JOURNAL OF VIROLOGY, May 1976, p. 672-684

Vol. 18, No. 2
Printed in U.S.A.

Multiple Structures of Adeno-Associated Virus DNA: Analysis of Terminally Labeled Molecules with Endonuclease R·*Hae*III

DAVID T. DENHARDT,* SHLOMO EISENBERG, KATALINA BARTOK, AND BARRIE J. CARTER

Department of Biochemistry, McGill University, Montreal, Quebec, Canada H3G 1Y6, and Laboratory of Experimental Pathology, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014*

Received for publication 25 November 1975

The double-stranded form of adeno-associated virus (AAV) DNA has about 20 sites sensitive to endonuclease *R·Hae*III from *Haemophilus aegyptius*; the fragments produced fall into about 13 size classes, 8 of which contain single fragments. The location of the *Hae*III-produced AAV fragments relative to the three *Eco*R1 fragments was determined. Using revised figures for the molecular weights of the *Hae*III cleavage products of ϕX174 replicative form DNA, we calculated that AAV DNA contains about 4,000 nucleotides. After *Hae*III digestion of duplex DNA terminally labeled with ^{32}P using polynucleotide kinase, the majority of fragments containing a 5′ ^{32}P label were about 40 nucleotides in length, and fragments of similar size were generated from each end, suggesting that the *Hae* site closest to the end is within the terminal repetition. Two more-slowly-migrating cleavage products also bore 5′ ^{32}P end label. These three terminally labeled species were also generated from single-stranded AAV DNA by digestion with *Hae*III, and evidence that one may have a nonlinear ("rabbit-ear") structure is presented. The predominant 5′ terminal base was identified as thymine for both the plus and minus strands of AAV. Single-stranded AAV molecules could not be efficiently covalently circularized by incubation with polynucleotide ligase or ligase plus T4 DNA polymerase.

Adeno-associated virus (AAV) has a genome composed of approximately 4,200 nucleotides (4,000, according to our data), sufficient to code for about 150,000 daltons of protein if all the nucleotides serve a coding function (see reference 2 for a review). However, because only about 75% of the genome is transcribed (8) into stable RNA, the aggregate amount of protein that can be coded for is approximately 10^5 daltons. The fact that the largest protein in the virus has an apparent molecular weight of about 87,000 to 92,000 (22, 31) is consistent with the possibility that the AAV genome contains only one gene and codes for only one protein that is processed to yield the virion polypeptides.

For replication and for transcription, AAV is normally dependent on adenovirus functions of unknown nature (9); presumably, host cell functions are also required. Because of this dependence on adenovirus and host functions, AAV should provide a useful tool for investigation of the replication of mammalian DNA. It is likely that there are similarities between the mechanisms of replication of the AAV genome and the eucaryote chromosome. In particular, AAV may yield clues as to how the ends of linear molecules are replicated; since DNA polymerases are incapable of replicating the extreme 3′ ends of DNA templates, some provision must be made so that information is not lost at these ends (11, 12).

The genome of the AAV virion comprises one single-stranded linear molecule of either plus or minus polarity, and when the DNA is extracted from the virions the complementary strands will reassociate to form a duplex, if conditions permit (26, 30). Koczot et al. (23) and Berns and Kelly (3) established that the single-stranded viral molecules could form circles that could be seen in the electron microscope; these studies suggested the presence of an inverted terminal repetition of at least 1.5% of the genome. Studies of the properties of double-stranded AAV DNA led Gerry et al. (16) to conclude that AAV DNA contained a limited number of permutations (perhaps only two) and a terminal repetition (not inverted) representing about 1% of the genome.

It is evident that the structure of AAV DNA is complex, perhaps so that the linear single-stranded viral molecule can be replicated. We

have investigated aspects of the structure of AAV DNA using a combination of end-labeling with polynucleotide kinase and digestion with the restriction enzyme from *Haemophilus aegyptius*, endonuclease *R·Hae*III (27, 29), that cleaves the sequence $^{5'}\begin{smallmatrix}GGCC\\CCGG\end{smallmatrix}{}_{5'}$ at its center of symmetry. We have used the terminology suggested by Smith and Nathans (34) except that we have shortened it slightly for convenience. Endonuclease R·*Hae*III will be referred to as *Hae*III; the fragments produced by *Hae*III are designated A through M in this article, ordered according to their mobility on our acrylamide-agarose gels. Further work is required before it can be concluded that this order also corresponds to their molecular weights (36). Evidence is presented that AAV molecules have complementary sequences near each terminus, but that the terminal regions themselves, though self-complementary, are not always complementary to each other. This appears to lead to the formation of a terminal forked structure in a portion of the population.

MATERIALS AND METHODS

Viral DNAs. AAV 2 DNA was prepared from AAV 2 virions produced and purified as previously described (10). ^{3}H-labeled and ^{33}P-labeled ϕX174 replicative form (RF) DNA was purified (32) and was generously provided by C. Hours.

Enzymes. The restriction endonuclease R·*Eco*R1 was obtained from Miles Laboratories. The restriction endonuclease R·*Hae*III was purified as described (27), except that the nucleic acids were removed by precipitation with streptomycin sulfate rather than chromatography on Bio-Gel A 0.5 M. The spectrum of ϕX fragments generated by this nuclease preparation and the salt concentration at which the enzyme eluted from phosphocellulose indicated that the activity was primarily that designated *Hae*III (29). Bacterial alkaline phosphatase and T4 polynucleotide kinase were prepared as described (14a). The *Neurospora crassa* single-strand-specific exo- and endonucleases were prepared and used as described (1).

Centrifugation. AAV DNA was purified from enzyme reactions, when necessary, by velocity sedimentation in neutral 5 to 20% sucrose gradients containing 1 M NaCl, 1 mM EDTA, and 50 mM Tris-hydrochloride (pH 7.5). Samples (100 μl) were layered on 5-ml gradients and centrifuged for 4 h at 50 krpm in the SW50.1 rotor of the Beckman L265B ultracentrifuge. Fractions were collected through a hole poked in the bottom of the tube, and those fractions containing the DNA were pooled and diluted with an equal volume of 50 mM Tris-hydrochloride (pH 8.0), and the DNA was precipitated with 0.1 volume of 3 M sodium acetate (pH 5.5) and 2 volumes of isopropanol. Precipitation was allowed to occur for 12 h or more at −20 C, and the DNA was collected by centrifugation at 10 krpm in the HB4 rotor of the Sorval RC2 for 60 min. The DNA was dissolved in 100 μl of 10 mM Tris-hydrochloride (pH 8).

Labeling of the 5′ terminus. The DNA was incubated with 1 U of bacterial alkaline phosphatase per ml at 65 C for 15 min in 30 mM Tris-hydrochloride (pH 8). EGTA [ethylene-bis-(β-aminoethyl ether) *N*, *N*′-tetraacetic acid] was added to 6.5 mM, and the incubation at 65 C was continued for another 15 min (21). The solution was then made 1.5 mM in potassium phosphate (pH 6.8), 7 mM $MgCl_2$, 14 mM in β-mercaptoethanol, and 60 mM in Tris-hydrochloride (pH 7.4). ATP labeled in the gamma position (17) was added in a 100- to 1,000-fold excess over the molar amount of 5′ ends present (0.1 to 5 pmol). The specific activity ranged between 2×10^4 and 2×10^5 counts/min per pmol in the different experiments; the final concentration of ATP was maintained between 5 and 10 μM. The amount of polynucleotide kinase required to quantitatively label the 5′ ends of a control polynucleotide preparation in 60 min under these conditions was determined, and this amount was added at 0 min and again at 30 min. After 1 h at 36 C, EDTA was added to terminate the reaction, and the DNA was purified by velocity sedimentation in neutral sucrose gradients. This step resulted in the separation of the linear duplex molecules, used in most of this work, from the faster-sedimenting circular and oligomeric forms (7).

Cleavage with restriction enzymes. Endonuclease R·*Eco*R1 was used as described (6). Duplex AAV DNA (0.01 to 1 pmol) was digested overnight at 37 C in 10 mM each NaCl, $MgCl_2$, mercaptoethanol, and Tris-hydrochloride (pH 8) with sufficient *Hae*III to give complete digestion in 12 h. EDTA was added to 20 mM, bromophenol blue was added to 0.004%, and solid sucrose was added to about 0.5 M.

