concentration was not critical. The decay of the labeled TR and A stem protein-DNA complexes was measured over a time span of 0.5 to 60 min by measuring the decrease in the fraction of bound labeled substrate as a function of time (Fig. 5). It was expected that a simple dissociation of free DNA from the bound complexes would proceed at an exponential rate. However, the observed dissociation rates for both the TR and the A stem did not follow a simple exponential curve (that is, a straight line on a semilogarithmic plot). One explanation for this curve (Fig. 5b) was that we were measuring two different dissociation rates for each substrate. Indeed, a close examination of the autoradiograph revealed that for the TR substrate, the two slowest-migrating PDCs were reduced in intensity at 0.5 min, while the intensity of the faster-migrating species had increased (compare Fig. 5a, lanes 3 and 5). After 0.5 min, the different PDC species appear to decay at the same rate. The A stem PDCs behaved similarly. Thus, the slowest-migrating species appear to be higher-order complexes which decay into the faster-migrating complexes and then to free DNA. The simplest explanation for the biphasic dissociation curve, therefore, was that there were two types of Rep-DNA complexes (for both substrates) that released free DNA at two different rates, one fast and one slow.

Despite the complex dissociation curve, the half-life of the Rep-DNA complexes containing either TR or A stem could be estimated. The half-lives of the TR and A stem complexes were 3.25 and 0.80 min, respectively, approximately a fourfold difference. Thus, the approximately 3-fold difference in the maximum amount of each substrate that a given amount of Rep could bind (discussed above) and the 125-fold difference in K_D for the two substrates could be explained, in part, by the 4-fold difference in the dissociation rates.

Sequence requirements for Rep binding within the A stem.

FIG. 5. (a) Time course of Rep-DNA dissociation. Labeled TR or A stem (0.01 pmol per lane) substrate (Sub) was incubated with Tween 20-treated Rep protein (0.5 pmol per lane) for 30 min and then challenged with 4 pmol of unlabeled TR competitor (Comp). Aliquots were subjected to electrophoresis on nondenaturing 4% acrylamide gels at the indicated times (lanes 5 to 18) after the addition of unlabeled competitor. Aliquots from a separate reaction without competitor were loaded at time zero (lanes 3 and 4) or 60 min later (lanes 19 and 20). Lane 21 contains 0.01 pmol of 3'-end-labeled competitor TR which was generated from plasmid DNA. (b) Plot of the fraction of labeled substrate bound as a function of time on a semilogarithmic scale.

The A stem substrate had been constructed with a six-nucleotide protruding 5' terminus at the left end which could be attached to the synthetic oligonucleotide comprising the remainder of the hairpin TR (Fig. 1). For the purpose of the preceding experiments, this end had been filled in when the A stem alone was used as a substrate. When we performed experiments using the same oligonucleotide without filling in the protruding 5' termini, relatively little binding to the Rep protein was observed (Fig. 6, lanes 7 and 12). This suggested that sequences within that six-nucleotide single-stranded region were important for recognition by Rep. To determine the minimum sequence required at the left end of the A stem for Rep binding, the protruding 5' end was filled incrementally, using one, two, three, or all four of the deoxynucleoside triphosphates. The addition of a single deoxyguanine residue to the terminus slightly increased the level of Rep binding (Fig. 6; compare lanes 6 and 7 or lanes 11 and 12). The further addition of the next residue, deoxycytidine, restored Rep binding to the level of the completely double stranded oligonucleotide (Fig. 6; compare lanes 3 and 5 or lanes 8 and 10).

FIG. 6. Requirement for duplex DNA at the left end of Rep binding site. Combinations of nucleotide residues were added to the protruding 5' termini of the A stem substrate as indicated above each lane. Each modified substrate was generated from annealed oligonucleotides labeled at the 5' end of the top strand and was tested for binding to Rep at high (360 ng of Rep per 10 μ l) or low (72 ng) concentration. Weak binding to the unmodified (None) A stem substrate is evident in lanes 7 and 12. Lane 1 is the fully duplex substrate in the absence of Rep. Lane 2 is a fully duplex substrate that was labeled at the 3' end and gel purified. The sequence of the unmodified substrate is shown below. The regions protected by Rep from DNase I digestion (12) are indicated by a line above the sequence. Arrows at the ends indicate the nucleotides that were partially or completely repaired with Klenow enzyme. The *trs* site is also indicated.

