JOURNAL OF VIROLOGY, Mar. 1982, p. 990–999
0022-538X/82/030990-10$02.00/0
Vol. 41, No. 3

DNA Sequence of the 5′ Terminus Containing the Replication Origin of Parvovirus Replicative Form DNA

SOLON L. RHODE III* AND BELIA KLAASSEN†
Institute for Medical Research of Bennington, Bennington, Vermont 05201

Received 6 August 1981/Accepted 13 October 1981

The nucleotide sequence of the 5′ terminus of the parvovirus H-1 was determined. There are two orientations of the 242-base-pair terminal palindrome in native replicative form DNA, one inverted with respect to the other. Adjacent to the terminal palindrome is an AT-rich region that is noncoding and contains a 55-base-pair tandem repeat. The addition mutant of H-1, DI-1, was also sequenced in this region and shown to have three copies of the tandem repeat sequence. Similarly, the related parvovirus H-3 contains only one copy of this repeat sequence. This region contains the replication origin for parvovirus replicative form DNA replication. Some of the implications of these results are discussed.

The nondefective parvoviruses are small icosahedral viruses that contain a linear single-stranded DNA of about 1.6×10^6 daltons (13, 34, 36). The virion DNA is predominantly, but not entirely, minus strand and is the template for all virus-induced mRNA's that have been identified (9, 30). During the replication process, the virion single-stranded DNA is converted to a linear duplex replicative form (RF) DNA (16, 35). This RF DNA undergoes amplification by a semiconservative mode of DNA replication that requires a viral RF rep gene product which has not been characterized (16, 22). The RF DNA serves as a template for viral RNA synthesis and the synthesis of single-stranded minus strand DNA that becomes encapsidated into progeny virions (17, 37).

The virion DNA contains self-complementary sequences at both its 3′ and 5′ termini, and RF DNA has been shown to contain these terminal palindromes in both extended and foldback configurations (3, 5, 19). The 5′ termini of both the plus and minus strands of RF DNA are covalently bound to a protein(s) whose function is unknown (15). The pools of RF DNA extracted from infected cells are also characterized by a significant fraction (≃25%) of dimer-length molecules that are linked at the 3′ terminus (18). The 3′ termini of several parvoviruses have been sequenced, and evidence was discussed which indicated that the orientation of the 3′-terminal palindrome is unique, that is, not inverted, in both virions and RF DNA (1, 2).

In this report we present the primary structure of the 5′ terminus of the parvovirus H-1 and portions of the terminal regions of DI-1, an addition mutant of H-1, and H-3, a parvovirus related to H-1. This region of the molecule has previously been shown by us to contain the replication origin for RF DNA replication (26, 27). Our evidence indicates that the 5′-terminal palindrome is present in two orientations in RF DNA, one inverted with respect to the other. This flip-flop relationship suggests that the mechanism for synthesis of 5′ termini of DNA proposed by Cavalier-Smith applies to this end of parvovirus RF DNA (4). Finally, we present a model for RF DNA replication that is compatible with the currently available facts pertaining to this process.

MATERIALS AND METHODS

Materials. Restriction endonucleases were obtained from New England Biolabs or Bethesda Research Laboratories and were used as directed by the supplier. The large fragment of *Escherichia coli* DNA polymerase I (pol I) was obtained from Boehringer-Mannheim. T4 DNA ligase was from New England Biolabs. The deoxynucleotide triphosphates and dideoxynucleotide triphosphates were from P-L Biochemicals, Inc. The *Hin*dIII linker and the M13-specific primer were purchased from Collaborative Research, Inc. [α-^{32}P]dATP was from New England Nuclear.

Cells and viruses. The parvoviruses used in this study were wild-type H-1, H-1 DI-1, and H-3 (20). Virus and RF DNA were obtained from infected NB cells as previously described (19). M13 mp2, mp5, and mp7, and JM103, were kindly provided by J. Messing.

Molecular cloning of parvovirus DNA. Approximately 20 μg of parvovirus RF DNA was digested with the appropriate restriction endonuclease as previously described (19). The reactions were terminated by the addition of 50 mM Tris, pH 8.0, 0.1 M NaCl, 1 mM EDTA, and 0.2% Sarkosyl, to a final volume of 200 μl, and were extracted once with phenol and once with

† Laboratorium voor Fysiologische Scheikunde, Leiden, The Netherlands.

phenol-chloroform-isoamyl alcohol (25:24:1). Phenol was removed by ether extraction; the solution was adjusted to 0.3 M Na^+ with 3 M sodium acetate, pH 5.5, and the DNA was precipitated with ethanol. *Hin*dII or *Hin*cII digests were ligated to a 10-mer *Hin*dIII linker as described (8). In some cases the linkers were phosphorylated with unlabeled ATP, and the mobility shift of the fragment during its isolation by gel electrophoresis provided a monitor of the effectiveness of the linker ligation step. Reactions were carried out in 66 mM Tris, pH 7.4, 6.6 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM spermidine, 0.5 mM ATP, 3 μM linker, 50 μg of bovine serum albumin per ml, and 10 to 20 U of T4 DNA ligase per ml at 14°C for 20 h. The reactions were stopped by heating at 65°C for 10 min. The reaction mixture was adjusted to 100 μl and 60 mM NaCl–0.03% sodium azide and was incubated with 10 U of *Hin*dIII for 16 h at 37°C. The digestion was terminated by the addition of EDTA to 20 mM, and the DNA was precipitated with ethanol. The restriction fragments were isolated by electroelution after fractionation by electrophoresis on a gel of 5% acrylamide. Slightly broadened bands were obtained at the gel positions expected after the addition of the linker sequences. After ethanol precipitation, the DNA fragments were ligated to *Hin*dIII-cleaved pBR322 or, in the case of *Hin*cII 83/91, directly into M13 mp5 by standard methods (8). *E. coli* HB101 was transformed by the method of Dagert and Ehrlich (6). Recombinant plasmids were identified by screening the Ap^r Tc^s colonies for the proper size insertions by gel electrophoresis. Plasmid DNA was prepared by the extraction method of Guerry et al. (10), concentrated by precipitation with 10% polyethylene glycol, extracted with phenol, and purified by centrifugation in CsCl/ethidium bromide gradients. Parvovirus fragments were subcloned in M13 mp5 by digestion of 0.5 to 1 μg of plasmid with *Hin*dIII and ligation of the fragments to 50 to 100 ng of *Hin*dIII-cleaved M13 mp5 RF DNA. We did not recover any pBR322-M13 mp5 recombinants.

