JOURNAL OF VIROLOGY, July 1989, p. 3095–3104
0022-538X/89/073095-10$02.00/0

Vol. 63, No. 7

Factors That Bind to Adeno-Associated Virus Terminal Repeats

DONG-SOO IM AND NICHOLAS MUZYCZKA*

Department of Microbiology, SUNY Stony Brook Medical School, Stony Brook, New York 11794-8621

Received 11 December 1988/Accepted 14 March 1989

We have identified and characterized a DNA-protein complex that forms with the adeno-associated virus (AAV) terminal repeats. The complex formed only if the terminal palindrome was in the covalently closed or hairpin configuration; little if any binding was detected with the open duplex form of the terminal repeat. This fact suggested that both secondary structure and primary sequence are essential elements of recognition. DNase I protection studies indicated that virtually all of the A-A′ palindrome and significant portions of the B-B′ and C-C′ palindromes are protected. The postulated terminal resolution site of AAV also is protected. Restriction mapping of the sequences necessary for binding indicated that almost all of the terminal palindrome must be present for binding to occur. Hairpins which are similar in size and shape to the AAV termini did not exhibit competition for binding, and the complex formed only if AAV-infected extracts were used. Thus, the binding reaction is specific for AAV sequences. The viral-coded nonstructural proteins Rep78 and Rep68 comigrated with the DNA-protein complex on neutral acrylamide gels, suggesting that one or both of these proteins are components of the complex. The characteristics of the complex suggested that it has a role in AAV DNA replication.

Adeno-associated virus (AAV) is a linear, single-stranded, human DNA virus (for a review, see reference 2). The viral DNA is believed to replicate by the mechanism illustrated in Fig. 1 (3, 5, 30). The model proposes that the terminal palindrome of AAV functions as a hairpin primer for DNA synthesis. This leads to the formation of a linear, duplex replicative intermediate in which one of the ends is covalently closed. The hairpin end is resolved by a site-specific, strand-specific nick at a position which is believed to be near the end of the terminal palindrome. For purposes of discussion we have called this position the terminal resolution site (*trs*). As a consequence of resolution, the terminal palindrome is inverted and a portion of the parental strand (the terminal palindrome) is transferred to the progeny strand. Following this event, the terminal palindrome is converted to the hairpin configuration once again and is used to prime strand displacement synthesis, thereby generating a single-stranded progeny molecule. Similar mechanisms have been proposed for the closely related autonomous parvoviruses (1, 31).

A number of observations support the model. The replicative intermediate with covalently joined ends is a major species in AAV-infected cells (30). In vivo labeling experiments (11) indicate that DNA synthesis begins and ends near the terminal repeats. Mutants within the terminal repeat are defective for DNA replication in *cis* (26, 27). Two orientations of the terminal palindrome (flip and flop) are found in virion DNA (20). Further, wild-type clones that contain both terminal repeats in only one orientation rapidly generate both orientations of the terminal sequence during replication in human cells (24). Finally, the orientation at either end of the viral DNA is independent of that at the other (19).

A unique feature of the model is the proposed mechanism of terminal resolution. If the model is correct, it is reasonable to expect that AAV-infected extracts should contain factors that can bind to the terminal repeat when it is in the hairpin or covalently closed form. Two observations suggest that an AAV-coded gene product may bind directly to the terminal repeat. At least one AAV open reading frame (29), the *rep* region, codes for a family of four proteins (Rep78, Rep68, Rep50, and Rep42), and mutations in this region are defective for DNA replication (12, 23, 32). Furthermore, when recombinant AAV/simian virus 40 genomes are constructed, simian virus 40 replication is inhibited unless either the AAV terminal sequence or the *rep* gene is inactivated by mutation (13). However, because the *rep* gene has a pleiotropic phenotype that includes the transactivation and repression of other viral genes (14, 15, 33), it is not clear whether these observations mean that one or more of the Rep proteins bind directly to the terminal repeat or induce some cellular or viral protein which binds. Finally, genetic evidence indicates that the ability of the terminal repeats to form the correct secondary structure is essential for DNA replication (4, 18).

As one approach to establishing in vitro AAV DNA replication, we have been isolating and characterizing factors that bind to AAV sequences. In this report, we describe the characterization of factors that bind to the AAV terminal repeat and are found only in AAV-infected extracts. At least two of these proteins are Rep78 and Rep68. Strong binding occurs only when the terminal repeat is in the hairpin configuration, and the binding depends on both the sequence and the secondary structure of the terminal repeats.

MATERIALS AND METHODS

Preparation of nuclear extracts and fraction II. HeLa suspension culture cells were grown at 37°C in Eagle minimal essential medium supplemented with 5% calf serum–1% glutamine–penicillin–streptomycin. To prepare adenovirus- and adenovirus-plus-AAV-infected HeLa cell extracts, 1 liter of cells at a density of 5×10^6 cells per ml was infected with adenovirus type 2 (multiplicity of infection [MOI] = 10) or coinfected with adenovirus type 2 virus (MOI = 5 to 10) and AAV (MOI = 20). The cells were harvested approximately 30 h after infection, and the nuclei were prepared as described by Challberg and Kelly (6) and frozen at −70°C. Frozen nuclei were thawed and extracted for 1 h with buffer A (25 mM Tris hydrochloride [pH 7.5], 0.1 mM EDTA [pH 7.5], 1 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl

* Corresponding author.

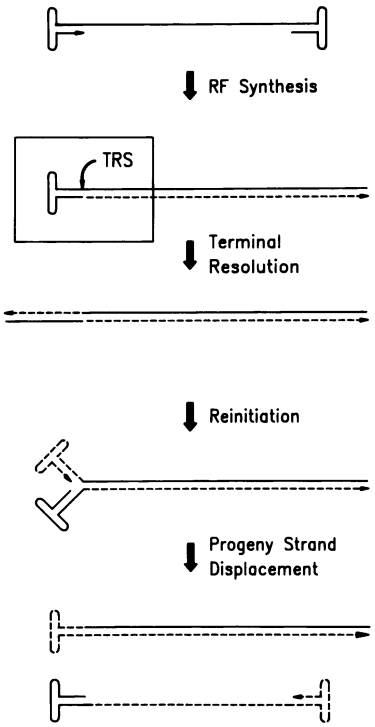


FIG. 1. The current model for AAV DNA replication. Solid line indicates parental DNA; dashed line indicates newly synthesized DNA. The box highlights the covalently joined T-shaped termini which are common to AAV RF intermediates and which were used in the binding assays in this study. See text for additional details.

fluoride, 20% glycerol) containing 0.2 M NaCl on ice. The nuclei were then removed by centrifugation (3,000 × *g* for 15 min), and the supernatant (0.2 M NaCl nuclear extract) was saved. The nuclear pellet was suspended in buffer A containing 1 M NaCl and incubated on ice for 1 h. Both the 0.2 M and 1 M nuclear extracts were centrifuged at 100,000 × *g* for 1 h. The pellets were discarded, and the supernatants were dialyzed against buffer A overnight at 4°C. Any insoluble precipitates in the nuclear extracts which appeared during dialysis were removed by centrifugation at 12,000 × *g* for 15 min. Similar extracts also were prepared from HN21, a cell line carrying a recombinant Rep^+ Cap^- provirus (22).