Electrophoresis. The composite acrylamide-agarose gels contained the designated percentage of acrylamide, *N*, *N*′ methylene-bis-acrylamide at 0.05 the concentration of acrylamide, 0.5% agarose (Seakem, product of Marine Colloids, distributed by Bausch and Lomb) and E buffer (14). Polymerization was initiated with 0.02% ammonium persulfate and 0.04% *N*,*N*,*N*′,*N*′-tetramethylethylenediamine. The gels were prerun for 30 min in E buffer, and then the DNA samples were injected onto the upper surface of the gels. The gels were 17 cm long in glass tubes with an internal diameter of 6 mm. The DNA was subjected to electrophoresis for 12 h at room temperature with a current of 3 mA induced by a 60-V potential difference across the electrodes. Under these conditions, the bromophenol blue dye marker migrated about 11 to 12 cm into the 17-cm gel. After completion of the electrophoresis, the gels were fractionated into 1- or 2-mm fractions using a Gilson Aliquogel gel fractionator. The fractions were digested with alkaline hydrogen peroxide overnight at 55 C and neutralized, and the radioactivity was determined with 10 ml of a Triton X-100–toluene–PPO (2,5-diphenyloxazole) cocktail in an Intertechnique scintillation spectrometer with appropriate settings and corrections for background and cross-talk where necessary. All gels are plotted with the cathode on the left so that the migration of the DNA fragments is from left to right.

Determination of the 5′-mononucleotides. The

DNA (about 0.1 μg) was digested with 100 μg of DNase I per ml (Worthington Biochem. Corp., EC 3.1.4.5, code D) for 45 min at 37 C in 5 mM $MgSO_4$, 50 mM Tris-hydrochloride (pH 8). Glycine buffer (pH 9.1) was added to a final concentration of 60 mM, and phosphodiesterase I (Worthington Biochem. Corp., EC 3.1.4.1, code VPH, treated [35] to inactivate the 5′ phosphatase) was added to 50 μg/ml. The digestion was continued for another 60 min and then frozen. A 10- to 15-μl sample of the digested DNA was applied to a washed (distilled water, 1 M LiCl, absolute methanol) polyethyleneimine-cellulose thin-layer plate (Macherey-Nagel Co., Cel 300 PEI) together with 20 μg each of the four deoxynucleotide monophosphates dissolved in water. The plate was developed in the first dimension with 1 M LiCl, washed with methanol, and then in the second dimension with 1 M acetic acid–3 M LiCl (9:1) (28). The location of the mononucleotide markers was determined by inspection under UV illumination. The entire plate was scanned for ^{32}P radioactivity by autoradiography, and the amount of radioactivity in the spots was quantitated in 10-ml of X-100–toluene–PPO scintillation fluid after elution of the mononucleotides with 1 ml of 0.5 M ammonium bicarbonate.

RESULTS AND DISCUSSION

Characterization of the fragments of the AAV duplex produced by *Hae*III. ^{32}P-labeled double-stranded AAV DNA was digested, together with [^3H]φX RF, with the restriction endonuclease R·*Hae*III from *H. aegyptius* as described in Materials and Methods. The products of the reaction were separated by electrophoresis in composite 4% acrylamide-0.5% agarose gels, and the distribution of radioactivity throughout the gel was determined by scintillation counting of the dissolved gel slices. Figure 1 shows the distribution of the ^{32}P-labeled AAV DNA fragments (solid circles) and the ^3H-labeled φX RF fragments (open circles) in the gel. The φX RF [^3H]DNA fragments were used as molecular-weight standards since their molecular weights have been determined (24); 11 φX

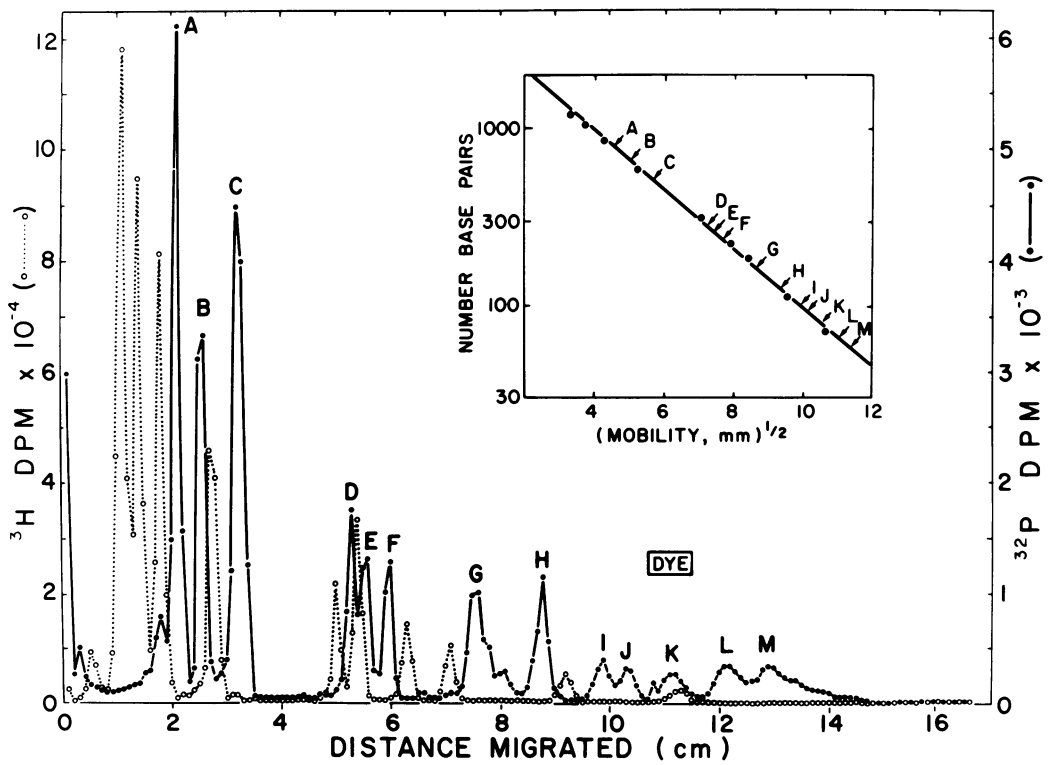

FIG. 1. *Acrylamide-agarose gel electrophoresis pattern of the HaeIII cleavage products of the AAV duplex. AAV duplex [^{32}P]DNA and φX174 RF [^3H]DNA were digested together with HaeIII and subjected to electrophoresis in a 4% acrylamide-0.5% agarose gel as described in Materials and Methods. Migration is from left to right; 1-mm fractions were collected. Symbols: (○) ^3H-labeled φX174 RF cleavage products, (●) ^{32}P-labeled AAV duplex cleavage products. Inset: Calibration curve of the size of the ^3H-labeled φX RF cleavage fragments (●) against the square root of the mobility. The size of the φX genome was taken as 5,500 base pairs (33).*

cleavage fragments were produced, two of which had very similar mobilities and were found in peak no. 6 at 5.4 cm.

The fragments produced from the AAV duplex fell into the 13 major mobility classes, identified by the letters A through M in order of increasing mobility as shown in Fig. 1. As illustrated in the inset to Fig. 1, an approximate linear relation exists between the logarithm of the size (e.g., the number of base pairs) of the ϕX markers and the square root of the distance each fragment has migrated into the gel. This allows an apparent molecular weight to be assigned to each of the fragments A through M, and these are given in Table 1; these molecular weights were calculated assuming that the correct size of the ϕX174 genome is 4,800 nucleotides (1a).