The *Hin*cII 95.7/99.9 fragment of H-1 was isolated from pBR322 by gel electrophoresis and restricted with *Taq*I; the single-stranded ends were filled in by treatment with the large fragment of *E. coli* pol I. *Hin*dIII linkers were added as described above, and the fragments were isolated by gel electrophoresis and finally were cloned in M13 mp5.

The 87/97.5 fragment of H-1 was obtained by cloning the *Taq*I 38/97.5 fragment in M13 mp7 at the *Acc*I site. RF DNA of this recombinant was restricted with *Sau*3A, which cuts H-1 at map position 87 and M13 mp7 at its *Bam* site. The proper fragment was isolated by gel electrophoresis and cloned at the *Bam* site of M13 mp7.

The strands represented in the M13 clones were determined by incubating approximately 1 μg of recombinant single-stranded DNA with 1 μg of H-1 minus strand DNA in 5× SSC (SSC: 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C for 1 to 2 h. Plus strand recombinants hybridize to H-1 DNA and show a marked reduction of mobility in 1.6% agarose gels. Some clones contained tandem repeats of the fragment. Inverted repeats were not recovered. Organisms containing parvovirus recombinant DNA were propagated under P1-EK1 conditions in accordance with the Guidelines for Recombinant DNA Research.

DNA sequencing. DNA sequencing was done by the chain-terminating method of Sanger et al. (23, 24). We used either the 96-base-pair (bp) primer as described (11) or a 12-bp synthetic primer. Electrophoresis was in 0.3-mm gels of 8% acrylamide at about 50°C (29). Most gels were 27 by 40 cm, but a few were 15 by 60 cm. Samples were applied in two loadings 3 h apart.

RESULTS

Orientation of the 5′-terminal palindrome. The physical map of H-1 RF DNA was determined for the restriction endonucleases *Hha*I and *Taq*I by partial digestions and reciprocal digestions with enzymes having known sites, as previously described (19) (Fig. 1). These studies revealed two small fragments generated by both of these enzymes that were present in less than equimolar amounts and which mapped to the 5′ terminus of the molecule. This can be illustrated by the double digestion of RF DNA with *Hha*I and *Hpa*II (Fig. 2). *Hpa*II is known to have a number of cleavage sites, all located in a small area of the 5′-terminal palindrome (19). The two *Hha*I fragments of about 140 and 100 bp were both cleaved by *Hpa*II, localizing them to the right end. Isolation of the *Hpa*I 91/100 fragment and then digestion with *Hha*I produced four fragments of 100, 140, 340, and 380 bp (data not shown). The sums of the first plus fourth and the second plus third of these equal the original 480-bp 91/100 fragment. The simplest interpretation of these results is that the 5′ terminus has *Hha*I sites at either 140 or 100 nucleotides from the end in approximately equal proportions. Furthermore, we would predict that although these sites are within the terminal foldback they are not in a palindromic portion of it which has its axis of symmetry at nucleotides 115 through 120. This interpretation was confirmed by demonstrating that *Hha*I did not cleave this region when in the foldback configuration (data not shown) and more directly by DNA sequencing as described below. A similar result was obtained with *Taq*I except that its cleavage sites were located about 135 or 105 nucleotides from the 5′ terminus. Thus, these data indicate that the 5′-terminal foldback region of H-1 exists in two orientations inverted with respect to each

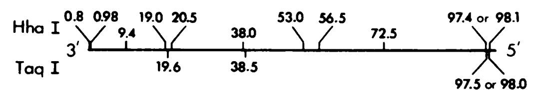


FIG. 1. Physical map of wild-type H-1 RF DNA for the restriction endonucleases *Hha*I and *Taq*I. It is assumed that the terminal palindromes are extended, with map position 0 at the 3′ end of the minus strand and 100 at the 5′ end. The *Hha*I sites at 0.8 and 0.98 were obtained from the DNA sequence (1). The remaining sites were derived by partial digestion and double digestions with enzymes with known cleavage sites.

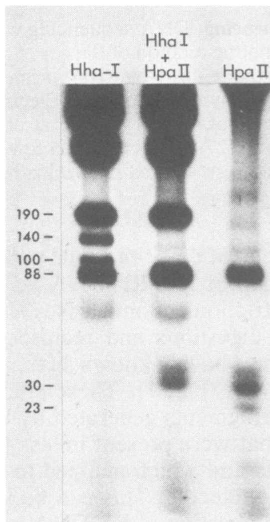


FIG. 2. Electrophoresis pattern of ^{32}P-labeled H-1 RF DNA fragments produced by: *Hha*I, *Hha*I + *Hpa*II, and *Hpa*II. Fragment sizes were estimated by using a *Hae*III digest of φX174 RF DNA in an adjacent lane. Electrophoresis was in a 15% acrylamide gel with a 1-cm top layer of 5% acrylamide containing the sample slots for 6 h at 150 V. The gel was stained with ethidium bromide, photographed, and then dried for autoradiography.