To prepare fraction II, the 1 M nuclear S100 extract was loaded onto a DEAE 52-cellulose column (5-ml bed volume) equilibrated with buffer A containing 0.05 M NaCl. Proteins were eluted with a 50-ml linear gradient of buffer A containing NaCl (0.05 to 0.6 M). The fractions which contained Rep78 and Rep68 were detected by Western immunoblotting, using anti-Rep antibody, and pooled (fraction I). Fraction I was dialyzed against buffer A and loaded onto a single-stranded DNA agarose column (2-ml bed volume) equilibrated with buffer A containing 0.05 M NaCl. The column was eluted with a 40-ml linear gradient of buffer A containing NaCl (0.05 to 1.0 M). The fractions which contained Rep78 and Rep68 were detected by Western blotting. The active fractions were pooled and concentrated by Amicon Centriflo 25 centrifugation as suggested by the manufacturer (fraction II).

Preparation of whole cell extracts. HeLa monolayer cells were grown in 100-mm-diameter dishes containing Eagle minimal essential medium supplemented with antibiotics and 10% fetal calf serum. The cells were infected with adenovirus type 2 (MOI = 5-10) and AAV type 2 (MOI = 20). At 40 to 48 h postinfection, the cells were washed three times with ice-cold phosphate-buffered saline and suspended in 0.5 ml of cold RIPA buffer (10 mM Tris hydrochloride [pH 8.2], 1% Triton X-100, 1% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS], 0.15 M NaCl, 0.02% NaN_3). The extract was passed several times through an 18-gauge needle and stored frozen at −20°C.

Preparation of radiolabeled DNA. No-end (NE) DNA was prepared from p*sub*201(+) plasmid DNA (25) as described elsewhere (R. Snyder et al., manuscript in preparation). The SSV16 isolate of p*sub*201(+) was used. Briefly, p*sub*201(+) DNA was digested with *Pvu*II and then partially digested with exonuclease III so that the 5′ strand of the terminal palindrome of AAV could adopt the hairpin configuration. The remaining gap was filled in and ligated by using T4 DNA polymerase and T4 DNA ligase. NE DNA is a linear AAV DNA molecule in which both ends are covently joined. NE DNA, therefore, is resistant to exonuclease III digestion, and any unreacted material was removed by exhaustive exonuclease III digestion and subsequent purification by CsCl-ethidium bromide centrifugation. Unless otherwise noted, enzymes were obtained from New England BioLabs, Inc., and used according to the instructions of the manufacturer.

NE DNA has two identical termini bounded by *Xba*I sites (see Fig. 2). To obtain end-labeled DNA termini, NE DNA was digested with *Xba*I and 3′ end labeled with [α-^{32}P]dATP, [α-^{32}P]dCTP, and avian myeloblastosis virus reverse transcriptase (International Biotechnologies, Inc.) in a buffer containing 50 mM Tris hydrochloride (pH 8.3), 135 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol at 37°C for 1 h. Before terminating the reaction, an excess amount of the four cold deoxynucleoside triphosphates were added. The *Xba*I fragment (261 nucleotides) was then purified by electrophoresis on a 6% native polyacrylamide gel and subsequent electroelution. To obtain labeled and nonlabeled linear (nonhairpinned) AAV DNA termini, p*sub*201(+) plasmid DNA was digested with *Pvu*II and *Xba*I and the 180-base-pair (bp) terminal DNA fragment was gel purified by low-melting-point agarose gel electrophoresis. The 3′ end labeling of the *Pvu*II-*Xba*I linear AAV terminal fragment was carried out as described above for NE termini. To obtain the hairpin conformation of the 180-bp *Xba*I-*Pvu*II fragment, it was boiled for 5 min and plunged into ice (Fig. 2).

Preparation of competitor DNAs. The 203-bp DNA fragment from pUC19 was prepared by digesting the plasmid with *Pvu*II and *Xba*I. The 3′-terminal fragment of bovine parvovirus (BPV), was isolated from a *Pst*I digest of the BPV plasmid pVT501 (a gift from R. Bates; 7). Shope fibroma virus termini were prepared from the plasmid pVCB-5 (a gift from M. Reddy; 8) after digestion with *Eco*RI and *Hin*dIII. All these DNA fragments were gel purified by low-melting-point agarose electrophoresis. Each of the fragments was boiled and quick chilled to obtain the hairpin conformation.

DNA-binding assay. The reaction mixtures contained (in a volume of 20 μl) 25 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2 ethanesulfonic acid)–KOH (pH 7.5), 10 mM $MgCl_2$, 1 mM dithiothreitol, 2% glycerol, 0.25 μg bovine serum albumin, 50 mM NaCl, 0.01% Nonidet P-40, either 1 μg of poly(dI-dC) or 0.8 μg of sonicated salmon sperm DNA, 1 to 2 μl of the indicated protein extract (0.85 μg of protein from a 1 M S100 nuclear extract or 0.18 μg of fraction II), and approximately 0.1 ng of the ^{32}P-labeled DNA fragment. The reaction mixtures were incubated at room temperature

(RT) for 20 min and then were loaded onto 4% polyacrylamide gels (39:1 acrylamide-bisacrylamide weight ratio) containing 0.5× TBE (89 mM Tris, 89 mM boric acid, 0.2 mM ethylenediaminetetraacetic acid) (pH 7.6). Gels were electrophoresed in the same 0.5× TBE buffer at RT for about 2 h at 25 mA. They were then transferred to Whatman 3MM filter paper, dried and autoradiographed.

DNase I footprinting. Fraction II was incubated with either 2.5 μg of poly(dI-dC) or with 2 μg of salmon sperm DNA–^{32}P-labeled DNA fragment (about 30,000 cpm) in a 50-μl DNA-binding reaction mixture (as described above) for 20 min at RT. After incubation, 25 μl of 10 mM $MgCl_2$–5 mM CaCl was added and the reaction mixture was treated with DNase I (Worthington Diagnostics; 3,388 U/mg) at a final concentration of 1 to 5 μg/ml for 1 min at RT. The DNase I reaction was terminated by adding 10 μl of stop solution containing 100 mM EDTA (pH 7.5), 5% SDS, and 25 μg/ml tRNA. Samples were extracted once with phenol and chloroform, precipitated with ethanol, dried, and dissolved in a sample-loading buffer containing 80% formamide, 1 mM EDTA, 10 mM NaOH, 0.1% xylene cyanol, and 0.1% bromphenol blue. The DNA samples were then denatured by boiling for 2 min and applied to an 8% polyacrylamide–8 M urea sequencing gel. The sequence of NE termini was determined as described previously (21).

Preparation of anti-Rep peptide antibody. Rabbit polyclonal antibody to the AAV Rep proteins was prepared by a modification of the procedure used by Mendelson et al. (23). The peptide was synthesized by the Center for Analysis and Synthesis of Macromolecules at SUNY/Stony Brook and consisted of the sequence CSTSDAEASINYADRYQNK. All but the first amino acid is encoded within the sequence of AAV (nucleotides 1866 to 1919). The cysteine was added to facilitate conjugation of the peptide to keyhole limpet hemocyanin. Immunization and preparation of the antiserum was done by Cocalico Biologicals, Inc., Reamstown, Pa.