As a first step towards locating these fragments on the AAV genome, the *Hae*III restriction enzyme pattern of the three AAV fragments produced by digestion with endonuclease R·*Eco*R1 was determined. This nuclease cleaves the AAV duplex into three fragments: *Eco*R1·A, *Eco*R1·B, and *Eco*R1·C, with molecular weights of 1.6×10^6, 1.1×10^6, and 1.3×10^5, respectively (6). *Eco*R1·A is the right-hand terminus, *Eco*R1·B is the left-hand terminus, and *Eco*R1·C originates from the middle of the genome (7); they can be separated by sedimentation velocity centrifugation on neutral sucrose gradients. The *Hae*III digestion pattern of the intact duplex and the two largest fragments produced by *Eco*R1 are illustrated in Fig. 2. ϕX RF DNA was added before the digestion with R·*Hae*III to provide molecular-weight markers, and the digests were subjected to electrophoresis on composite 3% acrylamide–0.5% agarose gels. The three panels show, from top to bottom, the distribution of radioactivity in the gels of the fragments resulting from intact duplex, the *Eco*R1·A segment, and the *Eco*R1·B segment. The resolution in these gels is not as good as in Fig. 1 because the acrylamide concentration was lower, and 2-mm rather than 1-mm fractions were collected; nevertheless, the similarity of the two patterns obtained from the intact duplex (^{32}P labeled) in Fig. 1 and the intact duplex [radioactively labeled with [^{3}H]thymidine and density-labeled with 5-bromouracil to about 90% substitution] in Fig. 2 is apparent. The two peaks labeled T_A and T_B in the lower two panels contain the fragments having as one terminus the end produced by *Eco*R1. (Additional data in support of this assignment are presented below in Fig. 5 and 6.)

The molecular weights of the various fragments were determined from a calibration curve established using the known molecular weights of the *Hae*III fragments of ϕX RF present in the gel. In this case, as illustrated in Fig. 3, a good linear relation was obtained between the logarithm of the number of base pairs in the fragment and the mobility; when plotted versus the square root of the mobility, the fit was not as good. We do not have any explanation for why a plot against the square root in one case (Fig. 1, 4% acrylamide) but not in the other (Fig. 2, 3% acrylamide) allows a better fit, but Williamson (37) reported a similar phenomenon. It may be a function of the acrylamide concentration. The molecular weights estimated from the plots are tabulated in Table 1.

TABLE 1. *Characteristics of the cleavage products of duplex AAV DNA produced by endonuclease R·HaeIII*

*Hae*III size class	No. of base pairs[a]	No. of fragments[b]	*Eco*R1 assignment
A	700 (640)	1	
B	584 (550)	1	A
C	454 (445)	2	A, B
D	253 (280)	1	B
E	227	1	A
F	210 (214)	1	A
G	148 (140)	2	A, A
H	113 (105)	2	A, B
I	92 (83)	1	B
J	83 (76)	1	B
K	71 (63)	1 1/2(?)	B (1/2A?)
L	59 (51)	3	A, B, B
M	51 (41)	4	A, A, B, B
T_A	257	1	
T_B	183	1	

[a] The first number was determined from the square root relationship plotted in Fig. 1 (normal duplex, ^{32}P labeled); the numbers in parentheses were derived using the linear relations shown in Fig. 3 [bromouracil-substituted duplex, [^{3}H]thymine-labeled]. The number of base pairs is calculated assuming ϕX174 RF has 4,800 base pairs (1a).

[b] Depends on the assumption that migration rate is proportional to the molecular weight only.

The number of fragments derived from one genome in each peak can be calculated from an analysis of the total radioactivity present in each band; if only one fragment is present in each band, then the amount of radioactivity in that band should be proportional to the size of the fragment. The assumption implicit in this analysis is that the mobility of the fragment is proportional to the size of the fragment and is not affected by the base composition or sequence; this may be a poor assumption, especially for the smaller fragments. An analysis of the two gels of the intact AAV duplex shown in Fig. 1 and 2 is presented in Fig. 4. For the data in Fig. 1, a better linear relation is obtained

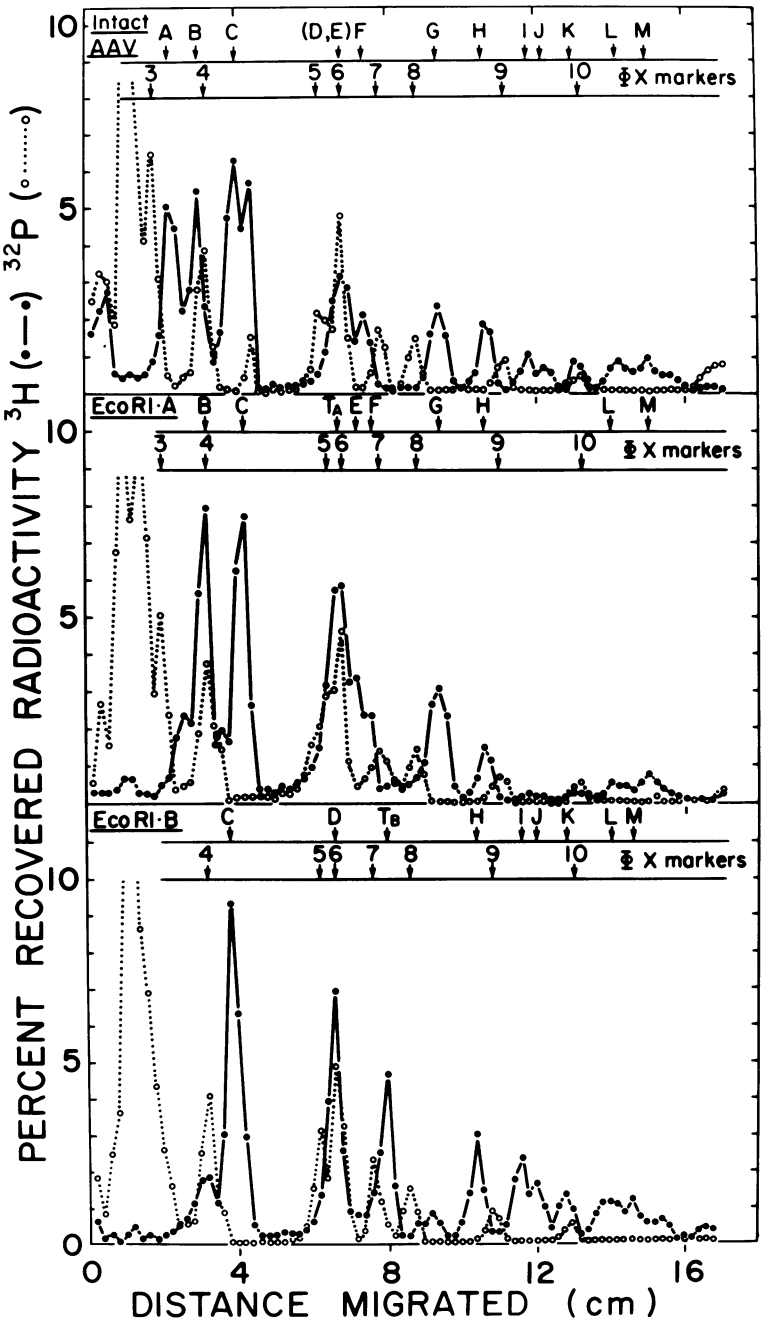

FIG. 2. *Acrylamide-agarose gel electropherogram of the HaeIII cleavage products of the AAV duplex and the EcoR1·A and EcoR1·B fragments. The AAV DNA preparations were labeled with [^{3}H]thymidine and substituted with 5-bromodeoxyuridine. ^{32}P-labeled ϕX RF was digested and subjected to electrophoresis simultaneously with the AAV DNA in a 3% acrylamide-0.5% agarose gel. The gel was fractionated into 2-mm portions. A to M and 3 to 10 indicate the distinctive cleavage products of the AAV duplex and ϕX174 RF fragments, respectively. T_A and T_B are the HaeIII cleavage products having one EcoR1-produced terminus. Symbols: (●) ^{3}H, (○) ^{32}P.*

when the logarithm of the total radioactivity in the band is plotted versus the square root of the distance migrated, whereas for the data of Fig. 2 a better fit is obtained when the mobility is used directly. From these plots, as well as similar plots of the data in panels B and C of Fig. 2 (data not shown), it appears that fragments A, B, D, E, F, I, and J are unique; only one fragment in each mobility class originates from the AAV duplex. There are two copies each of fragments C and H, one each derived from the *Eco*R1·A and *Eco*R1·B segments. Two fragments of the mobility class G are present in the duplex, and both are found in the *Eco*R1·A segment. The two smallest fragments, L and M, are difficult to quantitate accurately because of the broadness of the bands and evident heterogeneity. The best fit to all the gels is obtained, however, when it is assumed that there are three fragments in L and four in M, distributed as indicated in Table 1. Figure 1 also indicates that there may be one fragment running faster than M. The only fragment that is consistently ambiguous is K; possibly it is migrating anomalously because of an unusual base composition or possibly in the population of AAV duplexes a nonintegral number of fragments of that size are produced because of heterogeneity (note that the terminally labeled component beta [see below] migrates in this position) in the population. We cannot distinguish these alternatives at present.

