JOURNAL OF VIROLOGY, Apr. 1982, p. 337-341
0022-538X/82/040337-05$02.00/0

Vol. 42, No. 1

Structure of a Replication Intermediate in the Synthesis of Rous Sarcoma Virus DNA In Vivo

RONALD SWANSTROM,* J. MICHAEL BISHOP, AND HAROLD E. VARMUS

Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Received 5 October 1981/Accepted 16 November 1981

Intermediates in the synthesis of Rous sarcoma virus DNA in vivo contain a short second strand of DNA (plus strong-stop DNA) synthesized by using the region near the 5' end of the first (minus) strand of DNA as the template. In this report, we show that the 3' end of plus strong-stop DNA is extended about 15 to 20 nucleotides beyond the 5' end of the minus-strand DNA template, probably copying a portion of the $tRNA^{Trp}$ molecule that serves as primer for synthesis of the minus strand of DNA. The extra sequences present in plus strong-stop DNA may play a central role in the generation of the long terminal repeat present in mature forms of viral DNA.

The replication of retroviruses proceeds through a DNA intermediate (22, 26) synthesized by the virus-encoded, RNA-dependent DNA polymerase (24). The synthesis of viral DNA, using single-stranded genome RNA as the template, a process referred to as reverse transcription, has been studied extensively both in vitro and in vivo. As a result of these studies, a detailed model has been developed that can account for many of the features of this process (3; H. E. Varmus and R. Swanstrom, *in* R. Weiss, N. Teich, H. E. Varmus, and J. Coffin, ed., *Molecular Biology of Tumor Viruses Part 3, RNA Tumor Viruses*, in press). One of the unusual features of the model is that synthesis of the first strand of viral DNA (the minus strand) proceeds in a continuous manner on a discontinuous template. A consideration of the sequence organization of the RNA template and the DNA product suggests why this is a necessary feature of the model.

The structures of genome RNA and of the major unintegrated DNA product found in vivo, duplex linear DNA, are shown in Fig. 1A. The sequences within the RNA are rearranged in viral DNA in two significant ways. First, sequences that are at the ends of the RNA are joined in the DNA (U3-R-U5). Second, the block of sequences, U3-R-U5, is duplicated at the ends of the DNA, forming a terminal repeat (termed long terminal repeat [LTR]) that is not present in the RNA template (6, 14). It is clear that a continuous RNA template cannot give rise to the DNA product shown in Fig. 1A. A central objective of the current model of reverse transcription is to explain how multiple template strands are used.

The first discontinuity in the template appears when the minus strand reaches the 5' end of viral RNA, about 100 bases from the genome-bound tRNA that is used as a primer for minus-strand synthesis (5, 15, 20). Nucleotide sequence analysis of DNA extended beyond the 5' end of the template (10, 18) has shown that one copy of the short direct repeat present at the ends of viral RNA (R in Fig. 1A; 5, 12, 15) is sacrificed as transcription moves from the 5' end to the 3' end of the template. With the placement of the minus strand at the 3' end of the template, the fusion of the sequences U3-R-U5 is accomplished, and the bulk of the viral RNA template can be transcribed in an uninterrupted way. However, another template strand must be employed to generate a second copy of the LTR unit.

Before the minus strand is completed, the second, or plus, strand of viral DNA is initiated (9, 23). In particular, one prominent species of plus-strand DNA, termed plus strong-stop DNA [(+)ssDNA], is synthesized early after infection, using the region near the 5' end of the minus strand as template. Two observations implicated (+)ssDNA as the template for completion of the minus strand: (i) restriction endonuclease mapping of the sequences present in (+)ssDNA showed that they are the same as sequences present in the LTR (3, 9, 23), and (ii) completion of the minus strand is sensitive to actinomycin D, suggesting the necessity of a DNA template (11). These observations led to the suggestion that the growing end of the minus strand "jumps" to (+)ssDNA and, by displacement synthesis, copies the short plus strand to generate the LTR (3, 14).