Immunoblotting. Whole cell extract (10 to 100 μl) was mixed with 0.5 volume of sample buffer (16). The mixture was boiled for 5 min, centrifuged, and the supernatant was loaded onto a 10% SDS–polyacrylamide gel. After electrophoresis at 100 to 200 V, the gel was transferred to nitrocellulose, immunoblotted with anti-Rep antibody, and then incubated with 10^5 cpm of ^{125}I-labeled protein A per ml (36.3 μCi/μg; ICN Radiochemicals, Irvine, Calif.) for 1 h at RT. The nitrocellulose sheet was washed three times with phosphate-buffered saline containing 0.05% Tween 20 (20 min per wash), rinsed with phosphate-buffered saline for 20 min, dried, and autoradiographed for 16 to 72 h, without an intensifying screen, at RT.

RESULTS

Substrates. There are several replicative forms (RFs) of AAV DNA (Fig. 1), and in most of them the terminal palindrome is in the hairpin or covalently closed configuration. We therefore prepared the terminal sequences in the hairpin conformation for use as a substrate for binding in a gel-shift assay. Although there are several clones of wild-type AAV DNA available, the most convenient for our purpose was one constructed by Samulski et al. (25), p*sub*201(+). p*sub*201 (Fig. 2) has the advantage that the AAV sequences can be completely separated from vector DNA by *Pvu*II digestion. In addition, the terminal sequences are identical at both ends (including their orientations) and can be isolated on a 180-bp *Pvu*II-*Xba*I fragment. The fragment was labeled at its *Xba*I end (Fig. 2) and boiled to

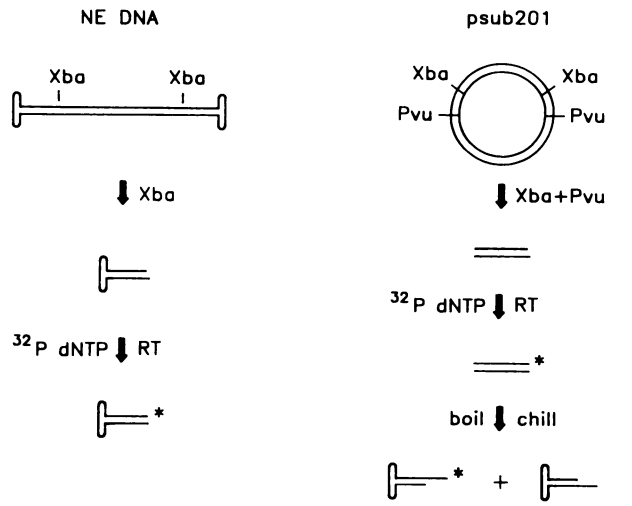


FIG. 2. Substrates for binding assay. Hairpinned termini (right panel) were prepared by digesting the plasmid p*sub*201 with *Pvu*II (which cuts at the junction of the AAV and vector sequences) and *Xba*I. The left and right *Pvu*II-*Xba*I fragments are identical and were separated from the remaining fragments by low-melting-point agarose gel electrophoresis. After elution from the gel, the fragment was labeled at the 3′ end with reverse transcriptase. The fragment was then boiled and chilled quickly to produce the hairpin conformation. Approximately 70 bases of the 180-base fragment are single-stranded in the hairpin substrate. NE DNA (left panel) was prepared from the AAV sequences of p*sub*201 as described in Materials and Methods. NE termini were prepared from NE DNA by digestion with *Xba*I, labeled at their 3′ ends with reverse transcriptase, and separated on neutral acrylamide gels.

generate the T-shaped hairpin conformation of the terminal repeat. The first 110 bases of the labeled fragment are duplex (i.e., hairpinned); the remaining sequences are single-stranded. The hairpin substrate is essentially similar to the right end of wild-type single-stranded AAV virion DNA, with the exception that the first 13 bases of the terminal repeat have been deleted in the construction of p*sub*201 (25) and are not present in the substrate.

To stimulate the end of an AAV RF in which the end is covalently joined, we used p*sub*201 to synthesize a linear AAV DNA molecule, NE DNA, in which both ends are continuous (Fig. 2). The synthesis of NE DNA will be described in detail elsewhere (Snyder et al., in preparation), and the sequences of the NE termini are confirmed below. As in the case of the parent molecule, p*sub*201, the *Xba*I fragments obtained from the ends of NE DNA are identical in sequence and orientation (flip). In the case of the NE fragment, however, the sequence is duplex and wild type over virtually its entire length (261 bases). Within the limitations of the gel-shift assay used here, we saw no significant differences in binding among the NE termini and the hairpin termini of p*sub*201 (data not shown).

Terminal binding factors within AAV-infected extracts. Except under special conditions (28, 34), AAV DNA does not replicate in mammalian cells unless they have been coinfected with a helper virus. A member of either the adenovirus or herpesvirus family can provide helper function (2). We, therefore, prepared nuclear extracts from uninfected cells and cells that had been infected with either adenovirus alone or adenovirus plus AAV. When we incubated these extracts with labeled NE termini, only the extract prepared from cells infected with both adenovirus

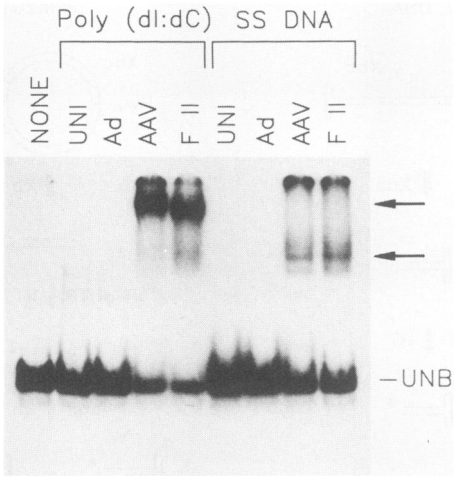


FIG. 3. Binding of extracts to NE termini. The 3′ labeled NE terminal *Xba*I fragment (0.1 ng) was incubated with 0.85 μg of 1 M nuclear extract from uninfected (UNI), adenovirus-infected (Ad), or adenovirus-plus-AAV-infected (AAV) cells in the presence of either 1 μg of poly(dI-dC) or 0.8 μg of salmon sperm DNA. The products of the reaction experiment were fractionated on a nondenaturing polyacrylamide gel. Lanes: F II, the NE substrate was incubated with 0.18 μg of fraction II; NONE, No extract was used; UNB, unbound starting substrate. Upper arrow, Position of the predominant complex seen in the presence of poly(dI-dC); lower arrow, predominant complex seen in the presence of salmon sperm DNA.