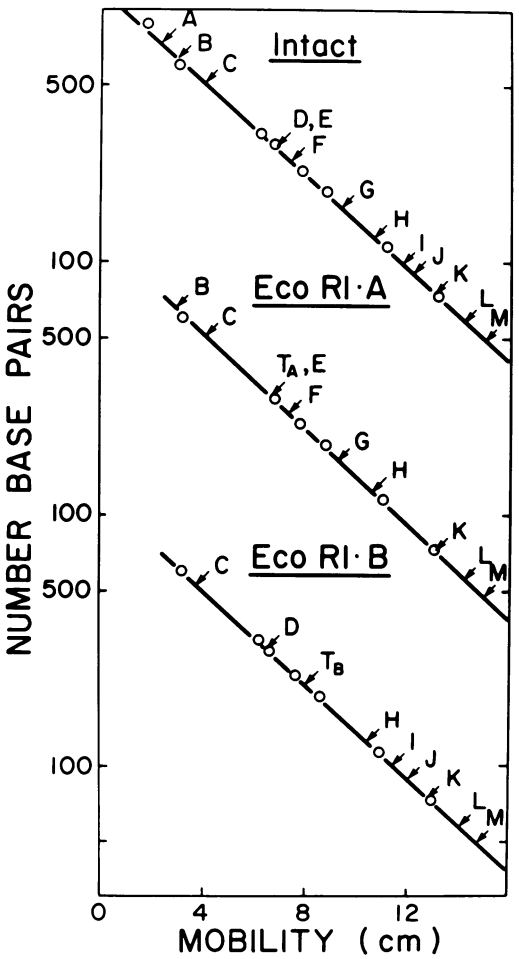

FIG. 3. *Size of the HaeIII cleavage products. From the data presented in Fig. 2, calibration curves were determined by plotting the number of base pairs in the φX174 fragments (assuming a total of 5,500 base pairs) against the distance each fragment migrated. From the position of each AAV cleavage product, the apparent number of base pairs in that fragment was estimated.*

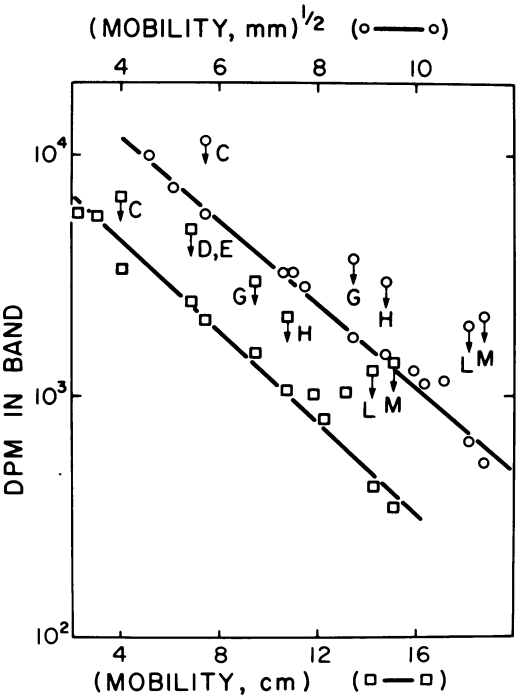

FIG. 4. *Number of fragments from one genome in each mobility class of AAV HaeIII cleavage products. The upper curve (○) is a plot of the radioactivity in each peak against the square root of the mobility; the data are from Fig. 1. The lower curve (□) is a plot of the radioactivity in each peak against the mobility; the data are from Fig. 2. Peaks C, G, and H (and D + E in the lower curve where they were not resolved) are replotted after division by 2; L is replotted after division by 3; and M is replotted after division by 4.*

Identification of the terminal fragments using polynucleotide kinase. The *Hae*III fragments of AAV DNA possessing the termini of the intact molecule were identified by labeling those termini with ^{32}P prior to the digestion with *Hae*III. The DNA was first dephosphorylated with alkaline phosphatase and then rephosphorylated using polynucleotide kinase and [γ-^{32}P]ATP. The terminally labeled molecules were digested with *Hae*III, and the resulting fragments were separated by electrophoresis in the composite acrylamide-agarose gels. The distribution of radioactivity in these gels is shown in Fig. 5. Most of the ^{32}P-labeled fragments derived from the intact molecules migrated on the leading edge of the peak containing the fragments in the size class M. Invariably a second peak containing 10 to 25% of the ^{32}P was observed to migrate just ahead of fragment G. Often there was also a third peak, very close to K, that contained 5 to 10% of the total ^{32}P. In most gels, at least 70% of the total ^{32}P label recovered was found in the two peaks near G and M. The apparent molecular weights of a duplex that would migrate at these positions were determined using the AAV fragments as standards; as shown in the inset (Fig. 5), a plot of log molecular weight versus mobility gave a reasonably straight line. A gel pattern similar (^{3}H distribution apparently identical; ^{32}P present in all three peaks in the above-mentioned proportions) to that shown in Fig. 5 was obtained from a digest of the circular and oligomeric DNA sedimenting ahead of the duplex DNA (equivalent to fractions 3 to 8 of Fig. 5a in reference 7). This suggests that the minor ^{32}P peaks are not derived from unusual, rapidly sedimenting molecules and that the oligomeric molecules have a similar set of terminal configurations.

The fact that between 60 and 80% of the ^{32}P from the intact duplex migrated in an appar-

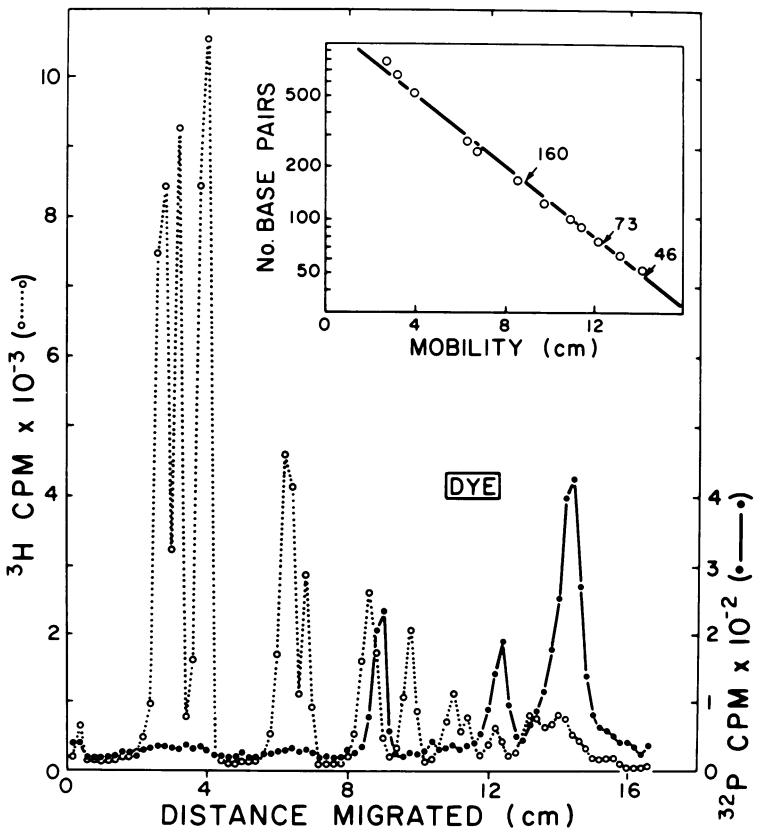

FIG. 5. *Acrylamide-agarose gel electropherogram of the cleavage products produced by digestion with HaeIII of terminally labeled AAV duplex DNA. The ^{3}H-labeled AAV duplex was labeled with ^{32}P at the 5'-termini using polynucleotide kinase, digested with HaeIII, and subjected to electrophoresis on 4% acrylamide-0.5% agarose gels as described in Materials and Methods. Symbols: (○) ^{3}H, (●) ^{32}P. Inset: Plot of the size of each AAV fragment versus its mobility. The arrows indicate the position and apparent size of the ^{32}P-labeled fragments.*

ently homogeneous peak suggested that the two *Hae*III terminal fragments were of the same size, and thus within the terminal repetition. This was confirmed by examining the digestion patterns of terminally labeled *Eco*R1·A and *Eco*R1·B fragments. The data are presented in Fig. 6; both *Eco*R1 fragments yielded ^{32}P-labeled *Hae* fragments migrating with the M mobility class, as well as the second component migrating ahead of G. An asymmetry in the labeling of the two ends is apparent. The termini produced by *Eco*R1, labeled T_A and T_B in the two panels, were labeled about 75% more efficiently than the "natural" end. This could be for conformational reasons, such as the presence of the protruding 5′ end produced by *Eco*R1 (19) or because the terminal A is labeled more efficiently than the terminal bromouracil (see below and reference 25). Comparison of the distribution of ^{32}P in these two panels also reveals that each fragment was contaminated with about 10% of the other fragment.