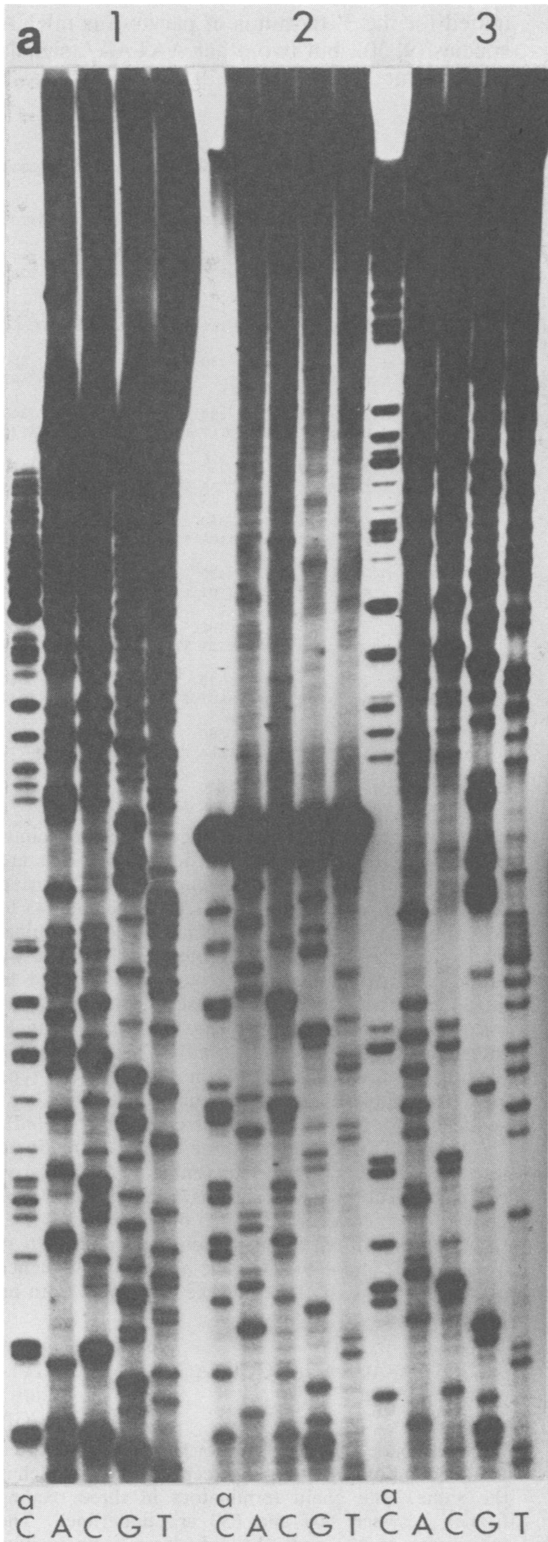
other such that there are two alternative positions for a *Hha*I or *Taq*I site that are equidistant from the axis of symmetry.

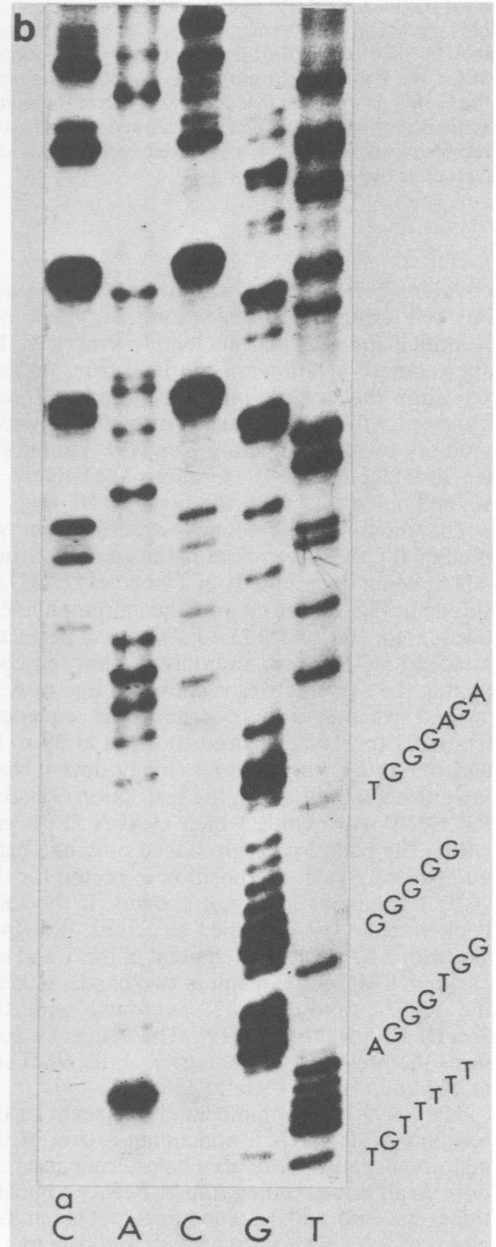
DNA sequence of the 5′ terminus of H-1. To determine the structure of the 5′-terminal palindrome and adjacent regions of H-1, we cloned the *Hin*dII 95.7/99.9 and 91/95.7 fragments with *Hin*dIII linkers into the plasmid pBR322. Subsequent subcloning of the 95.7/99.9 fragment into the single-stranded phage vector M13 mp5 resulted in a deletion of about 500 bp from the recombinants at high frequency. This instability was circumvented by subcloning 95.7/99.9 in two fragments using the *Taq*I site at 97.5 (for purposes of discussion only one orientation is assumed here), thus dividing the palindrome into parts. These fragments were sequenced by the chain-terminating method of Sanger et al. (24), and a representative gel is shown in Fig. 3. The polymerization reaction encountered a very inhibitory G-rich sequence in the palindrome when the reaction was in the direction 97.5 to 95.7 or 97.5 to 99.9, and it was not possible to read the sequence past this point on these strands. The sequencing reaction was often slightly ambiguous at this site on the opposite strands as a result of an underrepresentation of some of the G bands, as shown in Fig. 3a, but on some gels these regions were clearly determined, e.g., Fig. 3b. Another complication is that the 97.5/99.9 fragment contains a small region with a 9-bp palindrome with two *Hpa*II sites that yielded a different sequence depending on which strand of the subcloned fragment was the template (see Fig. 5). Three different clones were sequenced in the 5′ to 3′ direction 99.9 to 97.5 and two clones were sequenced from 97.5 toward 99.9 with the same result. The sequence derived in the 99.9 to 97.5 direction predicts a *Hpa*II fragment of 23 bp that is cleaved by *Hha*I. The alternative sequence, which is shown in the legend of Fig. 5, would have *Hha*I cleaving a *Hpa*II fragment of 36 bp. The results of such a digestion (Fig. 2) indicate that the former sequence is the correct one. The other sequence represents an inversion of 28 nucleotides (except the one underlined in the legend of Fig. 5). The sequence change in the fragment must have occurred during propagation of the original *Hin*dII 95.7/99.9 fragment cloned in pBR322 or during the subsequent subcloning in M13 mp5. This inversion is not to be confused with that occurring during the parvovirus RF DNA replication discussed above.