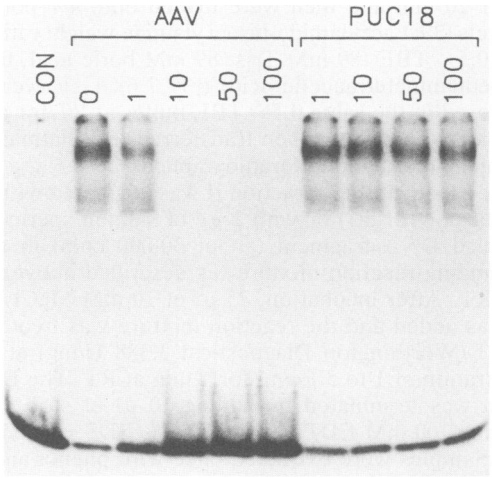


FIG. 4. Specific binding of fraction II to AAV DNA termini. The 3′-end-labeled terminal *Xba*I NE fragment was incubated with fraction II protein in the presence of poly(dI-dC) and either the unlabeled denatured (hairpinned) *Pvu*II-*Xba*I terminal fragment from p*sub*201 or the denatured *Pvu*II-*Xba*I (203 bp) fragment from pUC18. CON, No extract was used; the numbers indicate the molar ratio of unlabeled competitor fragment to labeled substrate. All other conditions were as described in the legend to Fig. 3.

and AAV contained proteins that bound to the AAV termini (Fig. 3, lanes AAV).

The apparent molecular weight of the DNA-protein complex and the number of bands depended on the type of nonspecific competitor DNA used in the assay. In the presence of salmon sperm DNA, only one set of bands was observed (Fig. 3, lower arrow) and these were often present as a doublet. These bands were also present when poly(dI-dC) was used as a competitor, but in this case, the predominant DNA-protein complex had a much higher apparent molecular weight (Fig. 3, upper arrow). This finding suggested that more than one kind of protein-DNA complex could form with the terminal sequences of AAV. A third band also was seen (below the lower arrow in Fig. 3), but this band was not unique to AAV-infected extracts.

Anticipating that at least one of the DNA-protein complexes might be caused by the binding of one of the AAV Rep proteins, we examined the ability of fraction II to bind to the terminal repeat. Fraction II was a partially purified preparation of Rep78 and Rep68 which contained little if any Rep52 and Rep40 and no detectable AAV capsid proteins (see below). Fraction II produced both sets of AAV-specific bands seen with the crude nuclear extracts (Fig. 3, lanes F II).

To demonstrate that binding was specific for the AAV terminal sequences, we used the unlabeled terminal fragment (*Pvu*II-*Xba*I) from p*sub*201 in the hairpin conformation and an unrelated *Pvu*II-*Xba*I pUC18 fragment (203 bp) as competitors in the binding assay (Fig. 4). Competition with the cold AAV fragment was seen at a 1:1 ratio of unlabeled to labeled NE fragment, and virtually complete competition was obtained at a 10:1 ratio. No competition was seen with the pUC18 fragment, even at a ratio of 100:1. In the experiment shown (Fig. 4), the pUC18 fragment was denatured before use in the binding assay to see whether binding to the AAV termini might be caused by a factor that binds nonspecifically to single-stranded DNA. In other experiments, the pUC18 fragment was used in the native duplex form and this form also did not result in competition for binding (not shown).

Effect of secondary structure on binding. Previous studies of mutations within the AAV terminal repeats had suggested that both the DNA sequence and the ability to form the T-shaped secondary structure were essential in order for the termini to serve as origins for DNA synthesis (4, 18, 26). To see whether binding depended on the secondary structure of the terminal repeat, we compared the ability of the *Pvu*II-

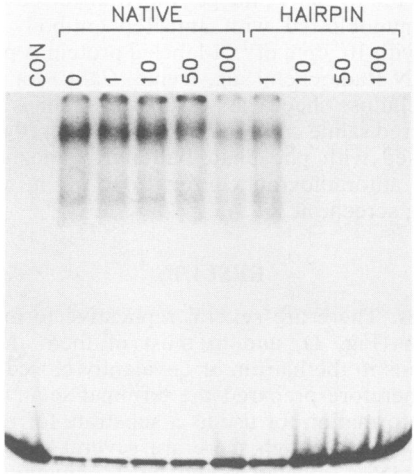


FIG. 5. Competition with native and hairpin termini. Labeled NE terminal *Xba*I fragment was incubated with fraction II protein in the presence of poly(dI-dC) and either the unlabeled duplex (NATIVE) form or the denatured (HAIRPIN) form of the *Pvu*II-*Xba*I fragment from p*sub*201. CON, No protein extract was used; the numbers indicate the molar ratio of unlabeled p*sub*201 fragment to labeled NE substrate. All other conditions were as described in the legend to Fig. 3.

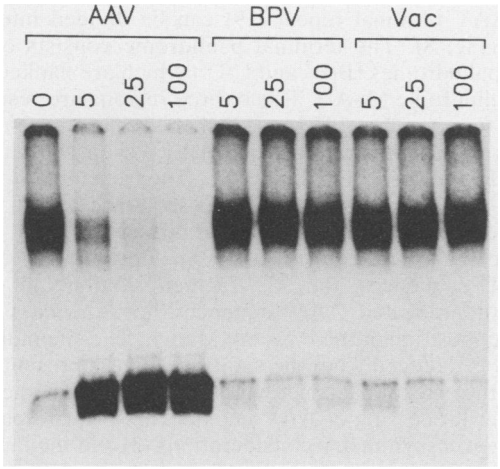
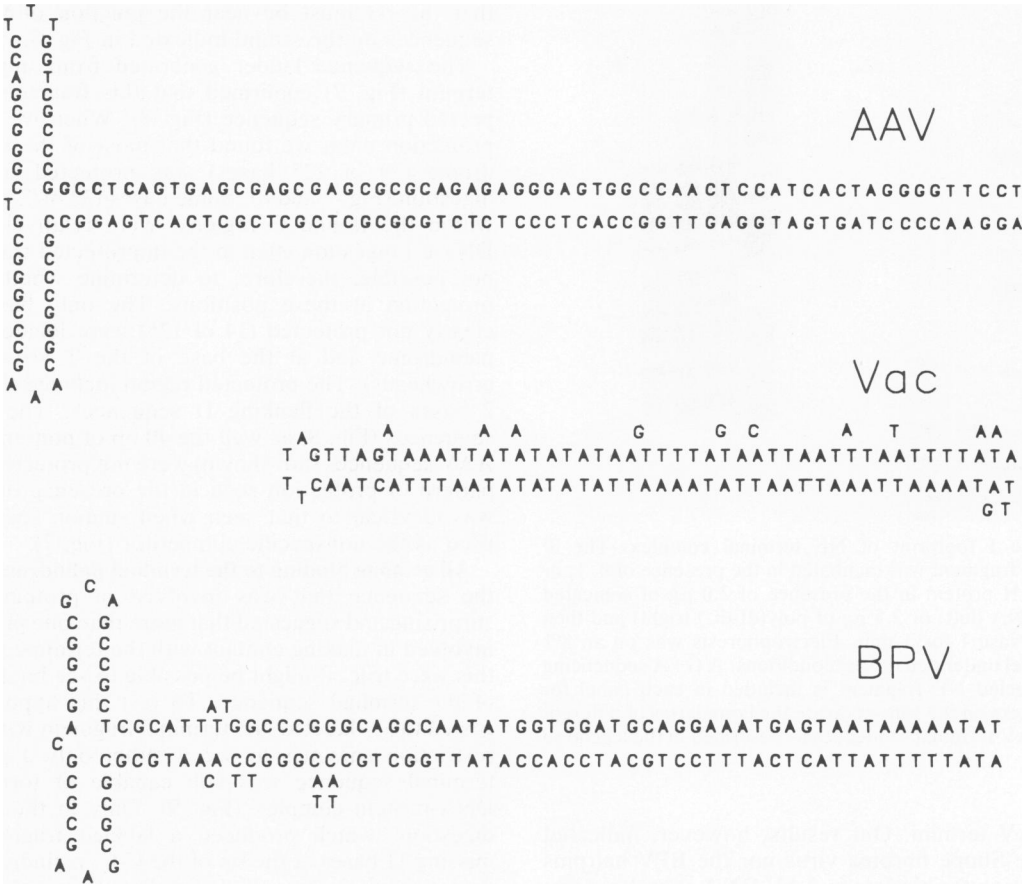