Location of fragment A in the genome. The size of the *Eco*R1·C fragment was obtained by subjecting it to electrophoresis together with ϕX RF markers. A value of 183 to 192 base pairs (assuming 4,800 nucleotides in ϕX DNA) was obtained depending on whether the data in Fig. 7 were plotted using the mobility or the square root of the mobility (the data fit equally well). After the *Eco*R1·C fragment was exposed to the *Hae*III nuclease, no change in mobility was detected, suggesting that there were no *Hae*III sites within that fragment; however, a site very close to an end would not have been detected. When the number of base pairs in *Eco*R1·C is added to the number of base pairs in *Hae*III·T_A and *Hae*III·T_B (Table 1), an aggregate size of 627 base pairs is obtained. This is close to the size of the fragment *Hae*III·A that was not present in either *Eco*R1·A or *Eco*R1·B. From this result and the data given in Table 1 we deduce the arrangement of fragments indicated in Fig. 8a.

Digestion of single-stranded AAV DNA with *Hae*III. The ability of single-stranded AAV DNA to form circular molecules (3) can be ex-

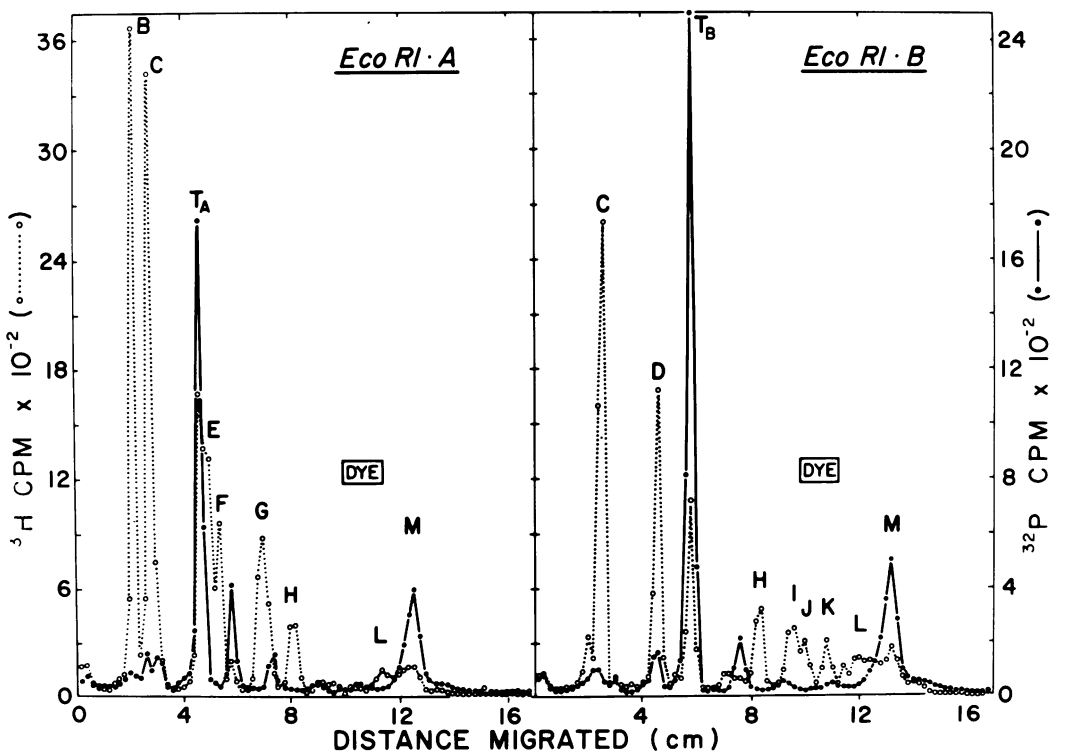

FIG. 6. *Acrylamide-agarose gel electropherogram of the cleavage products produced by digestion with HaeIII of terminally labeled EcoR1·A and EcoR1·B fragments. These fragments were purified from an EcoR1 digest of [^{3}H]thymidine-labeled, bromouracil-substituted AAV duplex. The EcoR1 cleavage fragments were labeled with ^{32}P using polynucleotide kinase, digested with HaeIII, and subjected to electrophoresis on 4% acrylamide-0.5% agarose gels as described in Materials and Methods. Symbols: (○) ^{3}H, (●) ^{32}P. The peaks are labeled as described in Fig. 2.*

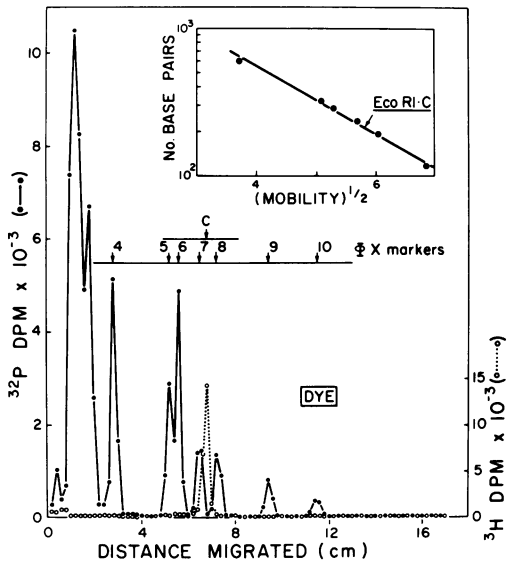

FIG. 7. *Size of the EcoR1·C fragment of AAV DNA. The ^{3}H-labeled EcoR1·C cleavage product was subjected to electrophoresis in 4% acrylamide-0.5% agarose gels as described in Materials and Methods. The HaeIII cleavage products of ^{32}P-labeled φX RF were included as markers and the data were plotted assuming a genome size of 5,500 nucleotides for φX174.*

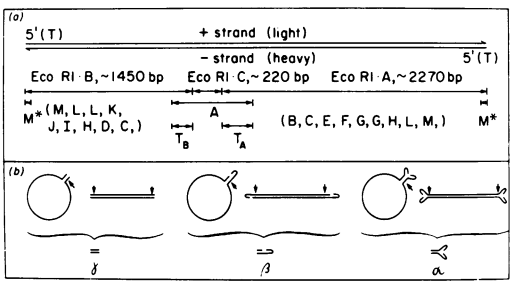

FIG. 8. *(a) Partial ordering of the cleavage products of AAV duplex DNA produced by HaeIII. The conventional representation of the AAV genome is shown at the top, and underneath the three EcoR1 cleavage products are drawn approximately to scale. The number of base pairs (bp) is calculated from the sum of the HaeIII fragments as given in Table 1. The location and relative sizes of A and the terminal fragments T_A, T_B, and M* are given. M* appears to migrate slightly ahead of M (Fig. 5). (b) Suggested structures for circular single-stranded and linear duplex AAV DNA. For reasons outlined in the text, the single-stranded molecules are believed to form the same terminal duplex structures as the linear duplex by base-pairing of the termini. Proposed locations of HaeIII cleavage sites are indicated by arrows. The various structures are not drawn to scale although α is the most-slowly-migrating component and γ is the fastest.*

plaind by the formation of structures like those drawn in Fig. 8b. If such structures are formed and if *Hae*III-sensitive sites are present in the duplex portions, then terminally labeled sequences should be cleaved from single-stranded DNA by the restriction enzyme and should have the same mobility as fragments derived from the duplex. Any single-stranded DNA fragments derived by cleavage of the single-stranded DNA by *Hae*III would be expected to migrate differently in these gels. To test these ideas, the experiment shown in Fig. 9 was performed.