The addition of the remaining three nucleotides to the terminus had no effect on binding (Fig. 6, lanes 3 to 5 and 8 to 10). This finding suggested that the Rep binding site was bracketed on the left by the G residue at nucleotide 95 of the AAV sequence.

DNase footprinting on a supercoiled plasmid template had indicated that three regions of the A stem were strongly protected by Rep (12). (These regions of the A stem are indicated at the bottom of Fig. 7 by lines above the sequence.) Two of the protected regions were composed of the sequence GAGC and were separated by one additional copy of GAGC. A fourth, imperfect copy of this motif, GCGC, was to the right of the protected regions. To determine the contribution of these regions to binding, we synthesized several different versions of the A stem oligonucleotide which contained transversion mutations spanning the protected regions. (The substituted sequences are shown in italic at the bottom of Fig. 7). Each mutated A stem substrate was tested for Rep binding with and without the attached segment containing the B and C palindromes; that is, each mutant was tested in the context of the complete hairpinned TR (Fig. 7a) or the A stem sequence alone (Fig. 7b).

In the context of the complete TR, the substitution of 7 bp from AAV nucleotides 92 to 98 (region 1) reduced Rep binding by approximately 50% (Fig. 7a, lane 3). This was consistent with the observation above that single-stranded DNA in this region inhibited binding. However, it was not clear why the difference in binding between the wild type (wt) and the 7-bp substitution mutant was less than the difference previously observed between the wt sequence with six nucleotides single stranded versus four nucleotides single stranded

FIG. 7. Rep binding to substrates with substitution mutations. Duplex oligonucleotides were synthesized with groups of transversions centered on regions protected by Rep protein from DNase I digestion (indicated by lines above the sequence). The substituted sequences are shown in italic below. Approximately 0.02 pmol of each substrate was used in 10 μ l of the standard gel shift assay as described in Materials and Methods. (a): A stem oligonucleotides containing each mutated area, 1, 2, or 3, and all combinations were ligated to the oligonucleotide containing the B-C hairpin segment and tested for binding to 120 ng of Rep protein as described in Materials and Methods. Lanes 1 and 2 are the wt sequence in the absence and presence, respectively, of Rep68. The substrates in lanes 3 to 5 have substitutions of the single region indicated. Lanes 6 to 9 have substrates with multiple regions substituted as indicated. (b) The mutated A stem oligonucleotides, without the attached B-C segment, were filled in with Klenow enzyme and assayed for binding to 240 ng of Rep68. All notations are the same as in panel a.

(Fig. 6, lanes 5 and 7). One possible explanation was that a minimum extent of duplex DNA was required for efficient binding. The substitution of 5 bp from nucleotides 101 to 105 (region 2) resulted in only a slight reduction in binding (Fig. 7a, lane 4). The mutation of 5 bp from nucleotides 118 to 122 in region 3 (nearest the *trs* site) had no detectable effect (Fig. 7a, lane 5).

We also synthesized multiple mutations in which combinations of two or all three of the protected regions were substituted. Substitution of both regions 1 and 2 nearly eliminated Rep binding (Fig. 7a, lane 6). This finding suggested that the sequences within regions 1 and 2 were redundant with respect to their Rep binding properties, but together they were essential. Oligonucleotides substituted in both regions 1 and 3 bound Rep at the same level as observed with substitution of region 1 alone (compare lanes 3 and 7 in Fig. 7a; see also Fig. 8). Similarly, mutant 2,3 reduced Rep binding to approximately the same level as mutant 2 alone (Fig. 7a and 8). This finding indicated that region 3 contributed little to the affinity of Rep protein for these substrates. Substitution of all three sites eliminated detectable Rep binding (Fig. 7a and 8).