FIG. 6. Competition with other hairpins. (Top) The 3′-end-labeled *Xba*I fragment of NE DNA was incubated with 0.35 μg of fraction II protein in the presence of the indicated molar ratios of unlabeled competitor fragment to labeled substrate. The competitor fragments were the 180-base *Pvu*II-*Xba*I terminal fragment from p*sub*201 (AAV), the 300-base *Pst*I fragment from pVT501 which contains the 3′ viral terminal sequence of BPV, and the *Hin*dIII-*Eco*RI fragment from pVCB-5 (300 bases) which contains the terminal sequence of Shope fibroma virus (Vac). Each competitor fragment was boiled and quickly chilled just before use to form the hairpin configuration. Incubation was carried out in the presence of 0.8 μg of sonicated salmon sperm DNA. Other conditions were as described in the legend to Fig. 3. (Bottom) The sequences and predicted secondary structures of the terminal hairpins within the AAV, Shope fibroma virus (Vac), and BPV fragments which were used as competitors.



*Xba*I terminal fragment (Fig. 2) to act as a specific competitor in either the native or hairpinned conformation. As shown in Fig. 5, competition for binding was seen with the hairpinned conformation of the *Pvu*II-*Xba*I fragment, as before. In contrast, competition with the native (nonhairpinned) conformation of the *Pvu*II-*Xba*I fragment occurred only at a ratio of 100:1. In other experiments, we were unable to detect binding directly to the labeled *Pvu*II-*Xba*I terminal fragment in the nonhairpin configuration (data not shown). This suggested that the secondary structure of the terminal sequences was an important element in the formation of the DNA-protein complexes.

Competition with other hairpins. In view of the importance of the secondary structure, we asked whether any hairpinned DNA end was sufficient for binding regardless of the DNA sequence. Two types of hairpinned fragments were isolated and tested for their competition ability in the gel-shift assay (Fig. 6, bottom). The first was a hairpin formed from the terminal sequences of Shope fibroma virus (a member of the poxvirus family) which was isolated from the plasmid pVCB-5 (8). The second fragment was obtained from pVT501 (7) and contained the 3′ terminal hairpin from BPV. BPV is an autonomous parvovirus whose nonstructural genes have significant similarity to the AAV *rep* genes (7). As shown in Fig. 6 (bottom), the potential secondary structure of the BPV end is quite similar in size and shape to

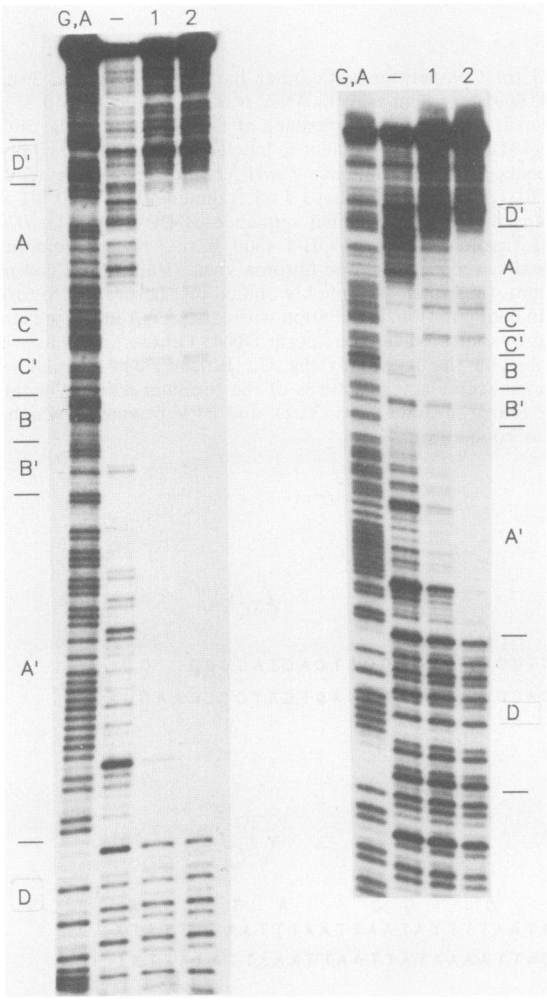


FIG. 7. DNase I footprint of NE terminal complex. The 3′ labeled NE *Xba*I fragment was incubated in the presence of 0, 1, or 2 μg of fraction II protein in the presence of 2.0 μg of sonicated salmon sperm DNA (left) or 2.5 μg of poly(dI-dC) (right) and then digested with DNase I for 1 min. Electrophoresis was on an 8% polyacrylamide gel under denaturing conditions. A G+A sequencing ladder of the labeled NE fragment is included in each panel for comparison. Letters on the sides indicate the boundaries of different regions of the AAV terminal sequences as described in the legend to Fig. 8.

that of the AAV termini. Our results, however, indicated that neither the Shope fibroma virus nor the BPV hairpins exhibit competition for binding to AAV DNA at molar ratios as high as 100:1 (Fig. 6, top). We concluded that both the primary sequences and the secondary structures of the AAV termini were essential for forming the DNA-protein complexes.

DNase I footprinting. We used the technique of DNase I footprinting to determine which sequences within the terminal repeat might be involved in the binding reaction. The DNA-protein complex was formed by using 3′ labeled NE termini in the presence of either poly(dI-dC) or salmon sperm DNA which were then treated with DNase I. The protected sequences were then compared with a sequence ladder obtained from unprotected NE termini. The results are illustrated in Fig. 7, and a summary of the protection data is presented in Fig. 8.

The AAV terminal repeat (19) can be divided into four regions (Fig. 8). The terminal palindrome consists of two smaller palindromes (B-B′ and C-C′) which are flanked by a larger palindrome (A-A′). These three regions are responsible for forming the 125-base T-shaped hairpin structure which is found in single-stranded AAV virion DNA and in some RF species (Fig. 1 and 8). The remainder of the terminal repeat consists of a 20-bp sequence called the D sequence which is present at both ends of the AAV DNA molecule in an inverted orientation. In addition to the A, B, C, and D sequences, the NE fragment contains 48 bp of internal (nonrepeated) AAV sequences (not shown in Fig. 8). The exact position of the *trs* is not known. The site indicated in Fig. 8 is deduced from the sequence that is present most often at the ends of packaged viral DNA (9). It assumes that no further processing of AAV ends occurs prior to packaging. However, symmetry considerations (3) and the fact that both ends of AAV are equivalent for resolution (19) suggest that the *trs* must be near the junction of the A and D′ sequences on the strand indicated in Fig. 8.

