JOURNAL OF VIROLOGY, June 1979, p. 946-950
0022-538X/79/06-0946/05$02.00/0

Vol. 30, No. 3

Nucleotide Sequence of the Self-Priming 3' Terminus of the Single-Stranded DNA Extracted from the Parvovirus Kilham Rat Virus

LOIS A. SALZMAN* AND PHYLLIS FABISCH

Laboratory of Biology of Viruses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20014

Received for publication 17 January 1979

The parvovirus genome is a linear, single-stranded DNA molecule with double-stranded hairpin termini. The 3' terminus can serve in vitro as a self-primer for the synthesis of a double-stranded viral DNA intermediate. We have sequenced the nucleotides in the 3' terminus and propose a model for the secondary structure of the terminus and the in vitro origin of replication for the complementary viral DNA strand.

The viruses in the family Parvoviridae are composed of two or three proteins and one molecule of DNA. The DNA genome is single stranded and linear, with a molecular weight of 1.2×10^6 to 1.8×10^6 (18). Studies utilizing DNA nucleases, affinity chromatography, and electrophoresis on gels of viral DNA fragments have all indicated that the single-stranded genome has double-stranded 3' and 5' termini (2, 16, 17). The viral DNA contains approximately 5,000 nucleotides, and the double-stranded regions of 3' and 5' termini are each approximately 100 to 150 nucleotide base pairs in length (3, 17, 19). These double-stranded termini are believed to be folded back on themselves in a hairpin configuration (2, 3, 11, 17). Experiments with *Escherichia coli* or phage T4 DNA polymerases demonstrate that the 3' hairpin double-stranded DNA can serve as a self-primer for the synthesis of a complementary linear viral DNA strand (2, 19). A double-stranded viral DNA molecule has been reported in infected cells with the properties of terminal hairpins and has been proposed as an intermediate in the replication of the viral genome (8, 10, 15, 21). Because of the importance of the 3' terminus of the viral DNA in viral DNA replication and possibly in transcription, we have undertaken the sequence analysis of the 3'-terminal nucleotides. This analysis has given us confirmation of a 3'-terminal hairpin structure, a proposed model for the secondary structure of the terminus, and the nucleotide sequence of the DNA which at least in vitro serves as the origin of replication of the complementary strand. We have compared the 3'-terminal nucleotide sequence of the DNA from the autonomous parvovirus, Kilham rat virus (KRV), with the nucleotide sequence found at the 3' terminus of the DNA from the defective parvovirus, adeno-associated virus (AAV) (6). KRV was plaque purified three times and then used for infection of a rat nephroma cell line (20). The virus was freed of most of the cellular contamination by two or three isopycnic centrifugations in cesium chloride. DNA was extracted from KRV by sedimentation at pH 13 in a 4.0-ml gradient containing 5 to 20% sucrose, 0.6 M NaOH, and 0.4 M NaCl-0.001 M EDTA in an SW60 rotor at 56,000 rpm for 3.5 h at 15°C (16). A single peak of 16S DNA was neutralized with 1 M HCl, brought to a final molarity of 0.1 M Tris, pH 7.6, and precipitated overnight at −20°C in 2.5 volumes of ethanol.

The 3' terminus of KRV was labeled by using T4 DNA polymerase (purified through the phosphocellulose step [13] and a gift of N. Nossal). The enzyme (2 μg; 1,000 U/mg) was added to an incubation mixture containing α-^{32}P-labeled deoxynucleotide triphosphate (500 pmol; 300 Ci/mmol; New England Nuclear Corp.), 50 mM Tris-chloride (pH 8.8), 5.6 mM $MgCl_2$, 14 mM $(NH_4)_2SO_4$, 6.7 μM EDTA, 0.2 mg of bovine serum albumin per ml, 10 mM β-mercaptoethanol, and 8 μg of KRV DNA in a final total volume of 0.1 ml. The incubation was at 11°C for 110 min. The acid-soluble radioactivity was removed by adding the incubation mixture to a 0.7-ml column of preswollen Sephadex G-50 in a 1-ml disposable syringe. The syringe was placed in a polyallomer centrifuge tube (9/16 by 3.5 inches [ca. 1.4 by 8.9 cm]; Beckman Instrument Co.) and centrifuged in a Sorvall RC5 centrifuge at 3,000 rpm for 20 min. Approximately 80% of the acid-precipitable radioactivity applied to the column was recovered in the effluent. The effluent was then added to the top