The *Hin*dII 91/95.7 fragments of H-1, the addition mutant, DI-1, and H-3 were similarly cloned in pBR322 and subcloned in M13 mp5. In addition, a *Rsa*I site revealed by sequencing was used to subdivide this region further for the sequencing of H-1 and DI-1. The sequencing strategy used is summarized in Fig. 4, and the sequence for wild-type H-1 minus strand is presented in Fig. 5.

The first seven nucleotides at the 5′ terminus were not directly determined but are inferred for reasons of symmetry since it is highly likely that they belong to the palindromic portions of the terminus which includes the *Hin*dII sites at positions 7 and 236. Because the 5′ terminus was

FIG. 3. (a) Sequencing gel of the *Hin*dII-*Taq*I fragment 95.7/98.0 (nucleotides 135 to 235) in both directions. The G-rich region at 188 to 201 produces a strong stop to the polymerization reaction, preventing confirmation of the sequence beyond that point (2). The complementary C-rich region which is adjacent to the pile-up band does not slow polymerization (3). Reading from bottom to top, panel 1 reads nucleotides 170 to 135 (plus strand), panel 2 reads 135 to 178 (minus strand), and panel 3 reads 235 to 180 (plus strand; numbers refer to those of Fig. 5). On panel 2 the H-1 sequence begins after the TTGG of the *Hin*dIII linker. In panel 3 the original *Hin*dIII linker was filled in by pol I and a second linker ligated so that the H-1 sequence begins after the TTGGAGCTTG. It was confirmed by sequencing across the *Hin*dII site at 235 and using the *Rsa*I fragment (nucleotides 157 to 313) that one G was lost from the linker during the first cloning of *Hin*dII 95.7/99.9. (b) Sequence from nucleotides 34 to 105 across the palindrome at 41 to 57. Symbols: aC, arabinofuranosyl cytosine; A, deoxyadenosine; C, deocycytidine; G, deoxyguanosine; T, thymidine.


a
1
2
3
a
C A C G T
a
C A C G T
a
C A C G T


b
TGGGAGA
GGGGG
AGGGTGG
TGTTTTT
a
C A C G T

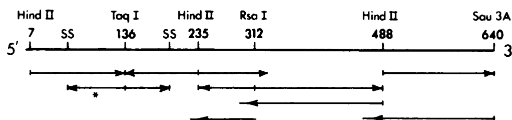


FIG. 4. Sequencing strategy used for determining the 5'-terminal sequences of parvovirus H-1. The *Hin*dII fragments: nucleotides 7 to 235 and 235 to 488 were cloned in pBR322 with *Hin*dIII linkers and 488 to ≈880 was cloned in M13 mp5 with *Eco*RI linkers. The *Hin*dII-*Taq*I fragments 7-136 and 136-235 and *Hin*dII-*Rsa*I fragments 235-312 and 312-488 were subcloned in M13 mp5. The *Sau*3A fragment 136-640 was subcloned in M13 mp7 from a *Taq*I fragment 136–≈3,270 cloned at the *Acc*I site in M13 mp7. The symbol ss indicates the G-rich palindrome that acted as a strong stop to the sequencing reactions. The clone marked with an asterisk had a region that was inverted within it, as discussed in the text.

covalently bound to protein and remained blocked after deproteinization, we have not identified the terminal nucleotide precisely. Its approximate location was derived from the observation that restriction of the terminal *Hpa*II fragment with *Hin*dII causes it to shift down in mobility by approximately 7 bp (19). Therefore, we are placing the 5' terminus tentatively at seven nucleotides from the first *Hin*dII site.

The most likely foldback structure that we predict for the 5'-terminal palindrome in virion DNA, using the methods of Tinoco et al. (32), is shown in Fig. 6. The stem of the hairpin, nucleotides 1 to 100 and 143 to 242, is a perfectly matched palindrome, and only the portion containing the asymmetries is shown in Fig. 6. A T-shaped structure is also possible if the sequences from 109 to 119 are paired to those at 89 to 94 and 149 to 154, but this is less likely since it has a lower free energy. Also, the restriction of native RF DNA with *Kpn*I, which cleaves at 76 and within the palindrome, produced only one band on agarose gels at the position expected for the 76/97.1 fragment (data not shown). If the foldback were in the T-shaped structure, the *Kpn*I site would be at the branch point of the T and not cleaved. This would result in two bands, namely the 76/97.1 and 76/100*, as found with the *Hae*III 83.5/100 bands (19). The sequence confirms the predicted asymmetries of the *Hha*I site at 100 and the *Taq*I site at 136.

The 88/95 region immediately adjacent to the terminal foldback is a noncoding region of the genome because there are chain-terminating codons in all three reading frames between nucleotides 600 and 640. A canonical polyadenylate signal of AATAAA (plus strand) followed by CA 21 nucleotides downstream is present at 277. This is compatible with the map position estimated for the 3' terminus of parvovirus mRNA species (9, 30), but two other AATAAA signals are present in this region. This region is also

```
 HindII 10         20         30         40         50
5'-TTCAGTTGAC CAACTGAACC TATAGTATCA CTATGTTTTT AGGGTGGGGG

        60         70         80   KpnI 90 HpaII  100
GGTGGGAGAT ACATACGTTC GCTATGGACC AAGTGGTACC GGTTGGTTGC

HhaI   110  HpaII 120        130   TaqI 140        150
GCTCAACCAA CCGGACCGGC TTAGCCGTCT TGTTCGAGCT TAGCAACCAA
HpaII
  KpnI 160        170        180        190        200
CCGGTACCAC TTGGTCCATA GCGAACGTAT GTATCTCCCA CCCCCCCACC

       210        220        230 HindII 240        250
CTAAAAACAT AGTGATACTA TAGGTTCAGT TGGTCAACTG AATAGACAGT

       260        270        280        290        300
AGTCCTTGTT CATTTTAATT ATTTTATTAT TTTTTTGGTC CCTAACATTC

       310   RsaI 320        330        340        350
AGTCTAAGGG AGTACCAACC AACCACCCAA CCACCCTTTT TATTTCCCAA

       360        370        380        390        400
ATACAAAATC TTTTTATATT CTATCTAACA TTAACCAACT AACTATATTT

       410        420        430        440        450
ATTTTAAAAT ACAAAATCTT TTTATATTCT ATCTAACATT AACCAACTAA

       460        470        480   HpaI 490        500
CTATATTTAT TCCAAAAGTT CTATTATTTT TTACAGTTAA CCTACCATAT

       510        520        530        540        550
TATGTTATTT TTAATAATCT ATTATTAAGT GTATATTTTT ATGTTGTAGT

       560        570        580        590        600
CTAGTTAGTT TAAGTATTAT GTGAAGCAAA CAGAGAAACA TAGTTGGTTG

       610        620        630              Sau3A
GTTAGTATGT CATGTGAGGC ACAGGTCTAC AAATCAATGG ATC-3'
```