The sequence ladder generated from unprotected NE termini (Fig. 7) confirmed that this fragment had the expected primary sequence (Fig. 8). When we examined the protection data, we found that most of the terminal palindrome (79 of 125 bases) was protected from DNase I digestion (Fig. 7 and 8). Some bases (32 of 125), particularly within the B and C regions, were relatively resistant to DNase I digestion even in the unprotected fragment. It was not possible, therefore, to determine whether there was protection at these positions. The only bases that were clearly not protected (14 of 125) were located in the C-C′ palindrome and at the base of the T structure (Fig. 8, arrowheads). The protected region included the *trs* and 1 to 2 bases of the flanking D sequences. The remaining D sequences (Fig. 8) as well the 40 bp of nonrepeated internal AAV sequences (not shown) were not protected. Finally, the pattern of protection seen in the presence of poly (dI-dC) was identical to that seen when salmon sperm DNA was used as the nonspecific competitor (Fig. 7).

All or none binding to the terminal palindrome. The size of the sequence that was involved in protein binding was surprising and suggested that more than one protein might be involved in making contact with the terminal palindrome. If this were true, it might be possible to see binding to subsets of the terminal sequence. To test this hypothesis, we digested the 3′ labeled NE-terminal fragment with a number of restriction enzymes to see whether only a portion of the terminal sequence was still capable of forming a stable DNA-protein complex (Fig. 9). Only in the case of *Sma*I digestion, which produced a labeled fragment that was missing 11 bases at the tip of the C-C′ palindrome, could we detect the formation of a complex. Removal of both the B and C palindromes (*Dde*I digestion) and cleavage in the middle of the A sequence (*Bss*HII digestion) eliminated binding. In the *Bal*I case, the digestion of the NE fragment was not complete. However, it was clear that only the undigested, intact NE terminal fragment was capable of forming a complex (Fig. 9). As expected, the shorter (labeled) *Bal*I fragment, which contained only the D sequences of the terminal repeat, did not form a complex.

We concluded that the only dispensable sequences were the 11 bases at the tip of the C-C′ palindrome, and that, if more than one protein were involved in the formation of the complex, binding must be coordinate and required the presence of nearly all of the terminal palindrome sequence. The fact that the tip of the C-C′ palindrome was not essential

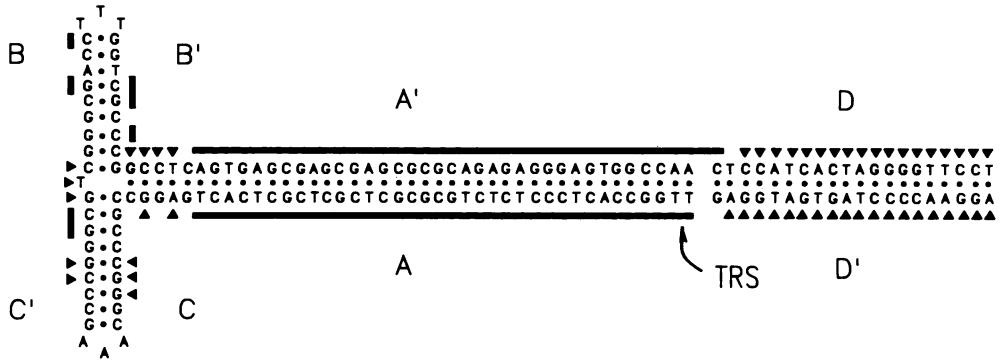


FIG. 8. DNase I protection map. A map of the protected regions within the NE terminal *Xba*I fragment as deduced from Fig. 7. Large letters (A-A′, B-B′, etc.) indicate the four sequence blocks which make up the AAV terminal repeat. The boundary between the A and D sequence blocks is indicated by an interruption in the sequence. TRS, The probable *trs* as deduced from the terminal AAV sequence found in viral DNA. A solid line adjacent to a base indicates that the base was protected from DNase I digestion; an arrowhead indicates that the base was clearly not protected. In some cases, it was not possible to determine whether protection had occurred; these bases have nothing next to them. The NE fragment has an additional 48 bp to the right of the D sequence which are not shown.

for binding was consistent with the protection pattern seen with DNase I (Fig. 8) in which most of the nonprotected bases were found in the C-C′ sequence. This suggested that there was an assymetry to the binding such that one of the T arms (the B sequence) had more contacts that the other one (the C sequence). We do not know whether the assymetry of the protection pattern is a function of the sequences within the B and C palindromes or a function of the orientation of the NE terminal fragment (i.e., flip or flop).

Presence of Rep78 and Rep68 in the terminal complex. We mentioned earlier that the one or more of the Rep proteins may have a direct role in AAV DNA replication. Although a partially purified preparation of Rep78 and Rep68 (fraction II) retained binding activity, we had not shown directly that any of the Rep proteins were present in the complex. Furthermore, the AAV terminal repeats have at least three other functions in the AAV life cycle in addition to their roles in DNA replication. Studies of AAV vectors (22)

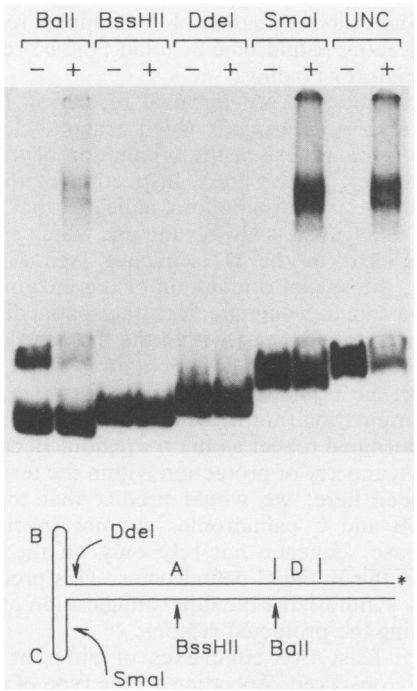


FIG. 9. Restriction mapping of terminal binding. The 3′ labeled terminal *Xba*I NE fragment was digested with the indicated restriction enzymes, extracted with phenol, precipitated with ethanol, and dried. Then, the digestion products were dissolved in 10 mM Tris-acetate (pH 7.5)–1 mM EDTA, and equivalent amounts of each digest were incubated with (+) or without (−) fraction II protein in the presence of salmon sperm DNA. Lane UNC, Uncut NE fragment was used. The diagram at the bottom illustrates the approximate positions of the restriction sites with respect to the four sequence blocks in the terminal repeat and the position of the label (*). Other conditions were as described in the legend to Fig. 3.