Figure 1B shows a diagram of the replication intermediate that is believed to mediate the jump of the growing minus strand to (+)ssDNA (3, 21;

Varmus and Swanstrom, in press). The important feature of this intermediate is the presence of complementary single-stranded regions at the 3′ ends of the minus and plus strands. Annealing of the complementary ends allows formation of a short duplex containing these ends. Subsequent extension of the 3′ end of the minus strand can occur by strand displacement synthesis. Using (+)ssDNA as template to complete the minus strand produces the LTR. The tRNA primer for the minus strand and the primer binding site in genome RNA [(−)PB] may play a crucial role in generating the complementary ends in the intermediate shown in Fig. 1B. Based on the model of retrovirus DNA synthesis (3; Varmus and Swanstrom, in press), we can predict the following series of events.

(i) When (+)ssDNA is synthesized by using the 5′ end of the minus strand as template, a portion of the minus-strand primer, covalently linked to the minus strand, is also copied into the plus strand. In the case of Rous sarcoma virus (RSV), 18 nucleotides can be copied from the $tRNA^{Trp}$ primer before transcription is blocked by a 1-methyl-adenine at position 19 from the 3′ end of the tRNA (2, 4). Removal of the RNA from the (+)ssDNA-tRNA hybrid by the polymerase-associated RNase H activity leaves a single-stranded tail at the 3′ end of (+)ssDNA composed of sequences representing the (−)PB site.

(ii) Sequences complementary to the (−)PB site appear in the minus strand when this site is copied from the RNA genome. RNase H removes the R-U5 sequences from the 5′ end of the RNA template (these sequences are in a hybrid with the 5′ end of the minus strand), causing synthesis of the minus strand to terminate after copying the (−)PB site. Again, RNase H removes the template RNA from the minus-strand DNA, leaving a single-stranded region in the DNA composed of sequences complementary to the (−)PB site. The single-stranded tail at the 3′ end of the minus strand can anneal to the single-stranded tail at the 3′ end of (+)ssDNA. After annealing, the minus strand is extended by using (+)ssDNA as template, resulting in the duplication of the sequences in (+)ssDNA at each end of the minus strand.

In this paper, we describe the structure of (+)ssDNA in a replication intermediate of RSV DNA found in vivo. On the basis of S1 nuclease sensitivity, the 3′ end of (+)ssDNA is shown to contain sequences not present in the minus-strand DNA template. The position of these extra sequences is consistent with the proposal that a portion of the minus-strand tRNA primer is copied during synthesis of (+)ssDNA.

Cultured quail tumor cells (QT-6) were infected with the B77 strain of RSV, and cytoplasmic DNA was isolated 3 h after infection (23). The sample was treated with RNase A in low salt and then fractionated by velocity sedimentation in a neutral sucrose gradient (23). DNA sedimenting between 8S and 20S was pooled and ethanol precipitated. Viral DNA prepared in this fashion after a brief infection is free of mature forms of viral DNA (linear or circular DNA), and any DNA sequences present in an RNA hybrid are rendered single stranded by the RNase treat-