FIG. 5. The nucleotide sequence of the 5' terminus of the parvovirus H-1. The first four nucleotides are inferred from the expected symmetry of the inverted repeat encompassing nucleotides 0 to 100 and 143 to 242, and were not directly determined. The 5'-terminal nucleotide was bound to a protein; its removal by protease digestion leaves a 5' terminus blocked to labeling by polynucleotide kinase. Its location was estimated by the mobility shift to the *Hpa*II terminal fragment after digestion with *Hin*dII. The sequences in the brackets are the 55-bp tandem repeats of wild-type H-1. The viable defective mutant of H-1, DI-1, was sequenced for the map region corresponding to *Hin*dII 236 to 488. DI-1 contained three copies of the 55-bp repeat sequence with two intervening TTA. Similarly, the parvovirus H-3 was shown to contain only one copy of this sequence. H-3 also differed from H-1 at three other positions in this region: nucleotides 291 = T, 362 = G, and 479 = G. The alternative sequence obtained from two clones marked in Fig. 4 with an asterisk was

```
100                         HpaII   HpaII
GCGCTCA-ACAAGACGGCTAGCCGGTCCGGTT-
                                 TaqI
                             GGT-TCGA
```

This sequence (between the dashes) is the inverted complement of that in Fig. 5 except for the T, which is the same. The chain terminators in three reading frames between 600 and 630 are underlined. The terminator at 602 ends the only large open reading frame, so it very likely marks the carboxy terminus of the H-1 coat proteins (Rhode, unpublished data).

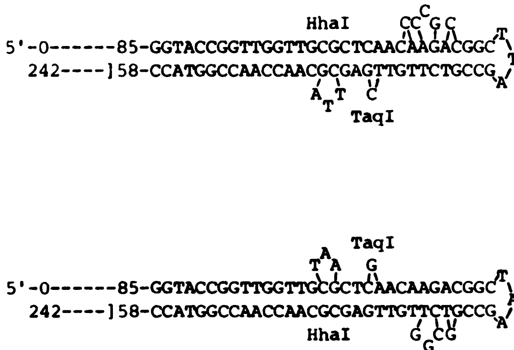


FIG. 6. Estimated structures for the two types of 5′-terminal palindromes in their foldback configurations (32). The nucleotides 0 to 85 and 158 to 242 are completely base paired. This structure was suggested by a reviewer of this manuscript and has a slightly lower free energy than the one originally proposed by us.

distinguished by its richness in AT (73%) and two regions of entirely A-T pairs at 261 to 285 and 471 to 481. H-1 contains a tandem repeat of 55 bp (348 to 402, 406 to 460) with an intervening TTA. The addition mutant DI-1 contains three tandem repeats of this sequence, including the intervening TTA, and H-3 possesses only one copy and is missing the intervening TTA. Other than the repeat sequence, H-3 is highly homologous to H-1 since it differed at only three positions between 235 and 488.

DISCUSSION

The fortuitous asymmetries of the *Hha*I and *Taq*I sites in the 5′-terminal region of H-1 RF DNA have demonstrated that native RF DNA contains approximately equal amounts of two types of terminal palindrome, one the inverted complement of the other. Since the RF DNA was derived from a freshly cloned virus, this finding indicates that the flip-flop configuration of the 5′-terminal palindrome is an active process, but it does not prove that it happens during each round of replication. This result has also been established for the defective parvovirus adeno-associated virus (12), and both results strongly support the foldback transfer process for generating 5′ termini described by Cavalier-Smith (4). In the case of the nondefective parvoviruses, this mechanism appears to apply to only one end of the DNA, the 5′ terminus of the minus strand. Because of the foldback transfer, the mere existence of covalent linkage of minus to plus strands at the 5′ terminus is not convincing evidence in favor of a DNA-primed mechanism of initiation of RF DNA replication (e.g., 31).

A hypothetical replication model that incorporates the known features of parvovirus RF replication is outlined in Fig. 7. In particular, it depends on the finding of Y-shaped replicative intermediates described for H-1 (26, 27). These data conflict with the failure of Tseng et al. to detect Okazaki fragments in nascent H-1 DNA (33). Resolution of this contradiction will require further study. The synthesis of parental RF DNA is hypothesized to initiate on the native 3′ terminus of virion DNA, step 2 (3). Direct evidence of this is lacking, but it has been shown that virion DNA chases into RF DNA with a covalently closed 3′-terminal foldback (35). The transcription of parental RF DNA has been inferred by indirect methods through its sensitivity to inhibition by substitution of bromodeoxyuridine for thymidine during parental RF synthesis (17, 37).

One principal feature of this model is that it utilizes a Cavalier-Smith type foldback transfer process to generate the 5′ termini of minus strands, which is shown at step 5 or 9. A second feature is that an asymmetrical processing of the minus strand 3′-terminal palindrome is executed on a dimer substrate for generating the 5′ termini of plus strands. In this way the unique orientation of the 3′-terminal palindrome of minus strands in virion DNA and RF DNA is conserved as previously reported (1), steps 5 and 6.

At step 3 replication is presented as an internally primed event on the plus strand template of RF DNA with the 5′ foldback region in its extended duplex configuration. There is no direct evidence to support or refute this hypothesis at this time. This step may be analogous to the D loop of mitochondria. In fact, we are proposing that the function of the G-rich sequence in the plus strand shown as f in Fig. 7 prevents minus strand synthesis on the 5′ side of this site as suggested for mitochondrial DNA replication (25). It is interesting that this site in H-1 is a G-rich palindrome, AGGGTGGGG GGGTGGGA, that is remarkably similar to the sequence AGGGGGTGGGGGGGT found adjacent to the origin of H-strand synthesis in rat mitochondrial DNA (25). By inhibiting DNA synthesis, this signal sequence would ensure that the plus strand palindrome remains free to renature as a foldback structure when initiation and rightward chain extension on the minus strand template drives the separation of the parental foldback region, as shown in steps 3 and 4. Progression of the replication fork toward the covalently closed 3′ foldback converts the monomer RF to a dimer, step 5. Synthesis of the plus strand would be by a discontinuous method that presumably uses Okazaki fragments as intermediates.