of a 3.8-ml gradient of 5 to 20% sucrose containing 0.4 N NaCl-0.001 M EDTA and 0.6 N NaOH at pH 13 and centrifuged in a Beckman Spinco rotor (type SW60) for 3.5 h at 56,000 rpm and 15°C. Fractions of 0.2 ml were collected. The peak fraction of 16S viral DNA containing 60 to 70% of the applied radioactivity was neutralized with 1 N HCl and precipitated in a solution containing 0.3 M sodium acetate, 10 μg of tRNA, and 2.5 volumes of ethanol at −20°C. This DNA was used for the nearest-neighbor analysis (4), the sequencing of the molecule by the procedure of Maxam and Gilbert (12), and the exchange reaction described by Englund (5).

The terminal 3′ nucleotides were determined by the exchange reaction (5). The reaction depends on the fact that at 11°C, in the presence of a single nucleoside triphosphate, T4 DNA polymerase degrades duplex DNA only until a nucleotide which can be replaced by the supplied triphosphate is removed. The enzyme does not penetrate deeper into the DNA strand. If a labeled nonterminal nucleoside triphosphate is supplied to the enzyme incubation alone and in the presence separately of each of the other three nonradioactive triphosphates, the presence of an unlabeled triphosphate inhibits the incorporation of the isotope into acid-precipitable radioactivity. Only those deoxynucleotide triphosphates which precede the radioactive deoxynucleotide triphosphate inhibit incorporation of the radioisotope. As Table 1 shows, the presence of unlabeled triphosphates has little or no effect on the incorporation of [α-^{32}P]dGTP into acid-precipitable radioactivity, indicating that it might be the terminal nucleotide. This result is supported by the finding that the presence of unlabeled dGTP in all of the other incubations resulted in the inhibition of the acid precipitation of the radioactive triphosphate. The incorporation of [α-^{32}P]dATP into acid-precipitable radioactivity is inhibited by dGTP and dTTP. Thus, the terminal trinucleotides at the 3′ OH terminus of KRV DNA are probably ApTpG. However, the incorporation of [α-^{32}P]dTTP was inhibited only 55 to 60% by dGTP. This increased level of [α-^{32}P]dTTP was found repeatedly and suggests that a fraction of the DNA molecules may terminate in dTMP. If this is the case, the dGMP may be added to the terminal dTMP by T4 DNA polymerase because of a dCMP at position 114 which is complementary to the dGMP. KRV DNA may also have some heterogeneity at the 3′ terminus. Some molecules could have a terminal dGMP and some dTMP. Heterogeneity of one or two terminal nucleoside monophosphates has been reported for the defective parvovirus AAV (6). It is also possible that the T4 DNA polymerase adds on a nucleotide monophosphate to precede the terminal dGMP. This nucleotide monophosphate would be dTMP since it would have to hydrogen bond with the dAMP found at nucleotide 115 (Fig. 1).

Nearest-neighbor base sequence analysis was determined by digestion of the 3′ [α-^{32}P]DNA

TABLE 1. *Competition reactions and nearest neighbors*[a]

α-^{32}P-labeled dNTP added[b]	Radioactivity (cpm) in the exchange reaction with the following additional dNTP:					Nearest neighbor analysis 1		Nearest neighbor analysis 2 ([5′-^{32}P]-pCp added)[d]	
	None	dATP	dGTP	dCTP	dTTP	dNMP[c]	% cpm added	dNMP	% cpm added
dATP	3,800		100	3,200	225	dAMP	87	dAMP	2
dTTP	3,029	2,714	1,272	2,894		dAMP	91	dTMP	87
dGTP	3,500	3,700		3,628	3,597	dTMP	78	dGMP	11
dCTP	3,045	210	240		240	dAMP	11	dCMP	<1
						dTMP	84		