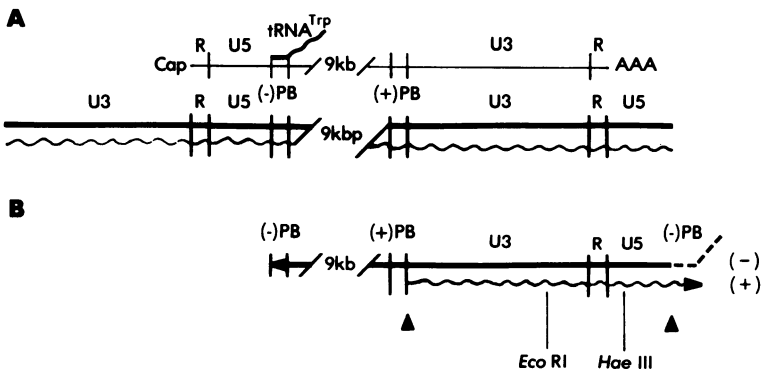

FIG. 1. (A) Genome RNA and unintegrated linear viral DNA. The symbols used in this diagram have been described previously (1; Varmus and Swanstrom, in press). Briefly, R refers to the short direct repeat present at the ends of viral RNA. U5 and U3 are unique sequences near the 5′ and 3′ ends, respectively, of viral RNA that are duplicated during the genesis of the long terminal repeat in viral DNA. (−)PB and (+)PB are the positions of the primers for minus-strand DNA and (+)ssDNA synthesis. (B) A proposed intermediate in the synthesis of viral DNA. (−) First (minus) strand of viral DNA, represented by the heavy line. (+) Second (plus) strand of DNA, represented by the light line. The arrows at the ends of the lines indicate the direction of polymerization (i.e., the 3′ ends). The dashed line shows the position of the tRNA primer for minus-strand synthesis, and the arrowheads show the possible positions of S1 nuclease-sensitive sites. The positions of the *Eco*RI and the *Hae*III restriction endonuclease sites in this region are also shown.

ment. Viral DNA was analyzed by digestion with the single-strand-specific nuclease S1 for 30 min at 37°C (25). The conditions used for S1 nuclease digestion allowed complete hydrolysis of single-stranded DNA, as measured by acid solubility of radiolabeled DNA, whereas 80% of the radioactivity in end-labeled duplex DNA was resistant to the nuclease (data not shown).

The strategy used to examine the structure of the replication intermediate was to determine the size of (+)ssDNA before and after S1 nuclease digestion. The region of (+)ssDNA in duplex with minus-strand DNA should be resistant to digestion. However, if (+)ssDNA also contains a copy of part of the $tRNA^{Trp}$ primer, then there should be a single-stranded tail at the 3′ end of (+)ssDNA susceptible to digestion by S1 nuclease. The size of (+)ssDNA was determined by electrophoresis in an agarose gel, followed by transfer of the DNA out of the gel onto nitrocellulose filter paper and detection of viral DNA by hybridization with a radioactive DNA probe (14, 16). The radioactive probe ($cDNA_{3'}$ [19]) was of minus-strand polarity so that it would anneal only to plus-strand sequences. The analysis of (+)ssDNA is shown in Fig. 2A.

Viral DNA was analyzed in both native and denatured forms. (+)ssDNA in native DNA, not treated with S1 nuclease, appeared in large heterogeneous DNA species migrating slowly in an agarose gel (Fig. 2A, lane 2), as expected of (+)ssDNA base-paired to minus-strand DNA with heterogeneous lengths (23). When this DNA was treated with S1 nuclease (Fig. 2A, lane 3), a duplex region of about 340 base pairs was observed, representing the duplex region of minus-strand DNA and (+)ssDNA (Fig. 1B). To determine the structure of the right end of this population of molecules, the duplex region containing (+)ssDNA was cleaved with the restriction endonuclease *Eco*RI before S1 nuclease treatment. This step eliminates the minus-strand size heterogeneity by giving all of the molecules of interest the same left end (see Fig. 1B). After cleavage with *Eco*RI, but before S1 nuclease treatment, fragments detected by $cDNA_{3'}$ were largely homogeneous in size, with an apparent molecular weight of 170 base pairs (Fig. 2A, lane 4). This size is larger than that expected for the *Eco*RI restriction fragment, since sequence analysis of two closely related viruses places the *Eco*RI site 150 nucleotides from the U5-(−)PB boundary (17; D. Schwartz, personal communication), the site believed to represent the right end of unintegrated viral DNA (see Fig. 1 and references 13, 17, and Varmus and Swanstrom, in press). When this DNA was treated with S1 nuclease (Fig. 2A, lane 5), the apparent size of the *Eco*RI fragment was reduced to about 150 base pairs. Thus, the duplex *Eco*RI fragment containing (+)ssDNA sequences must contain a region of single-stranded DNA to account for the shift in mobility after S1 nuclease treatment.