In this experiment, ^{32}P-end-labeled single-stranded AAV DNA (the minus strand) was digested with *Hae*III and subjected to electrophoresis on gels using the same conditions employed for the duplex DNA. As can be seen (Fig. 9), essentially all of the ^{32}P was found in the same three peaks as were obtained when end-labeled duplex DNA was digested. If the

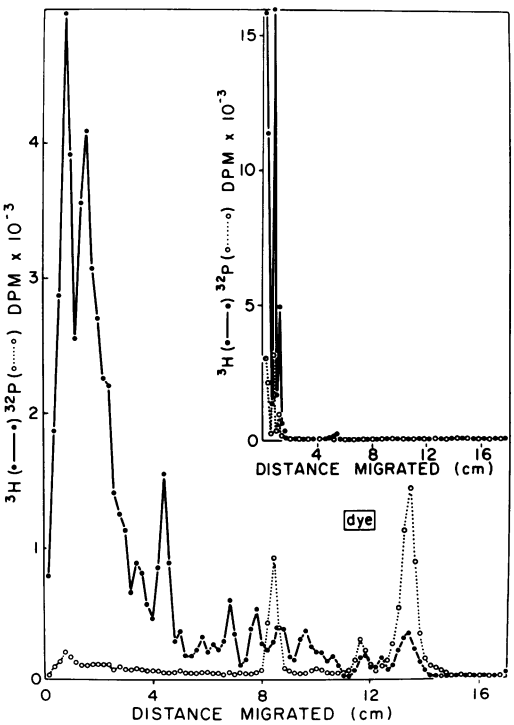

FIG. 9. *Acrylamide-agarose gel electropherogram of the HaeIII cleavage products of single-stranded ^{3}H-labeled AAV DNA (the minus strand) terminally labeled with ^{32}P using polynucleotide kinase. The DNA was digested and subjected to electrophoresis on 4% acrylamide-0.5% agarose gels using the same conditions as used in Fig. 5 and 6. The inset shows the electropherogram of the DNA before digestion with HaeIII. Symbols: (●) ^{3}H, (○) ^{32}P.*

DNA was not digested with the restriction endonuclease, no radioactivity entered the gel (Fig. 9, inset), thus providing strong evidence that the three peaks seen after *Hae* digestion were products of *Hae* action. We infer from these results that the same three types of fragments are generated by *Hae*III digestion of the single-stranded DNA as are generated from the duplex. Several discrete peaks of [^{3}H]DNA are also evident, suggesting that single-stranded AAV DNA, like f1 DNA (19) and ϕX DNA (5), can be cleaved at specific locations by *Hae*III. Since under our conditions the single-stranded DNA fragments migrated differently from the duplex fragments, we cannot with certainty determine which peaks correspond. A gel pattern similar to that seen in Fig. 9 was also obtained when terminally labeled plus-strand DNA was digested.

We propose that the three terminally labeled structures (alpha, beta, and gamma, in order of increasing mobility) seen in the gels shown in Fig. 5, 6, and 9 have the configurations alpha, beta, and gamma drawn in Fig. 8b. *Hae* sites located as indicated by the arrows would generate fragments of the various sizes found. We think the real situation is still more complex, but in the absence of more information further speculation is not useful. To determine if there was any single-stranded DNA in the terminal regions, the DNA was exposed to the single-strand-specific endonuclease from *N. crassa*, either before or after digestion with *Hae*III. In both experiments (data not shown), the amount of ^{32}P in all three positions was reduced to about the same extent. The loss of ^{32}P is mostly, if not entirely, the result of the fact that the nuclease will attack duplex DNA slowly from the ends, presumably as the result of "fraying." We conclude that at least by this criterion there were no single-stranded regions in the terminal fragments.

There are a number of experiments that can be done to check predictions based on the hypothetical structures illustrated in Fig. 8b. One of these is that the mobility of alpha should vary with the pore size of a gel in different fashion relative to beta, gamma, and normal linear duplex fragments. Because duplex DNA fragments migrate through the gel in an "end-on" fashion (13) the apparent molecular weight of alpha, relative to linear duplex markers, should increase as the acrylamide concentration is increased, since it will be more strongly retarded by virtue of its nonlinear structure. To test this prediction, a series of gels of different concentrations were run, and the apparent molecular weights of the three terminally labeled fragments were estimated relative to the known AAV fragments. The data are summarized in Fig. 10; in the 3.5% gel, alpha ran ahead of fragment G, whereas in the 6.5% gel it ran behind G. The other two fragments did not appear to change their mobility significantly. We therefore argue that a portion of both single-stranded and double-stranded AAV molecules have the "rabbit-ear" structure shown in Fig. 8b.

We deduce from these results that individual single-stranded AAV DNA molecules exist with two different terminal sequences, each of which is self-complementary and capable of forming "fold-back" structures, but which are not complementary to each other. In other words, a portion of the population does not have a complete inverted terminal repetition. If the terminal sequences were perfectly complementary to each other they would be expected to form the more stable single duplex (like gamma) rather than the double "fold-back" structure (alpha). It appears that about one-quarter of the molecules in the population contain non-complementary terminal sequences, although this could be an underestimate if the kinase labeling was not 100% effective. These results are not incompatible with the conclu-

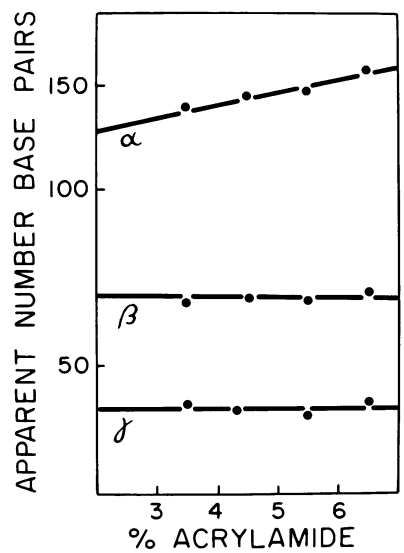

FIG. 10. *Apparent size of the terminally labeled fragments as a function of acrylamide concentration. ^{3}H-labeled AAV duplex DNA was terminally labeled with ^{32}P, digested with HaeIII, and subjected to electrophoresis on gels of different acrylamide concentrations, but all with 0.5% agarose. Calibration curves were established from the positions of the ^{3}H peaks (assuming an AAV genome size of 4,000 nucleotides), and the apparent size of the ^{32}P-labeled fragments was then estimated. α, β, and γ are believed to have the structures illustrated in Fig. 8b.*

sions (2, 4, 16) that there are two or more types of AAV molecules that differ by having a limited terminal permutation. However, it should be noted that for technical reasons it has not so far been possible to clone AAV.

Identification of the 5′ terminal nucleotide. In the initial attempt to identify the 5′ ends of AAV DNA, the plus and minus strands, separated by centrifugation in CsCl equilibrium density gradients after substitution of the thymine with 5-bromouracil, were individually investigated. The DNA was labeled with [γ-^{32}P]ATP using polynucleotide kinase and then digested with DNase I and venom phosphodiesterase. The resulting 5′ mononucleotides were resolved by two-dimensional thin-layer chromatography on polyethyleneimine-cellulose plates, and the ^{32}P was located on the plates by autoradiography. Both strands yielded similar autoradiograms. The spots from the minus-strand chromatogram were excised, and the radioactivity was determined. As can be seen from the data in Table 2, most of the ^{32}P was recovered in a region of the chromatogram that did not coincide with any of the four markers.

