JOURNAL OF VIROLOGY, Dec. 1990, p. 6286–6290
0022-538X/90/126286-05$02.00/0

Vol. 64, No. 12

Analysis of Long Terminal Repeat Circle Junctions of Human Immunodeficiency Virus Type 1

JEFFREY S. SMITH,[1] SUNYOUNG KIM,[2] AND MONICA J. ROTH[1]*

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854,[1] and New England Deaconess Hospital and Harvard Medical School, Boston, Massachusetts 02215[2]

Received 25 June 1990/Accepted 4 September 1990

Circle junctions of unintegrated human immunodeficiency virus type 1 strain IIIB were analyzed after polymerase chain reaction amplification. Among the 28 colonies sequenced, eight unique circle junction species were detected. Five of the eight species resulted in circle junctions with larger inserts than predicted. A majority of these could result from heterogeneity in generating the U5′ long terminal repeat terminus.

Retroviral replication involves the conversion of the RNA genome into a double-stranded DNA copy (for reviews, see references 26 and 27). Conversion of the RNA genome to DNA is complex and is catalyzed solely by the viral reverse transcriptase. This process involves two priming events, transposition (or jumping) of DNA, RNA- and DNA-dependent DNA synthesis, and RNase H removal of viral RNA and RNA primers (26, 27). In this process, sequences are duplicated, producing the long terminal repeats (LTRs) (8, 26).

Three characterized DNA products are synthesized: a linear molecule containing a copy of the LTR at each terminus and two circular forms containing either one or two copies of the LTRs (26, 27). The boundary between the two LTRs in the circular form is referred to as the circle junction (25).

Specific sequences located at the termini of the LTRs are required for viral integration. In the viral RNA, these sequences are internal and adjacent to the plus- and minus-strand primers. Imprecision by reverse transcriptase in the priming of DNA synthesis or removal of the primer could therefore alter the position of the *cis*-acting sequences. In this study, the circle junctions of newly synthesized, unintegrated human immunodeficiency virus type 1 strain IIIB ($HIV\text{-}1_{IIIB}$) viral DNA were analyzed.