To determine which strand in the duplex DNA contained the single-stranded tail, the same samples (Fig. 2A, lanes 2 through 5) were also analyzed under denaturing conditions (Fig. 2A, lanes 7 through 10). Denaturation of untreated viral DNA resulted in the clear identification of (+)ssDNA (Fig. 2A, lane 7), as has been previously described (7, 9, 23). (+)ssDNA had an apparent length of around 370 bases in this analysis. Treatment of the DNA with S1 nuclease before denaturation resulted in a slight shift downward in the size of (+)ssDNA to about 340 bases (Fig. 2A, lane 8). This experiment shows that the S1 nuclease-induced reduction in the size of the duplex molecules was due to digestion of a single-stranded region in (+)ssDNA. The fact that the shift occurred after digestion with *Eco*RI (Fig. 2A, lanes 4 and 5) places the single-stranded region at the 3′ end of (+)ssDNA. To determine the size of the single-stranded region more accurately, we examined the molecules denatured after *Eco*RI cleavage. S1 nuclease treatment reduced the length of the plus strand from 180 to 160 bases, or around 20 bases (Fig. 2A, lanes 9 and 10). Part of the shift may be attributed to the 4-base overhang at the site of cleavage by *Eco*RI in (+)ssDNA. Thus, the single-stranded tail at the 3′ end of (+)ssDNA appears to be between 15 and 20 bases long.

As an additional control to exclude digestion of flush-ended duplex DNA by S1 nuclease, we created a blunt end in the duplex region of (+)ssDNA and determined the S1 nuclease sensitivity of this site. The blunt end was obtained by cleavage with the restriction endonuclease *Hae*III, which cleaves within the U5 region of the B77 strain of RSV at position 57 from (−)PB (Fig. 1B; R. Swanstrom, unpublished observations). An appropriately shortened fragment containing part of (+)ssDNA was obtained after *Hae*III cleavage of the same sample of viral DNA used in the previous experiments (Fig. 2B, lanes 2 and 3). When untreated DNA was mixed with the *Hae*III-cleaved DNA and the mixture was subjected to S1 nuclease digestion, only the untreated DNA showed a size shift with nuclease treatment, whereas the size of the DNA containing the blunt end was unaffected by S1 nuclease digestion (Fig. 2B, lanes 4 and 5). The results from this experiment confirm the previous results and show that blunt ends are not significantly affected under the conditions of S1 nuclease digestion that were employed.

Other workers have identified unpaired sequences at the right end of duplex viral DNA synthesized in vitro by RSV and murine leuke-