As an alternative, a DNA-primed initiation using the plus strand 3′ terminus as primer

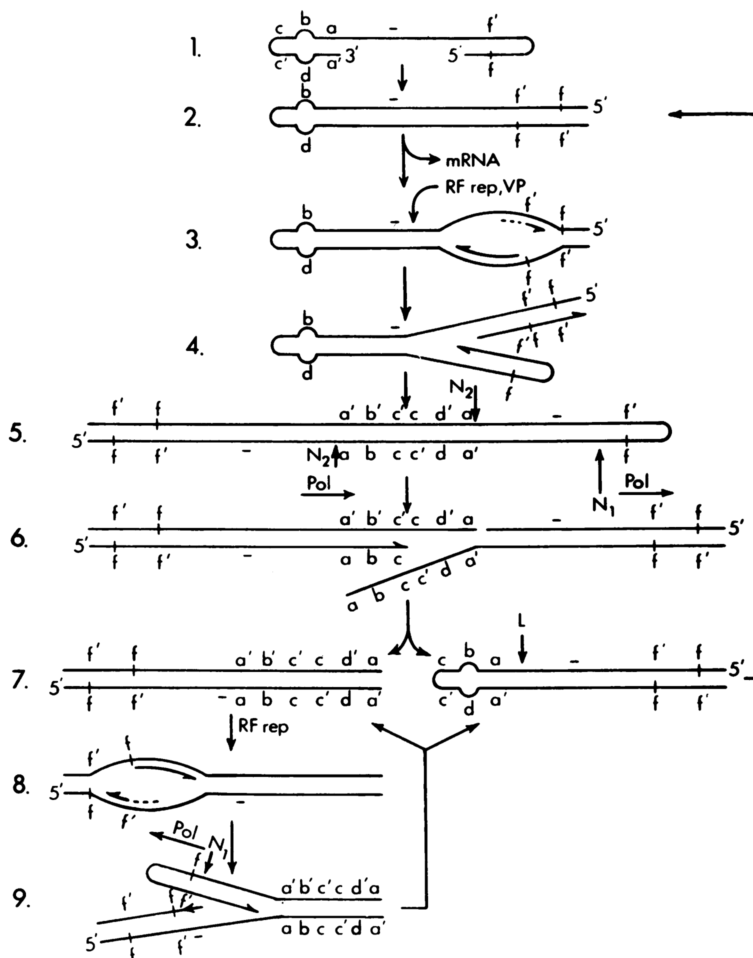


FIG. 7. Hypothetical model for the replication of parvovirus DNA. The viral DNA is not drawn to scale. Complementary sequences are represented by small letters and their primes. Abbreviations: RF rep, transacting viral product required for RF replication; VP, viral proteins; N_1 or N_2 site-specific single-strand endonuclease or nickase; Pol, DNA polymerization; L, ligation; f, the palindrome sequence GGGTGGGGGGGTGGG. All or nearly all of the 5′ termini are bound to a protein (15). Step 1 depicts the structure of virion DNA. At step 2 the synthesis of parental RF DNA has occurred, and the parental RF is template for viral mRNA synthesis. When RF rep or virion proteins, or both, are synthesized, RF replication proceeds at step 3. Initiation is proposed to begin with minus strand DNA just inboard of the terminal inverted repeat sequences. An early replicative intermediate is shown at step 3. Progression of the replication fork to the 3′ hairpin converts the parental RF to a dimer RF. A foldback transfer process proceeds at one of the daughter molecule's 5′ termini to produce a 5′ terminus for the nascent minus strand in steps 5 and 6. This is represented by nickase N_1 cleaving the plus strand to produce a 3′-OH primer site for strand displacement synthesis which extends the foldback. In a similar manner nickase N_2 cleaves at the boundary of the 3′ hairpin sequences (step 5), but strand displacement synthesis occurs at only one of these sites (step 6). In step 7 the dimer has converted to two monomer RF DNA molecules, and the orientation of the 3′ hairpin remains the same as for the parental virion DNA. RF DNA with the extended 3′ end replicates as in steps 3 and 4 except that two monomer RF molecules are generated (steps 7–9). Monomer RF DNA with the 3′ end in the hairpin configuration must recycle through the dimer intermediate when it replicates. This model was formulated to fit the following experimental data: (i) the replicative intermediates of RF DNA are double-stranded branched molecules (26); (ii) the orientation of the 3′ hairpin is unique (1); (iii) there are two orientations of the 5′ hairpin (this work; 1); (iv) dimer molecules are prominent members (25%) of the pool of replicating RF DNA (18, 26, 27); (v) a significant fraction of the 3′ foldback structures are covalently closed (18); (vi) only the extended and foldback configurations of the 5′ terminus have been identified (19). A rabbit-eared 5′ terminus has not been detected by gel electrophoresis of 5′-terminal restriction fragments.

would require a mechanism for converting the extended structure of the palindrome to a "rabbit-eared" structure. We have produced a terminal fragment tentatively identified as a "rabbit-eared" structure after denaturation and renaturation of the native terminal fragment. There is no evidence for such a terminal fragment in native RF DNA (S. L. Rhode, unpublished data).