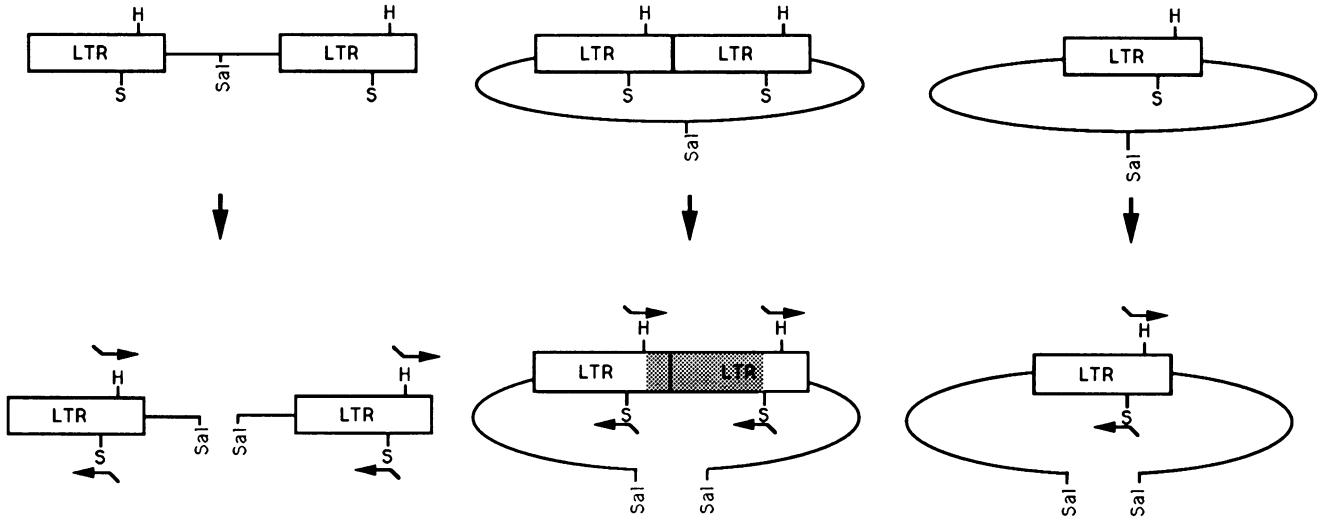

FIG. 1. Amplification of the circle junction sequence. The products of the polymerase chain reaction amplification procedure for the three unintegrated viral DNA species are shown. The linear molecule and the circular products containing one or two LTRs were digested with *Sal*I restriction enzyme. Two oligonucleotides containing the sequences 5′-*TACGAATTC*GCCTCAATAAAGCTTGCCTTG-3′ and 5′-*ATCGAA*T TCCCTAGTTAGCCAGAGAGCTCCC-3′ were synthesized. The former oligonucleotide encodes the *Hin*dIII (H) recognition site at nucleotide position 531 of HXB2 virus (17); the latter encodes the *Sac*I (S) recognition sequence at nucleotide position 487 of HXB2 (underlined). The nucleotides in italics do not hybridize with the viral DNA and were designed as possible *Eco*RI cloning sites for the amplified material. The arrows indicate the positions and orientations of the oligonucleotide primers. The circle junction sequence specifically amplified is stippled. The figure is not drawn proportionally; restriction sites only within the LTR are indicated.

Strategy for amplification of the circle junctions. To analyze the HIV-1 LTR termini, the circle junctions of newly synthesized circles were amplified by using the polymerase chain reaction. H9 cells were acutely infected with $HIV\text{-}1_{IIIB}$

* Corresponding author.

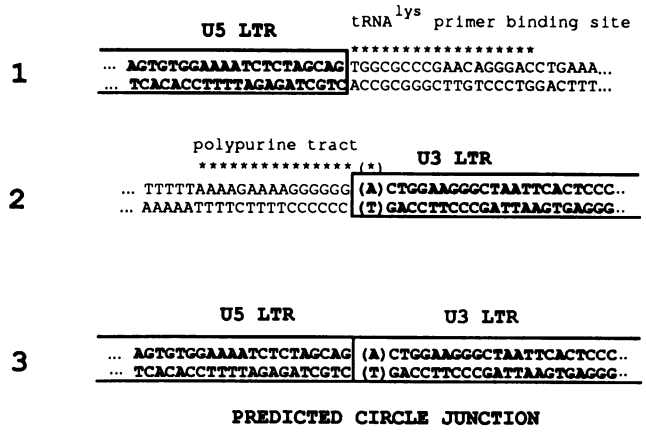

FIG. 2. Mechanism of generating the circle junction. The internal retroviral sequences which contribute to the circle junctions are indicated. Line 1 contains the U5 LTR sequence adjacent to the tRNA PBS. The 18 nucleotides that can base pair with the tRNA primer are indicated (*). Line 2 contains the U3 LTR sequence adjacent to the PPT. The sequences comprising the PPT are indicated (*). The A residue which may serve in the PPT or the U3 LTR is in parentheses. Line 3 contains the predicted circle junction produced after replication and ligation of the viral sequence.

and harvested 4 days postinfection when approximately 40% of the culture was infected (12). At this point, the largest amount of free viral DNA, including a low level of circular species (12), is found. Extrachromosomal viral DNA was isolated by the method of Hirt (9).

Figure 1 outlines the amplification strategy for the circle junctions. The DNA containing the three unintegrated viral forms was digested with *Sal*I, reducing the possibility that linear and single LTR circular species would be amplified. The region between and including the *Hin*dIII and *Sac*I restriction sites should be amplified, producing a 637-nucleotide product (Fig. 1, stippled region). The polymerase chain reaction mixture was assembled as described by Perkin Elmer Cetus. Because of the nonhomology of the primers in the initial steps, hybridization was performed at 55°C for five cycles and at 62°C for the remaining 25 cycles. After passage through a Centricon-30 filter (Amicon), DNA fragments 400 to 1,000 nucleotides in length were isolated from a 1.5% agarose gel by glass powder (28). This DNA was again amplified with the hybridization temperature at 68°C for 25 cycles. After two rounds of amplification, a 640-nucleotide species could be detected. This species decreased in size by approximately 45 nucleotides when digested with *Hin*dIII and *Sac*I and was the only unique product that hybridized with an LTR-specific probe on a Southern blot (data not shown). The DNA was digested with *Hin*dIII and *Sac*I, and the 590-bp fragment was subcloned into pTZ19U (13).