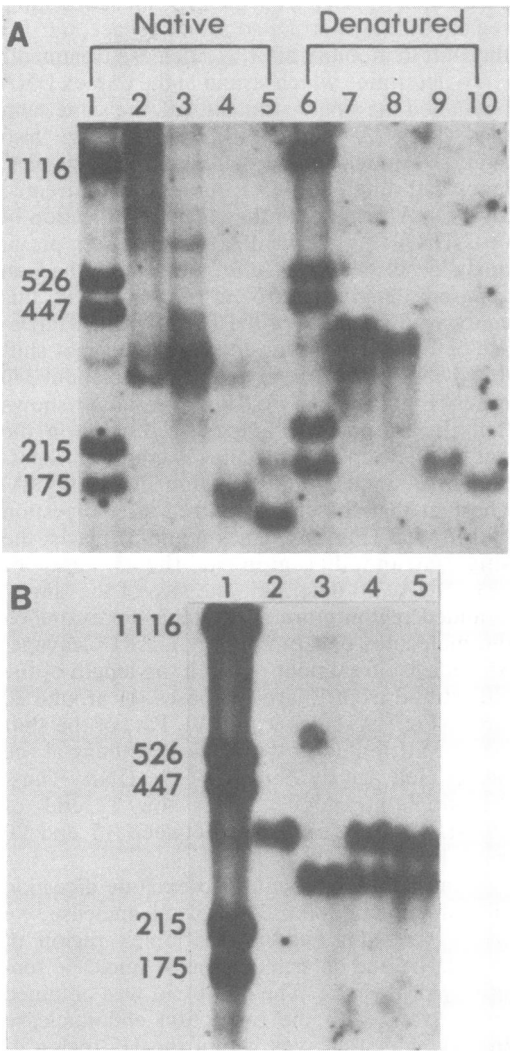

FIG. 2. Digestion of a replication intermediate of RSV DNA with S1 nuclease. Cytoplasmic DNA was isolated 3 h after infection of QT-6 cells with the B77 strain of RSV (23). Some of the DNA samples were treated with *Eco*RI or S1 nuclease or both, as indicated. The DNA was analyzed by electrophoresis in a 3% agarose gel (10 mM sodium phosphate, pH 6.8), using simian virus 40 DNA cleaved with *Hin*dIII and end labeled as size markers. Some samples (lanes 6 through 10) were denatured with 50% formamide, 50°C, for 30 min in the presence of glyoxal before electrophoresis (8). (Under these gel conditions, denatured DNA migrates slightly slower than does duplex DNA.) After electrophoresis, the gel was treated with alkali and neutralized, and the DNA samples were transferred to nitrocellulose filter paper as described (14, 16). Virus-specific sequences were detected by hybridization with radioactive RSV $cDNA_3$ (19), a probe complementary to the region of the RSV genome adjacent to the 3′ polyadenylate tract. (A) Lane 1, native size marker. Numbers on the side refer to the lengths of the fragments of the marker DNA in either base pairs (lanes 1 through 5) or bases (lanes 6 through 10). Lane 2, DNA from infected cells (untreated); lane 3, DNA from infected cells treated with S1 nuclease before electrophoresis; lane 4, DNA from infected cells treated with *Eco*RI before electrophoresis; lane 5, DNA from infected cells treated with *Eco*RI and S1 nuclease before electrophoresis; lanes 6 through 10 are identical to lanes 1 through 5, except the samples were denatured in the presence of glyoxal before electrophoresis. (B) DNA samples were treated with *Hae*III or S1 nuclease or both, as indicated. The samples were denatured and analyzed by electrophoresis in a 3% agarose gel as described above. The numbers on the side refer to the lengths of the marker DNA fragments measured in bases. Lane 1, Fragments of simian virus 40 DNA generated by cleavage with *Hin*dIII; lane 2, viral DNA sample (untreated); lane 3, viral DNA sample cleaved with *Hae*III; lane 4, a mixture of the samples treated as described for lanes 2 and 3; lane 5, same as the sample in lane 4, but treated with S1 nuclease before denaturation and electrophoresis.

mia virus (3, 21). In the studies of murine leukemia virus DNA, these sequences were shown to be in (+)ssDNA and complementary to part of the minus-strand primer, $tRNA^{Pro}$ (3). We have made a similar observation for RSV DNA synthesized in vivo, showing that (+)ssDNA is extended at its 3′ end 15 to 20 nucleotides beyond the 5′ end of its minus-strand DNA template; however, we have not attempted to show that the extra sequences are complementary to $tRNA^{Trp}$.

The structure of (+)ssDNA we observed for DNA synthesized in vivo corresponded to the structure of a predicted intermediate in the currently accepted model of reverse transcription (3; Varmus and Swanstrom, in press). This is the first demonstration of such a structure in viral DNA synthesized in vivo. The fact that virtually all viral DNA molecules had an extended (+)ssDNA (Fig. 2A, lanes 4 and 5) further supports the idea that this is an important structure in the synthesis of viral DNA.

We thank S. Ortiz for preparing the DNA used in these experiments and B. Cook for typing the manuscript.

This work was supported by Public Health Service grants CA 12705, CA 19287, and training grant 1T32 CA 09043 from the National Institutes of Health and by American Cancer Society grant MV48G.

LITERATURE CITED

1. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. 42:1–26.
2. **Cordell, B., F. M. DeNoto, J. F. Atkins, R. F. Gestland, J. M. Bishop, and H. M. Goodman.** 1980. The forms of $tRNA^{Trp}$ found in avian sarcoma virus and uninfected chicken cells have structural identity but functional distinctions. J. Biol. Chem. 255:9358–9368.