The foldback transfer processes which are symmetrical at the 5' terminus, that is, independent of the flip or flop orientation, and asymmetrical for the 3' terminus are shown as a site-specific nicking reaction by nickase N_1 or N_2 followed by strand displacement by the DNA replication apparatus, $\xrightarrow{pol}$. Therefore, DNA replication itself would drive the strand separations necessary for these reactions. Examination of the reported sequence for the various 3'-terminal structures of H-1 and comparison with the appropriate region of the 5' terminus suggests how the template could direct these reactions (Fig. 8). Sequence 1 is the plus strand sequence at the border of the 5' foldback with the foldback turnaround to the right. Sequences 2 and 3 are the two portions of minus strand sequences from the dimer junction, that is, the abc and ad'c of Fig. 7, and sequence 4 is the 3' terminus in foldback configuration. The minus strand 3'-terminal sequences at the monomer junctions in the dimer are shown one above the other for ease of comparison with the axis of symmetry at the right indicated by an asterisk. The tentative nicking and primer sites for strand displacement are again indicated by $N_{\downarrow}\xrightarrow{pol}$. A most striking feature of all four sequences is the homologous positions of the possible recognition signals TGAAC, TGACCAAC, and TGAA($^{C}_{A}$)C with respect to the postulated nicking sites. Equally important, the 3' foldback region of the dimer, which we have suggested is nicked but is not a substrate for strand displacement synthesis, is missing the TGAAAC sequence present in the other two templates. Similarly, the same signal site in the foldback configuration (4) is located at an interior loop rather than a duplex site, as found in the dimer. This model proposes that this configuration is processed by ligation rather than nicking. Therefore, the model predicts that a duplex TGAA($^{C}_{A}$)C signal is required for the strand displacement DNA synthesis. Thus, these regions contain the required symmetries and asymmetries of sequence necessary to regulate the foldback transfer mechanism as described in the model. It is possible that the nicking enzymes N_1 and N_2 are the same even though the nick sites do not seem to have identical relationships to the postulated recognition signals. It is also possible that the terminal bound proteins are the nickases (15).

This model predicts a recycling of some of the monomer RF DNAs through a dimer molecule and allows some of the monomer RF DNAs to replicate without a dimer intermediate, steps 7–9. Dimers are postulated to be formed by replication and not recombination. This is supported by our finding that the dimers produced in cells doubly infected with standard virus and a defec-

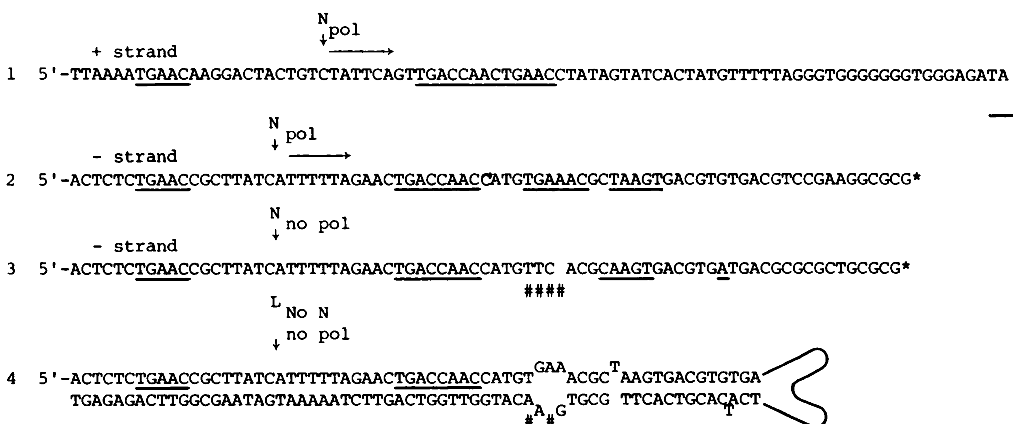


FIG. 8. Comparison of the 5'-terminal (1) and 3'-terminal (2–4) sequences in the region of the hypothetical sites for the nickases N_1 and N_2 or the ligation (L) of the replication model of Fig. 7. Sequence 1 is the plus strand sequence at the boundary of the 5'-terminal palindrome. Sequences 2 and 3 are the minus strand sequences at the dimer RF junction brought into vertical alignment with the axis of symmetry indicated by the asterisk. Sequence 4 is the 3' terminus in its foldback configuration (1). The # signs below the line mark points of asymmetry where the one side of the dimer sequence (3) or the foldback sequence (4) does not have a duplex TGAAAC which we postulate to be required for the strand displacement, $\xrightarrow{pol}$, to occur. The exact site of nicking for sequence 1 awaits the precise placement of the 5'-terminal nucleotide.

tive deletion mutant are of two types: standard length ×2 or deleted length ×2 (Rhode, unpublished data). This conclusion should be tempered by the fact that we cannot prove that the replicating pools of standard and deleted molecules are not compartmentalized within the nucleus, thus precluding recombination between them (28).

Because of the symmetry within 100 nucleotides of the nick site at the 5′ terminus, one might expect that the nicking-strand displacement mechanism might not distinguish between 5′ termini in foldback and extended configurations. If this is the case, this process acting on extended termini would strand displace free copies of the 5′ hairpin as a wasteful side reaction of the replication process, and we intend to search for these. Other predictions of this model which can be tested are that the 5′ termini of dimer molecules will be in the same orientation (step 6, Fig. 7) and that inhibition of DNA synthesis will block the interconversion of the various terminal structures.

Finally, the model does not show progeny minus strand synthesis. Virion DNA synthesis could occur by an interaction of an appropriately mature viral capsid with an early replication intermediate (just before step 3 or 8) in such a way as to block the antipolar plus strand synthesis. Evidence has been presented which suggests that encapsidation proceeds in close conjunction with the displacement of minus strand from the RF template (7).

The sequence derived for the 5′ terminus of H-1 includes the replication origin which maps to this site by electron microscope analysis of replicative intermediates (26, 27). The sequence data have confirmed our earlier hypothesis that the addition in DI-1 is a tandem repeat of a sequence that is repeated already in wild-type H-1 and is present in only one copy in H-3 (20). A similar relationship has been suggested by physical mapping data for minute virus of mice (MVM) and MVM (i) (14). It is thus very likely that the further additions reported in H-1 DI genomes between map positions 91 and 95.7 are tandem repeats of this same sequence (21). What role this sequence has in DNA replication is unknown, but it is clear from the example of H-3 that only one copy is required for efficient RF DNA replication. The tentative origin region has many features reported for other origins of DNA replication. The noncoding region inboard of the terminal palindrome 244 to 640 is A-T rich (73%) and contains a 25-nucleotide AT block at 262 to 286 and an 11-nucleotide AT block at 472 to 482. It also contains a 23-nucleotide AC stretch at 314 to 336. There are true palindromes of eight nucleotides or larger at 3 to 10, 25 to 35, 41 to 57, 111 to 119, 186 to 202, 208 to 218, 233 to 240, 267 to 278, and 313 to 324. We are attempting to develop a functional assay for origin activity so that structure-function studies can be carried out to determine which features of this sequence are important in the initiation of RF DNA replication.