[a] The 3′ terminus of KRV DNA (5γ) was labeled with α-^{32}P-labeled deoxynucleotide triphosphate by using T4 DNA polymerase (gift of N. Nossal) as described in the text. The incubation mixture was in a total volume of 0.1 ml and was incubated at 11°C for 110 min. In the exchange reaction, nonradioactive deoxynucleotide triphosphate was added to the incubation where indicated. Acid-precipitable radioactivity was then determined. Background radioactivity from incubation in the absence of DNA was subtracted. In nearest-neighbor analysis 1, the 3′ terminus of KRV DNA was labeled with α-^{32}P-labeled deoxynucleotide triphosphate and T4 DNA polymerase as described above. In nearest-neighbor analysis 2, the 3′ terminus of KRV DNA was labeled with [5′-^{32}P]cytosine 3′,5′-biphosphate and T4 RNA ligase. In both analysis 1 (total counts per minute, 2×10^4) and analysis 2 (total counts per minute, 3×10^5) the radioactive DNA was digested with micrococcal nuclease and spleen phosphodiesterase. The products were separated by paper chromatography and located by UV light, and the radioactive areas were counted in a liquid scintillation counter. Background radioactivity was subtracted.

[b] dNTP, Deoxynucleotide triphosphate.

[c] dNMP, Deoxynucleotide monophosphate.

[d] pCp, Cytosine 3′,5′-biphosphate.

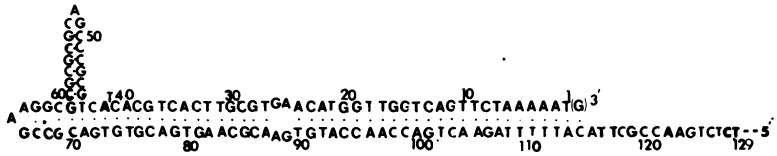

FIG. 1. *Nucleotide sequence of the 3′ terminus of KRV DNA in its most highly based-paired configuration. A, dAMP; G, dGMP; T, dTMP; and C, dCMP.*

with micrococcal nuclease (Worthington Biochemicals Corp.) and spleen phosphodiesterase (Worthington Biochemicals Corp.) to yield 3′ monophosphates (4). The 3′-terminal α-^{32}P is transferred from the 5′ position of the incoming nucleotide to the 3′ position of the adjacent nucleotide in the DNA chain. The 3′ monophosphates were separated by electrophoresis on Whatman 3MM paper (47 by 51 inches [ca. 1.2 by 1.3 m]) in 0.5% pyridine–0.88 M acetate buffer–1 mM EDTA (pH 3.5) at 2,300 V and 140 mA for 90 min at 4°C in a Savant high-voltage electrophoresis instrument. The radioactive spots were located after exposure to X-ray film (Kodak XR-2). The radioactive areas on the paper were cut out and counted in a Triton-toluene scintillation mixture. As Table 1 shows, the data from the nearest-neighbor analysis supports the 3′-terminal sequence of -ApTpG. dGMP is primarily next to dTMP; dTMP is next to dAMP, and dAMP is next to another dAMP. dCMP is not in the terminal tetranucleotide of -pApApTpG. The finding that dAMP may be the nearest base to dGMP again suggests some heterogeneity in the terminal 3′ sequence. As mentioned previously, heterogeneity of the 3′-terminal nucleotides has been reported for DNA extracted from the defective parvovirus AAV (6). In the case of AAV, the 3′ terminus is predominantly 3′AACCGGT- or 3′ ACCGGT-. The authors believe that this heterogeneity is due to a natural occurrence rather than to an artifact or nuclease activity (6).