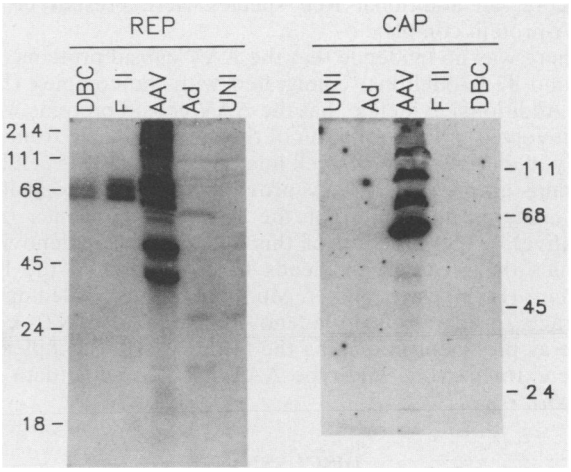


FIG. 10. Comigration of Rep proteins with the terminal complex. Identical volumes of uninfected (UNI), adenovirus-infected (Ad), and adenovirus-plus-AAV-infected (AAV) whole cell extracts as well as 2 μg of fraction II protein (F II) were electrophoresed on an SDS-polyacrylamide gel and then immunoblotted with anti-Rep (REP) or anti-AAV capsid (CAP) antibodies as described in Materials and Methods. The lanes marked DBC contain the protein which comigrates with the DNA-binding complex. To obtain this material, the labeled, denatured *Pvu*II-*Xba*I terminal fragment from p*sub*201 was incubated with fraction II protein in the presence of salmon sperm DNA and electrophoresed on a preparative neutral gel as described in the legend to Fig. 3. The gel-shifted DNA-protein complex was extracted from the gel, and the protein was denatured and applied to the lanes marked DBC.

indicate that the sequences present in the NE terminal fragment are the minimum sequences required for integration and rescue of AAV proviruses and for packaging of AAV viral DNA. The mechanism of AAV integration is unknown but is believed to require only cellular proteins and may involve an alternate form of AAV DNA replication (17, 22). Rescue also may involve a cellular enzyme and is enhanced by the presence of the first 13 bp of the AAV terminal repeat (10, 25). Packaging presumably requires the interaction of one or more of the AAV capsid proteins with the terminal repeats.

To determine whether the Rep proteins are present in the DNA-protein complex, we formed the complex by using fraction II in the presence of salmon sperm DNA and separated the complex from unbound material by electrophoresis on a native acrylamide gel. The complex was eluted from the gel and then denatured. Finally, the proteins present in the complex were separated on an SDS-acrylamide gel, transferred to nitrocellulose, and immunoblotted with anti-Rep and anticapsid antibodies (Fig. 10). The results indicated that the whole cell extracts from AAV-infected cells contain all four of the AAV Rep proteins and the three AAV capsid proteins. These proteins are not present in crude uninfected HeLa cell extracts or in extracts from cells infected with adenovirus. Fraction II contains little if any of the three AAV capsid proteins. We also could not detect the two smaller Rep proteins, Rep52 and Rep40, in fraction II. However, fraction II does contain substantial amounts of Rep78 and Rep68 and a third protein species which cross-reacts with anti-Rep antibody. We do not know the origin of this additional cross-reactive protein, but it is present in crude extracts as well and it occurs only in AAV-infected cells. Finally, when we examined the proteins present in the terminal complex, only the larger Rep proteins were detected. This finding suggested that Rep78 and Rep68 and, possibly, an additional Rep species were present in the DNA-protein complex.

There was no evidence that the AAV capsid proteins (61, 73, and 87 kilodaltons) comigrated with the complex (Fig. 10). Additional evidence that the AAV capsid proteins were not involved in the formation of the complex came from the study of an AAV proviral cell line, HN21 (22). This cell line contains an integrated AAV provirus which is *rep* positive but contains a deletion of the AAV capsid genes (*cap* negative). Superinfection of this cell line with adenovirus rescues the provirus and leads to production of the Rep proteins (D. McCarty and N. Muzyczka, unpublished data). Extracts prepared from adenovirus-infected HN21 cells were as proficient in making the terminal DNA complex as were extracts from wild-type AAV-infected cells (data not shown).

DISCUSSION

We have described factors which bind to the AAV terminal repeat and are present only in AAV-infected extracts. The DNA-protein complex is specific for the AAV terminal sequence and forms only when the terminal repeat adopts a covalently closed or hairpin configuration. The primary AAV sequence protected by the complex consisted of nearly the entire terminal palindrome, including the *trs* region. Similar experiments, using dimethyl sulfate protection (Arun Srivastava, personal communication), have essentially confirmed these results. Although we have not demonstrated it conclusively, our results suggest that at least one protein involved in binding is one of the larger AAV-coded Rep proteins (Rep78 and Rep68) or an antigenically similar protein. The relatively large region of DNA protected within the complex (60 to 90 bp) suggests that more than one molecule of Rep may be involved in the complex. The binding activity copurified with Rep78 and Rep68, and these larger Rep proteins were found to comigrate with the complex on neutral gels. In addition, the complex formed in the absence of detectable AAV capsid proteins. Taken together, these results suggest that the complex described here is likely to be involved in AAV DNA replication. We cannot exclude, however, the possibility that the complex is involved in DNA packaging or transcriptional control.

Secondary structure as an element in recognition. One of the biochemical problems in the mechanism of AAV terminal resolution is how the endonuclease which cuts at the *trs* region can recognize the cleavage site and also sense the secondary structure of the covalently joined end of the AAV RF molecule, which is approximately 60 bp away. In principle, cleavage of RF molecules which already have been resolved would be nonproductive and would interfere with the subsequent steps in AAV DNA replication. Therefore, the enzyme complex involved in initiating terminal resolution must be capable of recognizing both the sequences in the immediate vicinity of the *trs* as well as the secondary structure of the terminal repeat. The complex we have characterized essentially fulfills these requirements. Although most of the contact points would appear to be within the A palindrome, the complex does not form efficiently unless the contact points within the B and C palindromes are in a particular spatial orientation with respect to the *trs* region. Furthermore, formation of the complex requires that virtually the entire palindrome be intact; the *trs* region alone is not sufficient for binding.

A second issue in AAV terminal resolution is how the resolution complex determines which strand to cleave within the *trs* region, regardless of the orientation of the B and C palindromes (i.e., flip or flop). Inspection of the terminal sequences in the two orientations indicates that the strand which must be cleaved is always the one which contains the A sequence next to the D′ sequence (see Fig. 8), and presumably, the spatial orientation of the enzyme complex in this region should be unique. Yet, the spatial distributions of the B and C palindromes within the flip and flop orientations are not unique. In the flip orientation which was tested here (Fig. 8), the B palindrome is opposite the *trs*, whereas in the flop orientation (not shown), it would lie on the same side of the terminal repeat as the *trs* region. Because of the apparent assymmetry of protection within the terminal complex described here, we would predict that the contacts within the B and C palindromes are not specific for the particular base sequence but rely only on the secondary structures of the B and C palindromes. This prediction can be tested by synthesizing the flop configuration of NE DNA and examining the protected regions.

Finally, at least two complexes of different molecular weights were observed depending on the type of competitor DNA used in the binding experiments, salmon sperm or poly(dI-dC). However, the AAV sequences that were protected were identical in both cases. One possible explanation for this observation is that only one protein (or protein complex) is responsible for making primary contact with the DNA and that this initial complex then promotes the assembly of additional factors. The additional factors are, presumably, capable of binding to DNA themselves because they produce effective competition in the presence of salmon sperm DNA. These factors could include enzymes such as

DNA polymerases, helicases, single-stranded DNA-binding proteins, or topoisomerases. The exact nature of the proteins involved in binding should become apparent as the binding assay is used to purify the proteins to homogeneity.