If two regions contributed to Rep binding, it was of interest to determine whether a different pattern of PDCs would be seen when region 1 or 2 was missing. Conceivably, if regions 1 and 2 each bound a separate Rep molecule (or Rep complex), mutations in one of these regions would produce an altered gel shift pattern. To determine if this was the case, we incubated the wt substrate and mutant 2,3 or 1,3 with two different concentrations of Rep68 so that we could detect the presence

FIG. 8. Electrophoretic pattern of Rep bound to mutated substrate DNA. Oligonucleotide substrates (0.02 pmol each) with multiple substitution mutations in the A stem (see Fig. 7) were ligated to the B and C palindromes and bound to 12 ng (lanes 1, 4, 7, and 10), 120 ng (lanes 2, 5, 8, and 11), or 0 ng (lanes 3, 6, 9, and 12) of Rep protein under standard EMS conditions as described in Materials and Methods. The samples were electrophoresed on 4% polyacrylamide-0.5× TBE gels.

of most of the PDC species if they were formed (Fig. 8). The substrate missing region 2 or region 1 formed essentially the same pattern of complexes as the wt substrate. This finding suggested that regions 1 and 2 cooperated in binding the Rep protein or complex of proteins.

The results of Rep binding assays with the mutated A stem substrates lacking the B-C hairpin (but with the protruding 5' ends filled in) were different from those seen with the same mutations in the context of the complete TR (Fig. 7b). Both mutations in region 1 and region 2 abolished binding nearly completely. Thus, in the absence of the secondary structure element, both of these regions were essential. The mutation in region 3, again, had little effect on binding by Rep protein.

DISCUSSION

We have compared the binding of homogeneously pure Rep68 to the complete hairpinned TR with binding to a linear Rep binding sequence. The linear sequence is contained in a 25-bp region of the A stem identified in the accompanying report (12) and predicted by our earlier genetic analysis of the TR (17). In our earlier studies, we were unable to detect direct interactions between nonhairpinned A stem sequences and Rep protein (6, 17). This was probably due to the lower concentrations of Rep available at the time from AAV-infected HeLa cells and is consistent with the fact that the affinity for the A stem is much lower than that for the complete TR in the hairpin configuration. Our comparisons of the Rep binding affinity for the synthetic A stem, with and without the B and C palindromes attached, support this conclusion. When we titrated Rep protein in the presence of these substrates, only 10% of the A stem substrate could be bound under conditions in which virtually all of the complete TR (>95%) was bound (Fig. 2). In the accompanying report (12), we demonstrated that we could use an A stem DNA affinity column for purification of the Rep protein but were unable to quantitatively bind all of the Rep protein in our extracts. Our inability (in this study) to bind all of the A stem substrate even at very high Rep concentrations is consistent with the results obtained from DNA affinity chromatography. Even when the Rep-to-DNA ratio was 200:1, only 18% of the A stem substrate was bound (Fig. 2).

Our competition analysis between TR and A stem indicates

that there is a large difference in affinity for Rep. The A stem competitor DNA does not significantly compete against TR for Rep binding at 125-fold excess. Conversely, Rep binding to the labeled A stem is essentially zero in the presence of an equimolar amount of the TR competitor. This finding suggests a difference in the dissociation constant of at least 125-fold between the TR and A stem substrates.

We have used the data from our competition experiments to estimate the apparent dissociation constants of the TR and A stem complexes. Scatchard plots yield K_D s of 6.3×10^{-9} and 2.0×10^{-8} for the TR and A stem, respectively. These values should be interpreted with caution because we have made no attempt to account for the formation of Rep multimers either prior to or resulting from binding, nor for the possible complexity of the binding site. When these characteristics are understood, more meaningful values for the dissociation constants may be derived.

The nature of the Rep multimers. We and others have commented on the fact that stable Rep-DNA complexes probably contain more than one Rep DNA molecule (1, 6, 15). Our Scatchard analysis was used to calculate the average number of molecules of Rep that are bound within a stable TR or A stem complex. For the complete TR this ratio was 3.7 Rep molecules for each TR. The calculations for the A stem revealed a 2.7-fold difference in the number of molecules of Rep per substrate molecule (10:1 for the A stem). Since the pattern of Rep-DNA complexes was not significantly different when the A stem or TR was used as a substrate, it is likely that the 10:1 ratio found for the A stem reflects the stability of this complex rather than the actual average number of Rep molecules associated with each A stem molecule. For example, the difference may stem in part from the difference in dissociation rates which we observed in a time course of PDC decay. In addition, the titration of Rep protein against the TR substrate reveals that the pattern of protein DNA species, which was noted previously (1, 6, 15), consists of at least 6 complexes which vary in intensity with protein concentration. We have observed these six complexes clearly in a number of related experiments. The distribution of Rep protein among these different complexes presumably accounts for the average of 4 Rep molecules to one substrate molecule observed in this study so that some of these complexes must have more than 4 Rep molecules per substrate.