3. **Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell **18**:93–100.
4. **Harada, F., R. C. Sawyer, and J. E. Dahlberg.** 1975. Primer ribonucleic acid for initiation of *in vitro* Rous sarcoma virus deoxyribonucleic acid synthesis-nucleotide sequence and amino acid acceptor activity. J. Biol. Chem. **250**:3487–3497.
5. **Haseltine, W. A., A. M. Maxam, and W. Gilbert.** 1977. Rous sarcoma virus genome is terminally redundant: the 5′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:989–993.
6. **Hsu, T. W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and J. M. Taylor.** 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. J. Virol. **28**:810–818.
7. **Kung, H.-J., Y. K. Fung, J. E. Majors, J. M. Bishop, and H. E. Varmus.** 1981. Synthesis of plus strands of retroviral DNA in cells infected with avian sarcoma virus and mouse mammary tumor virus. J. Virol. **37**:127–138.
8. **McMaster, G. K., and G. G. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A. **74**:4835–4838.
9. **Mitra, S. W., S. Goff, E. Gilboa, and D. Baltimore.** 1979. Synthesis of a 600-nucleotide-long plus-strand DNA by virions of Moloney murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **76**:4355–4359.
10. **Omer, C., J. T. Parsons, and A. J. Faras.** 1981. Direct proof of the 5′ to 3′ transcriptional jump during reverse transcription of the avian retrovirus genome by DNA sequencing. J. Virol. **38**:398–402.
11. **Rothenberg, E., and D. Baltimore.** 1977. Increased length of DNA made by virions of murine leukemia virus at limiting magnesium ion concentrations. J. Virol. **21**:168–178.
12. **Schwartz, D. E., P. C. Zamecnik, and H. L. Weith.** 1977. Rous sarcoma virus genome is terminally redundant: the 3′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:994–998.
13. **Scott, M. L., K. McKereghan, H. S. Kaplan, and K. E. Fry.** 1981. Molecular cloning and partial characterization of unintegrated linear DNA from gibbon ape leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **78**:4213–4217.
14. **Shank, P. R., S. H. Hughes, H.-J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell **15**:1383–1395.
15. **Shine, J., A. P. Czernilofsky, R. Friedrich, H. M. Goodman, and J. M. Bishop.** 1977. Nucleotide sequence at the 5′ terminus of the avian sarcoma virus genome. Proc. Natl. Acad. Sci. U.S.A. **74**:1473–1477.
16. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
17. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Sequencing of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
18. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1981. The terminal redundancy of the retrovirus genome facilitates chain elongation by reverse transcriptase. J. Biol. Chem. **256**:1115–1121.
19. **Tal, J., H.-J. Kung, H. E. Varmus, and J. M. Bishop.** 1977. Characterization of DNA complementary to nucleotide sequences adjacent to poly(A) at the 3′-terminus of the avian sarcoma virus genome. Virology **79**:183–197.
20. **Taylor, J. M.** 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. Biochim. Biophys. Acta **473**:57–71.
21. **Taylor, J. M., and T. W. Hsu.** 1980. Reverse transcription of avian sarcoma virus RNA into DNA might involve copying the tRNA primer. J. Virol. **33**:531–534.
22. **Temin, H. M., and D. Baltimore.** 1972. RNA-directed DNA synthesis and RNA tumor viruses. Adv. Virus Res. **17**:129–186.
23. **Varmus, H. E., S. Heasley, H.-J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop, and P. R. Shank.** 1978. Kinetics of synthesis, structure, and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. J. Mol. Biol. **120**:55–82.
24. **Verma, I. M.** 1977. The reverse transcriptase. Biochim. Biophys. Acta **473**:1–38.
25. **Vogt, V. M.** 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. Eur. J. Biochem. **33**:192–200.
26. **Weinberg, R.** 1977. Structure of the intermediates leading to the integrated provirus. Biochim. Biophys. Acta **473**:39–55.