Genetic studies of the terminal repeats. Our results are largely consistent with previous genetic studies of mutations within the terminal repeats. These studies indicate that the terminal repeats are essential for DNA replication (4, 18, 24–27). Deletion and substitution mutations in the C palindrome indicate that the 11-base sequence between the *Sma*I sites of this region can be changed provided the substitution preserves the secondary structure of the T-shaped hairpin (4, 18). The substitution mutations are often as effective for DNA replication as the wild-type sequence. The simple deletion, however, is severely defective for DNA replication (4, 26). This probably does not conflict with our finding that digestion of NE termini with *Sma*I does not prevent binding. In the experiment reported here, the ends of the *Sma*I-digested DNA were not covalently joined, so the T-shaped secondary structure of the remaining B and deleted C palindrome probably was maintained. In contrast, the *Sma*I deletion pSM1205 (26) and its derivatives (4) are probably incapable of forming the T-shaped structure in vivo because so little of the C palindrome is left. Instead, the terminal repeat of a simple *Sma*I deletion may exist in the form of a linear hairpin with a single-stranded substitution loop (see reference 4 for additional discussion). This question can be settled by examining the ability of various terminal deletion mutants to bind AAV-specific factors.

Role of Rep proteins. Mutations which affect exclusively the larger Rep proteins (Rep78 and Rep68) have been found to be defective for AAV DNA replication and the transactivation of AAV viral promoters (12, 15, 33). These mutations are also defective for repression of heterologous viral promoters, albeit this defect is not as severe as that in mutations which affect all four Rep proteins (14). The fact that either Rep78 or Rep68 comigrates with the terminal complex suggests that the larger Rep proteins probably have a direct role in AAV DNA replication. We can only speculate about the enzymatic function of the Rep protein in the terminal complex. One possibility is that the Rep protein is the site-specific endonuclease required for terminal resolution. Another possibility is that Rep protein promotes the assembly of cellular proteins required for AAV replication. Finally, computer analysis of the Rep amino acid sequence has revealed a similarity between Rep and procaryotic helicases as well as the presence of a purine nucleotide binding site (McCarty and Muzyczka, unpublished). This suggests the intriguing possibility that Rep functions as a site-specific helicase. In other experiments, we have used NE DNA to devise an in vitro functional assay for AAV terminal resolution (Snyder et al., in preparation) in which the ends of NE DNA are converted to an open duplex conformation. Hopefully, this assay will help resolve the question of Rep function. As yet, it is not clear whether the complex reported here is a prelude to resolution in this assay.

It is even less clear how Rep proteins might control transcription. In preliminary experiments thus far, we have been unable to detect binding of the larger Rep proteins to internal sequences containing the three AAV transcriptional promoters (D.-S. Im and N. Muzyczka, manuscript in preparation). One possible explanation is that the Rep proteins induce transcriptional activation by binding to the terminal repeat, but studies of plasmids in which the terminal palindrome has been deleted have not revealed a major effect on transcription (15). Alternatively, the Rep proteins may exert their effects on viral transcription indirectly by interacting with other factors that are bound to promoter regions, rather than by binding AAV sequences directly. In this sense, the larger Rep proteins would be analogous in function to the adenovirus E1A proteins.

ACKNOWLEDGMENTS

This work was supported by a Program Project grant from the National Cancer Institute (5 P01 CA2814607) and by Public Health Service grant R01GM3572302 from the National Institute of General Medical Sciences.

We thank A. Srivastava for reading the manuscript and for sharing unpublished information with us. We also thank R. Snyder, D. McCarty, and P. Hearing for a critical reading of the manuscript.

LITERATURE CITED

1. **Astell, C. R., M. B. Chow, and D. C. Ward.** 1985. Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J. Virol. **54:**171–177.
2. **Berns, K. I., and R. A. Bohenzky.** 1987. Adeno-associated viruses: an update. Adv. Virus. Res. **32:**243–306.
3. **Berns, K. I., and W. W. Hauswirth.** 1979. Adeno-associated viruses. Adv. Virus Res. **25:**407–449.
4. **Bohenzky, R. A., R. B. LeFebvre, and K. I. Berns.** 1988. Sequence and symmetry requirements within the palindromic sequences of the adeno-associated virus terminal repeat. Virology **166:**316–327.
5. **Cavalier-Smith, T.** 1974. Palindromic base sequences and replication of eukaryotic chromosome ends. Nature (London) **250:** 467–470.
6. **Challberg, M. D., and T. J. Kelly.** 1979. Adenovirus DNA replication in vitro. Proc. Natl. Acad. Sci. USA **76:**655–659.
7. **Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates.** 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. J. Virol. **60:**1085–1097.
8. **DeLange, A. M., M. Reddy, D. Scraba, C. Upton, and G. McFadden.** 1986. Replication and resolution of cloned poxvirus telomeres in vivo generates linear minichromosomes with intact viral hairpin termini. J. Virol. **59:**249–259.
9. **Fife, K. H., K. I. Berns, and K. Murray.** 1977. Structure and nucleotide sequence of the terminal regions of adeno-associated virus. Virology **78:**475–487.
10. **Gottlieb, J., and N. Muzyczka.** 1988. In vitro excision of adeno-associated virus DNA from recombinant plasmids: isolation of an enzyme fraction from HeLa cells that cleaves DNA at poly(G) sequences. Mol. Cell. Biol. **6:**2513–2522.
11. **Hauswirth, W. W., and K. I. Berns.** 1977. Origin and termination of adeno-associated virus DNA replication. Virology **78:** 488–499.
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:**329–333.
13. **Labow, M. A., and K. I. Berns.** 1988. The adeno-associated virus *rep* gene inhibits replication of an adeno-associated virus/simian virus 40 hybrid genome in cos-7 cells. J. Virol. **62:** 1705–1712.
14. **Labow, M. A., L. H. Graf, and K. I. Berns.** 1987. Adeno-associated virus gene expression inhibits cellular transformation by heterologous genes. Mol. Cell. Biol. **7:**1320–1325.
15. **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.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
17. **Laughlin, C. A., C. B. Cardellichio, and H. C. Coon.** 1986. Latent infection of KB cells with adeno-associated virus type 2. J. Virol. **60:**515–524.
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. **Lusby, E., R. Bohenzky, and K. I. Berns.** 1981. Inverted terminal repetition in adeno-associated virus DNA: independence of the orientation at either end of the genome. J. Virol. **37**:1083–1086.
20. **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.
21. **Maxim, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific cleavages. Methods Enzymol. **65**:499–560.
22. **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.
23. **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.
24. **Samulski, R. J., K. I. Berns, M. Tan, and N. Muzyczka.** 1982. Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. Proc. Natl. Acad. Sci. USA **79**:2077–2081.
25. **Samulski, R. J., L.-S. Chang, and T. Shenk.** 1987. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. J. Virol. **61**:3096–3101.
26. **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.
27. **Senepathy, P., J. Tratschin, and B. J. Carter.** 1984. Replication of adeno-associated virus DNA. Complementation of naturally occurring Rep^- mutants by a wild-type genome or an Ori^- mutant and correction of terminal deletions. J. Mol. Biol. **179**:1–20.
28. **Shlehofer, J. R., M. Ehrbar, and H. zur Hausen.** 1986. Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helper virus dependent parvovirus. Virology **152**:110–117.
29. **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.
30. **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.
31. **Tattersall, P., and D. C. Ward.** 1976. Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. Nature (London) **263**:106–109.
32. **Tratschin, J.-D., 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.
33. **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.
34. **Yacobson, B., T. Koch, and E. Winocour.** 1987. Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. J. Virol. **61**:972–981.