The Rep-TR complex had a half-life of dissociation of 3.25 min, while that of the Rep-A stem complex was 0.8 min. Neither of the substrates exhibited a simple exponential decay curve as would be expected for a single type of PDC dissociating to its component protein and DNA molecules. Rather, the curves appeared to reflect at least two decay events, one rapid and one slow. The biphasic nature of this decay was also observed in the changes in intensity of the fast- and slow-migrating PDCs during the time course. While it is possible that there were two populations of primary Rep-DNA complexes, it seems more plausible that the rapid decay reflects the dissociation of complexes containing two molecules of DNA and Rep protein, to yield one molecule of free DNA and a complex containing the remaining DNA molecule. The dissociation of the complex containing a single DNA molecule would then release the remaining DNA molecule at a much slower rate. The idea of a Rep-DNA network is also consistent with the Rep titration experiment of Fig. 2, which demonstrates that the electrophoretic pattern of PDCs changes with the protein concentration. This finding suggests that the number of protein molecules per DNA substrate can be increased beyond the minimum required for binding, or that the increased concentration of PDCs at high Rep concentrations

results in networking of PDCs. These interpretations are not mutually exclusive, and both suggest that the Rep protein contains multiple binding sites for Rep or for DNA. We have not presented any direct evidence for complexes containing two substrate molecules. We note, however, that direct evidence for a network of two DNA molecules held together by a Rep protein bridge has been presented recently by Weitzman et al. (23) in the context of an interaction between the TR, Rep, and the AAV integration site from human chromosome 19. We also proposed previously the idea of Rep acting as a protein bridge between two DNA molecules as a mechanism for explaining how Rep transactivation might require more than one promoter element (10).

The nature of the Rep recognition sequence. Our experiments have not unequivocally identified the sequence recognized by Rep. In the accompanying report, we demonstrated that a 25-bp region of the A stem near the B and C palindromes was sufficient for binding (12). The results of the mutant analysis presented in this report are consistent with this 25-bp region containing most or all of the binding sequence. The most prominent feature of this region is a repeating GAGC motif present in three perfect copies followed by an imperfect GCGC motif. In the accompanying report, we demonstrated that two of the three regions protected by Rep in a nonhairpinned A stem substrate contained copies of the GAGC repeat (Fig. 6 and 7, regions 1 and 2). The reductions in Rep binding caused by the substitution mutations in the GAGC repeat region of the A stem (Fig. 7, region 1 and region 2 mutants) indicated that this portion of the 25-bp region was important for binding. The observation that two of the four GAGC tetramers can be disrupted (as with the mutation in region 1) without a complete loss in TR binding suggested that these elements are redundant. This was consistent with our finding that Rep binds to a site in the p5 promoter, which contains only one perfect and one or two related tetrameric motifs (see accompanying report [12] for alignment). Interestingly, the mutants in regions 1 and 2, while reduced in binding activity, produced the same electrophoretic pattern of PDCs as the wt TR. This finding suggested that the six PDCs were probably not the result of independent binding sites being occupied by different numbers of Rep molecules. Rather, it suggested that the different PDCs are largely produced by a different number of Rep molecules binding to a single site or by Rep bridges forming protein-DNA networks. Also in this study, we found that Rep binding was weak to an oligonucleotide in which the leftmost GAGC repeat was single stranded and the remaining three copies of the repeat were duplex (Fig. 6). Using A stem substrates that were partially repaired with Klenow enzyme, we showed that the addition of the GC half of the leftmost GAGC motif was sufficient to restore binding to the level of the fully duplex molecule. This implied either that the binding motif started with GC rather than GA or that three GAGC motifs (two perfect, one imperfect) were sufficient for maximum binding as long as at least two nucleotides of flanking duplex DNA were present. However, binding experiments in which we used synthetic oligonucleotides containing either three perfect GAGC or three perfect GCGA repeats failed to detect a stable Rep complex (data not shown). Taken together, our results have led us to conclude that most if not all of the Rep binding sequence is within a 25-bp sequence of the A stem (Fig. 1) that contains the GAGC repeats. However, it is not clear from our results whether the recognition element is a tandem array of the GAGC repeats or a more complex sequence within this region.