ACKNOWLEDGMENTS

We gratefully acknowledge Jeanne Helft for expert technical assistance, Virginia Haas for secretarial duties, and Eleanor Spicer for helpful discussions.

This work was supported by Public Health Service grants CA-25866 and CA26801 from the National Cancer Institute.

LITERATURE CITED

1. **Astell, C., M. Smith, M. Chow, and D. Ward.** 1979. Structure of the 3′ hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at origins of DNA replication. Cell 17:691–703.
2. **Astell, C., M. Smith, M. Chow, and D. Ward.** 1979. Sequence of the 3′-terminus of the genome from Kilham rat virus, a non-defective parvovirus. Virology 94:669–674.
3. **Bourguignon, G., P. Tattersall, and D. 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.
4. **Cavalier-Smith, T.** 1974. Palindromic base sequences and replication of eukaryote chromosome ends. Nature (London) 250:467–470.
5. **Chow, M. B., and D. C. Ward.** 1978. Comparison of the terminal nucleotide structures in the DNA of nondefective parvoviruses, p. 205–217. *In* D. C. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6:23–28.
7. **Gautschi, M., and P. Reinhard.** 1978. Synthesis of a mammalian parvovirus on Brij 58-lysed cells. J. Virol. 27:453–456.
8. **Goodman, H. M., and R. J. MacDonald.** 1979. Cloning of hormone genes from a mixture of cDNA molecules. Methods Enzymol. 68:75–89.
9. **Green, M., R. Lebovitz, and R. Roeder.** 1979. Expression of the autonomous parvovirus H-1 genome: evidence for a single transcriptional unit and multiple spliced polyadenylated transcripts. Cell 17:967–977.
10. **Guerry, P., D. J. LeBlanc, and S. Falkow.** 1973. General method for the isolation of plasmid deoxyribonucleic acid. J. Bacteriol. 116:1064–1066.
11. **Heidecker, G., J. Messing, and B. Gronenborn.** 1980. A versatile primer for DNA sequencing in the M13 mp2 cloning system. Gene 10:69–73.
12. **Lusby, E., K. H. Fife, and K. I. Berns.** 1980. Nucleotide sequence of the inverted terminal repetition in adenoassociated virus DNA. J. Virol. 34:402–409.
13. **McGeoch, D. J., L. V. Crawford, and E. A. C. Follett.** 1970. The DNAs of three parvoviruses. J. Gen. Virol. 6:33–40.
14. **McMaster, G. K., P. Beard, H. D. Engers, and B. Hirt.** 1981. Characterization of an immunosuppressive parvovirus related to minute virus of mice. J. Virol. 38:317–326.
15. **Revie, D., B. Tseng, R. Grafstrom, and M. Goulian.** 1979. Covalent association of protein with the replicative form DNA of parvovirus H-1. Proc. Natl. Acad. Sci. U.S.A. 11:5539–5543.
16. **Rhode, S. L.** 1974. Replication process of the parvovirus H-1. II. Isolation and characterization of H-1 replicative form DNA. J. Virol. 13:400–410.
17. **Rhode, S. L.** 1974. Replication process of the parvovirus H-1. III. Factors affecting H-1 RF DNA synthesis. J. Virol. 14:791–801.

18. **Rhode, S. L.** 1977. Replication process of the parvovirus H-1. VI. Characterization of a replication terminus of H-1 replicative-form DNA synthesis. J. Virol. **21**:694–712.
19. **Rhode, S. L.** 1977. Replication process of the parvovirus H-1. IX. Physical mapping studies of the H-1 genome. J. Virol. **22**:446–458.
20. **Rhode, S. L.** 1978. Replication process of the parvovirus H-1. X. Isolation of a mutant defective in replicative-form DNA replication. J. Virol. **25**:215–223.
21. **Rhode, S. L.** 1978. Defective interfering particles of parvovirus H-1. J. Virol. **27**:347–356.
22. **Rhode, S. L.** 1977. H-1 DNA synthesis, p. 279–296. *In* D. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
25. **Sekiya, T., M. Kobayashi, T. Seki, and K. Koike.** 1980. Nucleotide sequence of a cloned fragment of rat mitochondrial DNA containing the replication origin. Gene **11**:53–62.
26. **Singer, I. I., and S. L. Rhode.** 1977. Replication process of the parvovirus H-1. VII. Electron microscopy of replicative form DNA synthesis. J. Virol. **21**:713–723.
27. **Singer, I. I., and S. L. Rhode.** 1977. Replication process of the parvovirus H-1. VIII. Partial denaturation mapping and localization of the replication origin of H-1 replicative form DNA with electron microscopy. J. Virol. **21**:724–731.
28. **Singer, I. I., and S. L. Rhode.** 1978. Ultrastructural studies of H-1 parvovirus replication. VI. Simultaneous autoradiographic and immunochemical intranuclear localization of viral DNA synthesis and protein accumulation. J. Virol. **25**:349–360.
29. **Smith, A. J. H.** 1980. DNA sequence analysis by primed synthesis. Methods Enzymol. **65**:560–579.
30. **Tal, J., P. Ron, P. Tattersall, S. Bratosin, and Y. Aloni.** 1979. About 30% of minute virus of mice RNA is spliced out following polyadenylation. Nature (London) **279**:649–651.
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. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, P. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) **246**:40–41.
33. **Tseng, R. Y., R. H. Grafstrom, D. Revie, W. Oertel, and M. Goulian.** 1979. Studies on early intermediates in the synthesis of DNA in animal cells. Cold Spring Harbor Symp. Quant. Biol. **43**:263–270.
34. **Usategui-Gomez, M., F. Al-Lami, S. M. Hopkins, N. Ledinko, and H. Toolan.** 1969. Single-stranded DNA from the parvovirus H-1. Virology **39**:617–621.
35. **Ward, D. C., K. Dinshaw, and D. K. Dadachanji.** 1978. Replication of minute virus of mice DNA, p. 297–313. *In* D. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Ward, D. C., and P. Tattersall (ed.).** 1978. Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **Wolter, S., R. Richards, and R. W. Armentrout.** 1980. Cell cycle-dependent replication of the DNA of minute virus of mice, a parvovirus. Biochim. Biophys. Acta **607**:420–431.