As a control, *Escherichia coli* DNA substituted with 5-bromouracil and uniformly labeled with ^{32}P was similarly digested and quantitated. Figure 11a shows an autoradiogram of the ^{32}P-labeled 5-bromouracil-substituted DNA; 5′ bromodeoxyuridylic acid migrated to the same location as the 5′ terminal nucleotide of the 5-bromouracil-substituted AAV DNA. That the terminal nucleotide of unsubstituted AAV DNA was thymine was confirmed by labeling the 5′ end of AAV (the duplex this time) not substituted with bromouracil; 50% of the recovered radioactivity was found in thymidylic acid. The autoradiogram is shown in Fig. 11b, and it can be seen that the only radioactivity detected was contained in the four common mononucleotides. The spots were excised, the nucleotides were eluted, and the radioactivity present was measured. The results are given in Table 2.

TABLE 2. *Determination of the 5′ ends of AAV DNA*

Nucleotide	% ^{32}P recovered	
	Expt 1[a]	Expt 2[b]
5′dAMP	5.9	13.6
5′dTMP	4.4	50
5′dGMP	22	18.5
5′dCMP	8.8	16.9
5′dBrUMP	59	

[a] The minus strand of 5-bromouracil-substituted AAV DNA was used. In this experiment, the ends were quantitatively labeled as calculated from the ^{3}H and ^{32}P specific activities. Autoradiograms of the plus and minus strands were similar, with most of the radioactivity in 5′-dBrUMP and a lesser amount in 5′dGMP.

[b] The AAV duplex not substituted with 5-bromouracil. In this experiment, 50% of the input ^{3}H (the DNA was labeled with [^{3}H]thymidine) and 46% of the input ^{32}P (representing 5′ terminal phosphates) was recovered in the four mononucleotides. The autoradiogram is shown in Fig. 11b.

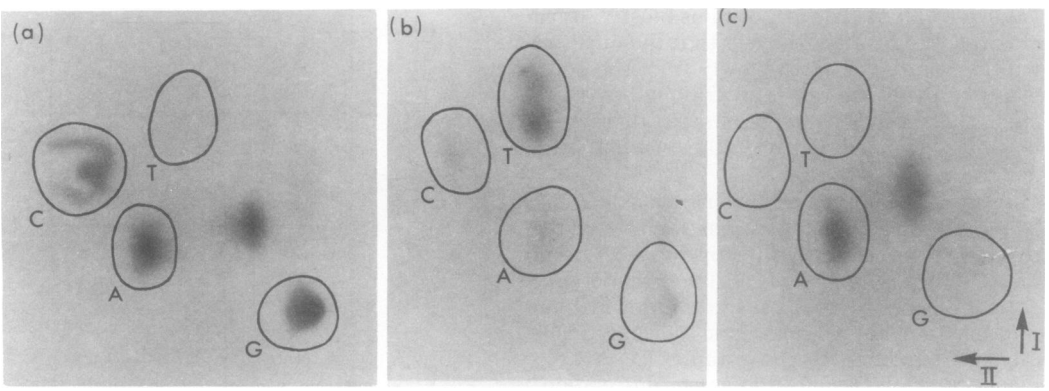

FIG. 11. *Autoradiograms of two-dimensional chromatograms of enzymatic hydrolysates of AAV DNA. Digestions with DNase I and phosphodiesterase I and two-dimensional chromatography on thin-layer polyethyleneimine-cellulose plates were performed as described in Materials and Methods. The reactions were complete since the only detectable radioactivity was in mononucleotides; less than 5% of the input radioactivity remained at the origin, and 50 to 75% of input radioactivity was accounted for as mononucleotides. (a) ^{32}P-labeled, 5-bromouracil-substituted E. coli DNA. (b) AAV duplex DNA not substituted with 5-bromouracil but labeled internally with [^{3}H]thymidine and labeled at the 5′ terminus with ^{32}P. (c) EcoR1·A restriction enzyme fragment terminally labeled with ^{32}P, substituted (about 90%) with 5-bromouracil, and labeled internally with [^{3}H]thymidine.*

TABLE 3. *Formation of protected 5' ends by incubation with DNA polymerase and DNA ligase*[a]

Prepn	φX*ts*7 DNA[b]		Random linear DNA[c]		AAV DNA[d]	
	lig	lig + pol	lig	lig + pol	lig	lig + pol
Control	2.5	4.9	2	1.6	3.35	6.8
Incubated	7	27.5	3.3	8.7	3.35	11.3
No phosphatase	205	206	91	89	43.5	59
% Protection	2.2	11	1.4	8.0	0	7.6

[a] The DNA, internally labeled with [^{3}H]thymidine and the 5' end labeled with ^{32}P, was incubated with polynucleotide ligase (lig) or polynucleotide ligase plus T4 DNA polymerase (lig + pol) and then treated with phosphatase (20). The numbers shown in the first three lines are the ratio of ^{32}P counts per minute to ^{3}H counts per minute after acid precipitation. The "control" was not incubated with ligase or ligase plus polymerase. The "No phosphatase" preparation was not reacted with phosphatase. "% Protection" is [(Incubated-Control)/No Phosphatase] × 100.

[b] The φX linear DNA was purified from phage grown under conditions of ligase deficiency in *E. coli ts7* (20). The specific activity was 8.1×10^4 counts/min per μg, and the P/H ratios given in the table have been multiplied by 10^3.

[c] The random linear φX DNA molecules were prepared by nicking circular single-stranded φX DNA with DNase I (15). Less than one nick per molecule was introduced, and the linear molecules were purified by sedimentation on alkaline sucrose gradients. The specific activity of the DNA was 2.5×10^3 counts/min per μg, and the P/H ratios given in the table have been multiplied by 10.

[d] The AAV DNA in this experiment was the 5-bromouracil-substituted H strand. Its specific activity was 4.4×10^4 counts/min per μg, and the P/H ratios in the table have been multiplied by 10^2.

In another experiment, the separated *Eco*R1·A and *Eco*R1·B fragments were terminally labeled with ^{32}P and examined. In both cases, two spots were observed on the autoradiograms: one in the position of 5'-deoxyadenosine monophosphate, as expected, since *Eco*R1 leaves a 5' A (18), and the second in the position of 5-bromodeoxyuridylic acid. An autoradiogram of the *Eco*R1·A chromatogram is shown in Fig. 11c. This experiment confirms that the 5' ends of both the plus and minus strands of AAV DNA are thymine. Consistent with the observation that T_A is labeled more efficiently than the natural end (Fig. 6a) is the observation that dAMP is somewhat more heavily labeled than deoxybromouridine monophosphate.

Circularization of AAV DNA. Iwaya et al. (20) described an experiment that suggested that linear single-stranded φX DNA molecules could be converted to circular molecules under appropriate conditions. The experiment involved isolating linear DNA from virions made under conditions of polynucleotide ligase deficiency, labeling the 5' ends with ^{32}P using polynucleotide kinase, and then demonstrating that some of the ^{32}P could be made resistant to bacterial alkaline phosphatase by incubation with T4 DNA polymerase and T4 DNA ligase. This could have occurred if a structure like beta in Fig. 8b were formed. To see if AAV could form such a structure, the same experiment was performed with single-stranded AAV DNA. The details of the experiment are given in Table 3. The result was that AAV DNA could be "circularized" to about the same extent (5 to 10%) as the linear φX DNA molecule. The criteria for circularization included resistance to alkaline phosphatase and velocity sedimentation in alkaline sucrose gradients to the position of circular DNA; this is not rigorous, however, since any dimers that might be formed by the polymerase-ligase reaction would also exhibit these properties. As a control, the same experiment was performed with linear φX DNA molecules having randomly located ends generated by nicking circular single-stranded DNA with DNase I; unexpectedly, protection of the 5' ends to about the same level was achieved. Thus it is not possible to conclude that the small amount of "circularization" observed for AAV DNA and the linear φX DNA from the ligase-defective host is evidence for any particular structure at the ends. This result also suggests that a majority of the molecules cannot have their termini in the β conformation.

ACKNOWLEDGMENTS

We thank H. E. David Lane for the ^{32}P-labeled, bromouracil-substituted *E. coli* DNA, Christian Hours for the ^{32}P-labeled φX174 RF DNA, Carol Kerr for the [γ-^{32}P]ATP, and Linda Pallett for typing the manuscript.

This research was supported by the National Cancer Institute of Canada, the Medical Research Council of Canada, and the United States Public Health Service.