Two other points are worth discussing. First, the contribution toward Rep binding of a separate region of protection

near the *trs* site (Fig. 7, region 3) appeared to be small. A mutation at this site had no detectable effect on Rep binding to the TR or the A stem. This was consistent with our previous observation that the *trs* site can be single stranded or its sequence can be altered without significantly affecting Rep binding (17). The second point is that mutants in regions 1 and 2 had different effects when they were tested for binding in the context of the A stem alone rather than the context of the complete TR containing the B and C palindromes. In the presence of the B and C palindromes, both mutants retained partial activity, while in their absence, neither mutant formed a stable complex. This finding suggested that the A stem recognition sequence and the B and C palindromes cooperated in binding Rep and supports the idea we proposed previously that multiple elements are involved in recognition of the hairpinned TR by Rep protein (12, 17). Thus the A stem binding sequence appears to be the primary element required for Rep binding. The B and C palindromes, in the hairpinned configuration, clearly improve the efficiency of binding. Site-specific *trs* endonuclease activity, however, appears to require the juxtaposition of an additional sequence, the *trs* site, within the A stem (12, 17). Finally, as shown in the accompanying report (12), cutting at the *trs* site will occur even in the absence of the B and C palindromes. This type of substrate is cut at a lower efficiency which is comparable to its relative binding affinity. We conclude by suggesting that other Rep binding sites, such as the one in the p5 promoter, may also contain multiple recognition elements that remain to be discovered.

ACKNOWLEDGMENTS

This work was supported by a Program Project grant (N.M.) from the National Cancer Institute (PO1 CA2814607), by a grant (RO1 GM3572302) from the National Institute of General Medical Sciences (N.M.), by a Public Health Service training grant (T32 AI25530) from the National Cancer Institute (D.M.M.), and by a Public Health Service training grant (T32 GM07963) from the Institute of General Medical Sciences (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. 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.
3. 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.
- 3a. Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urceley, 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.
4. Hauswirth, W. W., and K. I. Berns. 1979. Adeno-associated virus DNA replication: non unit length molecules. *Virology* **93**:57-68.
5. 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**:329-333.
6. Im, D. S., and N. Muzyczka. 1989. Factors that bind to adeno-associated virus terminal repeats. *J. Virol.* **63**:3095-3104.
7. 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.
8. 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.
9. Lusby, E., K. H. Fife, and K. I. Berns. 1980. Nucleotide sequence

- of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* **34**:402–409.
10. **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.
 11. **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.
 12. **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.
 13. **Mendelson, E., J. P. Trempe, and B. J. Carter.** 1986. Identification of the *trans*-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. *J. Virol.* **60**:823–832.
 14. **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.
 15. **Owens, R. A., M. D. Weitzman, S. R. M. 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.
 16. **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.
 17. **Snyder, R. O., D. S. Im, T. H. 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.
 18. **Snyder, R. O., R. J. Samulski, and N. Muzyczka.** 1990. In vitro resolution of covalently joined AAV chromosome ends. *Cell* **60**:105–113.
 19. **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.
 20. **Straus, S. E., E. D. Sebring, and J. A. Rose.** 1976. Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis. *Proc. Natl. Acad. Sci. USA* **73**:742–746.
 21. **Tratschin, J., I. L. Miller, and B. J. Carter.** 1984. Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function. *J. Virol.* **51**:611–619.
 22. **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.
 23. **Weitzman, M. D., S. R. M. Kyostio, R. M. Lotin, and R. A. Owens.** Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. USA*, in press.
 24. **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.