In an effort to obtain a more unambiguous determination of the 3′-terminal nucleotide, we utilized the method of Hinton et al. (9). In this procedure a 2′ ribonucleoside 3′5′-biphosphate is joined by T4 RNA ligase in an ATP-dependent reaction to DNA. [5′-^{32}P]cytosine 3′,5′-biphosphate (pCp) (2,300 dpm/pmol) was a gift from S. Venkatesan. T4 RNA ligase (100 U/0.1 ml) was purchased from P. L. Biochemicals, Inc., Milwaukee, Wis. The incubation mixture contained 50 mM Tris (pH 8.3), 10 μM $MnCl_2$, 250 μM KRV DNA (16S), 250 μM ATP, 2.5 mM [5′-^{32}P]pCp, and 5 U of RNA ligase in a total volume of 0.02 ml and was incubated at 17°C for 120 h. The acid-soluble radioactivity was removed in a Sephadex G-50 column, and the nearest-neighbor analysis assay was carried out as described above. This assay demonstrates more clearly the heterogeneity of the 3′ terminus of KRV DNA (Table 1, nearest-neighbor analysis 2). Utilizing this method, we found that 80 to 90% of the KRV DNA molecules terminated in dTMP, 8 to 15% of the molecules terminated in dGMP, and less than 5% terminated in dAMP or dCMP. We have decided, therefore, to label dTMP as the terminal 3′ nucleotide and precede this nucleotide with dGMP in parentheses to denote that dGMP can be terminal in a minority of DNA molecules. The reason for the heterogeneity is still not known. We decided to continue the sequence analysis of KRV DNA by using the chemical method of Maxam and Gilbert (12). We have determined the sequence of 129 nucleotides in the 3′-terminal repetition region. The method utilizes four chemical reactions that cleave DNA preferentially at guanine (G > A), at adenine (A > G), at cytosine and thymine equally (C = T), and at cytosine alone (C). The products of the four reactions are resolved by size by using electrophoresis on polyacrylamide gels. The DNA sequence can be read from the pattern of radioactive bands on an autoradiogram. The sequences can be read from both double-stranded and single-stranded DNA after strand separation. We used KRV DNA prepared as described above and labeled the 3′ terminus with T4 DNA polymerase and [α-^{32}P]dGTP. The sequence gels for KRV DNA are shown in Fig. 2. We encountered some difficulty in discriminating between the C and T residues, as shown in nucleotides 99 to 119. The sequence in the original radioautograph was easier to interpret than the picture indicates. Repeated gel separation has led us to assign the nucleotide sequence with confidence. The sequence is similar to that found by D. Ward (personal communication). We have illustrated the sequence in the form which gives the maximal amount of base pairing when the DNA is folded back on itself (Fig. 1). Nucleotides 1 to 44 are paired with the nucleotides 71 to 114, with the exception of nucleotides 25, 26, 41, 88, and 89. Nucleotides 45 to 50 are paired with nucleotides 54 to 59 by hydrogen bonding, and nucleotides 60 to 63 are paired with nucleotides 67 to 70. Only 9 of the