LITERATURE CITED

1. **Bartok, K., B. Harbers, and D. T. Denhardt.** 1975. Isolation and characterization of self complementary sequences from φX174 viral DNA. J. Mol. Biol. 99:93-105.

1a. **Berkowitz, S. A., and L. A. Day.** 1975. Molecular weight of single-stranded bacteriophage fd DNA. High speed equilibrium sedimentation and light scattering measurements. Biochemistry 13:4825-4831.

2. **Berns, K. I.** 1974. Molecular biology of the adeno-associ-

ated viruses. Curr. Top. Microbiol. Immunol. 65:1-20.
3. **Berns, K. I., and T. J. Kelly, Jr.** 1974. Visualization of the inverted terminal repetition in adeno-associated virus DNA. J. Mol. Biol. 82:267-271.
4. **Berns, K. I., J. Kort, K. H. Fife, E. W. Grogan, and I. Spear.** 1975. Study of the fine structure of adeno-associated virus DNA with bacterial restriction endonucleases. J. Virol. 16:712-719.
5. **Blakesley, R. W., and R. D. Wells.** 1975. Single-stranded DNA from ϕX174 and M13 is cleaved by certain restriction endonucleases. Nature (London) 257:421-422.
6. **Carter, B. J., and G. Khoury.** 1975. Specific cleavages of adenovirus-associated virus DNA by a restriction endonuclease R·*Eco*R1-characterization of cleavage products. Virology 63:523-538.
7. **Carter, B. J., G. Khoury, and D. T. Denhardt.** 1975. Physical map and strand polarity of specific fragments of adeno-associated virus DNA produced by endonuclease R·*Eco*R1. J. Virol. 16:559-568.
8. **Carter, B. J., G. Khoury, and J. A. Rose.** 1972. Adeno-associated virus multiplication. IX. Extent of transcription of the viral genome in vivo. J. Virol. 10:1118-1125.
9. **Carter, B. J., F. J. Koczot, J. Garrison, J. A. Rose, and R. Dolin.** 1973. Separate function provided by adenovirus for adeno virus-associated virus multiplication. Nature (London) New Biol. 244:71-73.
10. **Carter, B. J., and J. A. Rose.** 1974. Transcription in vivo of a defective parvovirus: sedimentation and electrophoretic analysis of RNA synthesized by adeno-associated virus and its helper adenovirus. Virology 61:182-199.
11. **Cavalier-Smith, T.** 1974. Palindromic base sequences and replication of eucaryote chromosome ends. Nature (London) 250:467-470.
12. **Denhardt, D. T.** 1972. A theory of DNA replication. J. Theor. Biol. 34:487-508.
13. **Dingman, C. W., M. P. Fisher, and T. Kakefuda.** 1972. Role of molecular conformation in determining the electrophoretic properties of polynucleotides in agarose-acrylamide gels. Biochemistry 11:1242-1250.
14. **Edgell, M. H., C. A. Hutchison III, and M. Sclair.** 1972. Specific endonuclease R fragments of bacteriophage ϕX174 DNA. J. Virol. 9:575-582.
14a. **Eisenberg, S., B. Harbers, C. Hours, and D. T. Denhardt.** 1975. The mechanism of replication of ϕX174 DNA. XII. Non-random location of gaps in nascent ϕX174 RF II DNA. J. Mol. Biol. 99:107-123.
15. **Fiers, W., and R. L. Sinsheimer.** 1962. The structure of the DNA of bacteriophage ϕX174. III. Ultracentrifugal evidence for a ring structure. J. Mol. Biol. 5:424-434.
16. **Gerry, H. W., T. J. Kelly and K. I. Berns.** 1973. The arrangement of nucleotide sequences in adeno-associated virus DNA. J. Mol. Biol. 79:207-226.
17. **Glynn, I. M., and J. B. Chappell.** 1964. A simple method for the preparation of ^{32}P-labeled adenosine triphosphate of high specific activity. Biochem. J. 90:147-149.
18. **Hedgpeth, J., H. M. Goodman, and H. W. Boyer.** 1972. DNA nucleotide sequence restricted by the R1 endonuclease. Proc. Natl. Acad. Sci. U.S.A. 69:3448-3452.
19. **Horiuchi, K., and N. D. Zinder.** 1975. Site specific cleavage of single-stranded DNA by a *Hemophilus* restriction endonuclease. Proc. Natl. Acad. Sci. U.S.A. 72:2555-2558.
20. **Iwaya, M., S. Eisenberg, K. Bartok, and D. T. Denhardt.** 1973. Mechanism of replication of single-stranded ϕX174 DNA. VII. Circularization of the progeny viral strand. J. Virol. 12:808-818.
21. **Johnson, P. H., A. S. Lee, and R. L. Sinsheimer.** 1973. Production of specific fragments of ϕX174 replicative form DNA by a restriction enzyme from *Haemophilus parainfluenzae*, endonuclease HP. J. Virol. 11:596-599.
22. **Johnson, F. B., H. L. Ozer, and M. D. Hoggan.** 1971. Structural proteins of adeno-associated virus type 3. J. Virol. 8:860-863.
23. **Koczot, F. J., B. J. Carter, C. F. Garon, and J. A. Rose.** 1973. Self-complementarity of terminal sequences within plus or minus strands of adeno-associated virus DNA. Proc. Natl. Acad. Sci. U.S.A. 70:215-219.
24. **Lee, A. S., and R. L. Sinsheimer.** 1974. A cleavage map of bacteriophage ϕX174 genome. Proc. Natl. Acad. Sci. U.S.A. 71:2882-2886.
25. **Lillehaug, J. R., and K. Kleppe.** 1975. Kinetics and specificity of T4 polynucleotide kinase. Biochemistry 14:1221-1225.
26. **Mayor, H. D., K. Torikai, J. L. Melnick, and M. Mandel.** 1969. Plus and minus single-stranded DNA separately encapsidated in adeno-associated satellite virions. Science 166:1280-1282.
27. **Middleton, J. H., M. H. Edgell, and C. A. Hutchison III.** 1972. Specific fragments of ϕX174 DNA produced by a restriction enzyme from *Haemophilus aegyptius*, endonuclease Z. J. Virol. 10:42-50.
28. **Randerath, K., and E. Randerath.** 1967. Thin-layer separation methods for nucleic acid derivatives, p. 323-347. *In* L. Grossman and K. Moldave (ed.), Methods in enzymology, vol. 12A. Academic Press Inc., New York.
29. **Roberts, R. J., J. B. Breitmeyer, N. F. Tabachnik, and P. A. Myers.** 1975. A second specific endonuclease from *Hemophilus aegyptius*. J. Mol. Biol. 91:121-123.
30. **Rose, J. A., K. I. Berns, M. D. Hoggan, and F. Koczot.** 1969. Evidence for a single-stranded adeno-associated virus genome: formation of a DNA density hybrid on release of viral DNA. Proc. Natl. Acad. Sci. U.S.A. 64:863-869.
31. **Rose, J. A., J. V. Maizel, Jr., K. Inman, and A. J. Shatkin.** 1971. Structural proteins of adenovirus-associated viruses. J. Virol. 8:766-770.
32. **Schekman, R. W., M. Iwaya, K. Bromstrup, and D. T. Denhardt.** 1972. Mechanism of replication of ϕX174 DNA. III. An enzymic study of the structure of the replicative form II DNA. J. Mol. Biol. 57:177-199.
33. **Sinsheimer, R. L.** 1968. Bacteriophage ϕX174 and related viruses. Prog. Nucleic Acid Res. Mol. Biol. 8:115-170.
34. **Smith, H. O., and D. Nathans.** 1973. A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. J. Mol. Biol. 81:419-423.
35. **Sulkowski, E., and M. Laskowski.** 1971. Inactivation of 5′ nucleotidase in commercial preparations of venom exonuclease (phosphodiesterase). Biochim. Biophys. Acta 240:443-447.
36. **Thomas, M., and R. W. Davis.** 1975. Studies on the cleavage of bacteriophage lambda DNA with *Eco*R1 restriction endonuclease. J. Mol. Biol. 91:315-329.
37. **Williamson, R.** 1970. Properties of rapidly labeled DNA fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells. J. Mol. Biol. 51:157-168.