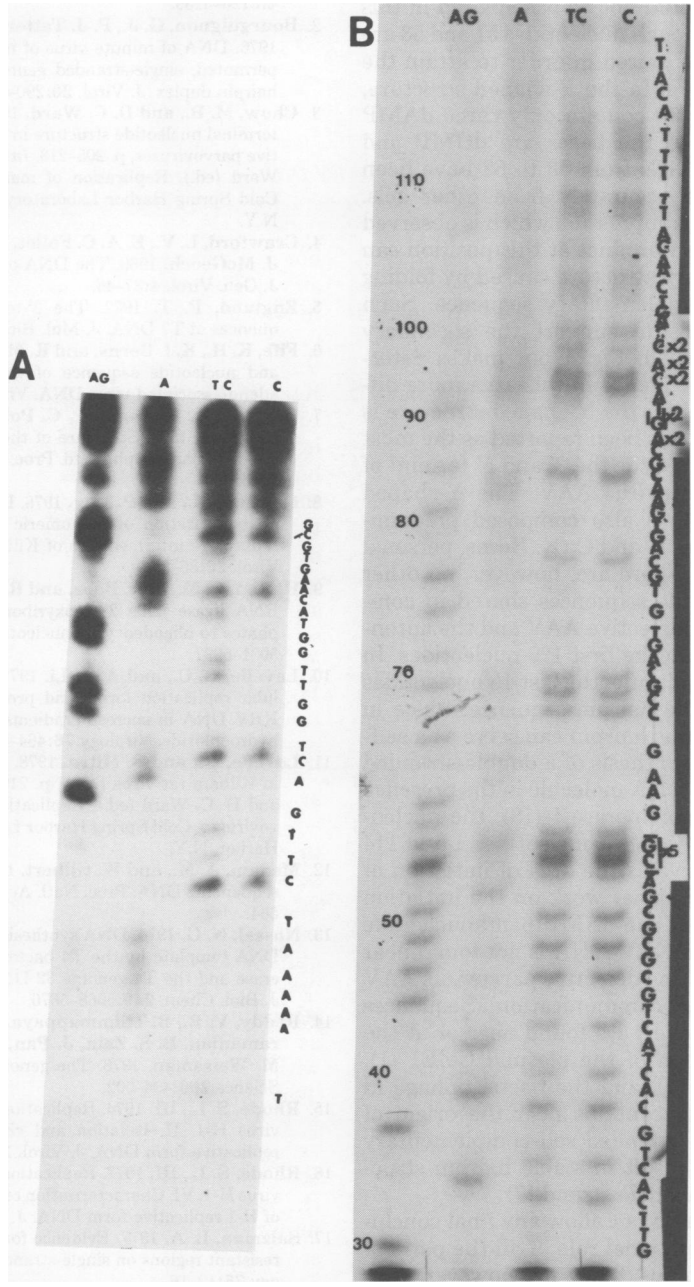

FIG. 2. *Autoradiogram of sequencing gels according to Maxam and Gilbert (12). KRV DNA labeled with ^{32}P on the 3′ terminus with a total of 40,000 to 80,000 cpm was divided into four portions for specific chemical reactions that cleave DNA at (i) guanine, (ii) adenine, (iii) cytosine and thymine, and (iv) cytosine alone. The reaction products were fractionated on gels containing 20% (A) or 10% (B) acrylamide, 0.6% methylene bisacrylamide, 7 M urea, 100 mM Tris-borate (pH 8.3), 1 mM EDTA, and 0.007% N,N,N′,N′-tetramethylethylenediamine. The samples were electrophoresed for 6 h at a constant voltage of 500 V (A) or 1,000 V (B). The gels were then placed in contact with Kodak X-ray film XR-2 (14 by 17 inches [ca. 35.6 by 43.2 cm]) at −70°C for 2 to 5 days and developed. Nucleotide sequence starts at the 3′ terminus.*

129 nucleotides could not be base paired in this arrangement, although nucleotides 51 and 53 are probably not base paired in order to attain the Y-shaped structure. In the Y-shaped structure, nucleotides 45 to 70 contain only three dAMP bases. The rest of the bases are dGMP and dCMP. The five GCs from 53 to 62 have been determined more accurately from other gels. This is an area of compression which is observed when the digestion product at this position can still have secondary structure caused by folding over of a self-complementary sequence. Such compression tends to support the secondary structure proposed in Fig. 1 but makes determination of the sequence in this area more difficult. The reason for the Y-shaped structure is not known, but it has been reported as the most reasonable structure for the 3′ and 5′ termini of the defective parvovirus AAV. The Y-shaped structure in AAV is also composed predominantly of the bases G and C (K. Berns, personal communication). There are, however, no other apparent nucleotide sequences shared in common between the defective AAV and the autonomous KRV within the first 129 nucleotides. In the KRV DNA molecule, the last 15 nucleotides are not part of the hairpin sequence. Since in vitro the 3′ terminus hairpin can serve as a self-initiation for the synthesis of a double-stranded covalently linked DNA molecule in the presence of *E. coli* DNA polymerase I (19), the nucleotides close to KRV DNA nucleotides 1 and 114 could possibly serve as the area of initiation of DNA synthesis. Further work on the initiation site is in progress. The dAMP pentanucleotide sequence found in KRV DNA is not found near the 3′ terminus of the defective parvovirus AAV (K. Berns, personal communication). Sequences of this type, however, are found near or at the origin of replication of the plasmid ColE1 (1), simian virus 40 (14), and the bacteriophage fd (7). Phage fd also contains near the origin of DNA replication extended self-complementary sequences which result in stable hairpin structures in the DNA single strand (7).

The above data do not allow any final conclusions as to the functional role of all the parts of the 3′ DNA terminus. They do, however, confirm the probability of a highly structured terminal sequence which is capable of folding back on itself and forming a hairpin turn. This hairpin-shaped double-stranded DNA could serve as the self-primer for initiation of the viral double-stranded replicative intermediate.

LITERATURE CITED

1. **Bastia, D.** 1977. The nucleotide sequence surrounding the origin of DNA replication of Col. E1. Nucleic Acids Res. 4:3123–3139.
2. **Bourguignon, G. J., P. J. Tattersall, and D. C. Ward.** 1976. DNA of minute virus of mice: self-priming, nonpermuted, single-stranded genome with a 5′-terminal hairpin duplex. J. Virol. **20**:290–306.
3. **Chow, M. B., and D. C. Ward.** 1978. Comparison of the terminal nucleotide structure in the DNA of non-defective parvoviruses, p. 205–218. *In* P. Tattersall and D. C. Ward (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Crawford, L. V., E. A. C. Follet, M. G. Burdon, and P. J. McGeoch.** 1969. The DNA of minute virus of mice. J. Gen. Virol. **4**:37–46.
5. **Englund, P. T.** 1972. The 3′-terminal nucleotide sequences of T7 DNA. J. Mol. Biol. **66**:209–224.
6. **Fife, K. H., K. I. Berns, and K. Murray.** 1977. Structure and nucleotide sequence of the terminal regions of adeno-associated virus DNA. Virology **78**:475–487.
7. **Gray, C. P., R. Sommer, C. Polke, E. Beck, and H. Schaller.** 1978. Structure of the origin of DNA replication of bacteriophage fd. Proc. Natl. Acad. Sci. U.S.A. **75**:50–53.
8. **Gunther, M., and P. May.** 1976. Isolation and structural characterization of monomeric and dimeric forms of replicative intermediates of Kilham rat virus DNA. J. Virol. **20**:86–95.
9. **Hinton, C. M., J. A. Baez, and R. I. Gumport.** 1978. T_4 RNA ligase joins 2′ deoxyribonucleoside 3′5′-biphosphates to oligodeoxyribonucleotides. Biochemistry **17**: 5091–5097.
10. **Lavelle, G. C., and A. T. Li.** 1977. Isolation of intracellular replication forms and progeny single-strands of KRV DNA in sucrose gradients containing guanidine hydrochloride. Virology **76**:464–467.
11. **Lavelle, G., and S. Mitra.** 1978. Double-helical regions in Kilham rat virus DNA, p. 219–229. *In* P. Tattersall and D. C. Ward (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
13. **Nossal, N. G.** 1974. DNA synthesis on a double-stranded DNA template by the T4 bacteriophage DNA polymerase and the T4 genome 32 DNA unwinding protein. J. Biol. Chem. **249**:5668–5676.
14. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, M. L. Celma, and S. M. Weissman.** 1978. The genome of simian virus 40. Science **200**:494–502.
15. **Rhode, S. L., III.** 1974. Replication process of the parvovirus H-1. II. Isolation and characterization of H-1 replicative-form DNA. J. Virol. **13**:400–410.
16. **Rhode, S. L., III.** 1977. Replication process of the parvovirus H-1. VI. Characterization of a replication terminus of H-1 replicative-form DNA. J. Virol. **21**:694–712.
17. **Salzman, L. A.** 1977. Evidence for terminal S_1 nuclease-resistant regions on single-stranded linear DNA. Virology **75**:12–16.
18. **Salzman, L. A.** 1978. The parvoviruses, p. 539–589. *In* D. P. Nayak (ed.), The molecular biology of animal viruses. Marcel Dekker, Inc., New York.
19. **Salzman, L. A., P. Fabisch, R. Parr, C. Garon, and T. Wali.** 1978. In vitro synthesis of double-stranded DNA from the Kilham rat virus single-stranded DNA genome. J. Virol. **27**:784–790.
20. **Salzman, L. A., and L. A. Jori.** 1970. Characterization of the Kilham rat virus. J. Virol. **5**:114–122.
21. **Salzman, L. A., and W. White.** 1973. In vivo conversion of the single-stranded DNA of Kilham rat virus to a double-stranded form. J. Virol. **11**:299–305.