Generation of circle junctions. The LTR termini are determined by the positions of the plus- and minus-strand origins of replication. For HIV, the first 18 nucleotides of the 3′ terminus of the $tRNA^{Lys}$, including the CCA sequences added posttranscriptionally, base pair with the viral RNA (24) and serve as the primer for minus-strand synthesis (Fig. 2, primer-binding site [PBS]). As with other retroviruses, the first deoxynucleoside monophosphate added to the end of the primer tRNA is therefore identical to the 5′ nucleotide of the U5 terminus (Fig. 2, line 1).

The product of minus-strand DNA synthesis is an RNA-DNA hybrid. A site-specific cleavage catalyzed by the reverse transcriptase-encoded RNase H occurs at the 3′ terminus of a polypurine tract (PPT; Fig. 2, asterisks) located next to U3. This cleavage produces the primer for DNA-dependent DNA synthesis catalyzed by reverse transcriptase. The plus-strand cleavage site of HIV has not been directly identified. The RNase H cleavage site has been predicted to occur either immediately downstream of the string of six G residues (18) or one base into the U3 region (25). In the latter case, the A residue adjacent to the six G's would be part of the PPT rather than U3 and is marked with parentheses in Fig. 2, line 2.

On the basis of the positions of the PBS and the PPT, it is possible to predict the HIV LTR termini (Fig. 2, line 3). The U5 LTR terminal sequences would be $\cdots\frac{CAG}{GTC}$. Because of the uncertain position of the plus-strand origin, the U3 terminus could be either $\frac{ACTG}{TGAC}\cdots$ or $\frac{CTG}{GAC}\cdots$. The circle junction could therefore be either CAGACTG or CAGCTG (Fig. 2, line 3).

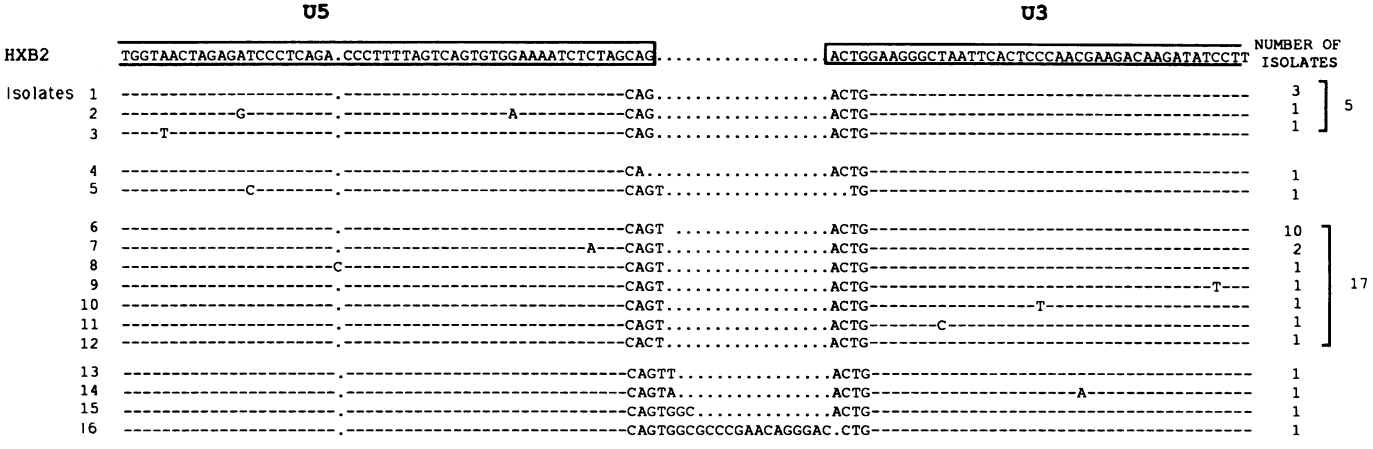

FIG. 3. Sequences of HIV-1_{IIIB} circle junctions. The sequences of the 28 circle junctions characterized are shown. Double-stranded plasmid DNA was used for sequencing, using the method of Sanger et al. (20) with Sequenase reagents (U.S. Biochemicals). For sequencing initiating at the *Hin*dIII (U5) terminus, the reverse primer (5′-AACAGCTATGACCATG-3′) was used. An oligonucleotide with the sequence 5′-GGTCAGTGGATATCTG-3′, corresponding to nucleotides 9194 to 9209 of HXB2, was synthesized and used as a primer for sequencing of the opposite strand. The U5 and U3 sequences of HXB2 are shown on the top line. Symbols: ---, sequence identical to that of HXB2; ···, gaps introduced for the alignment. The number of isolates containing the identical sequence is shown on the right.

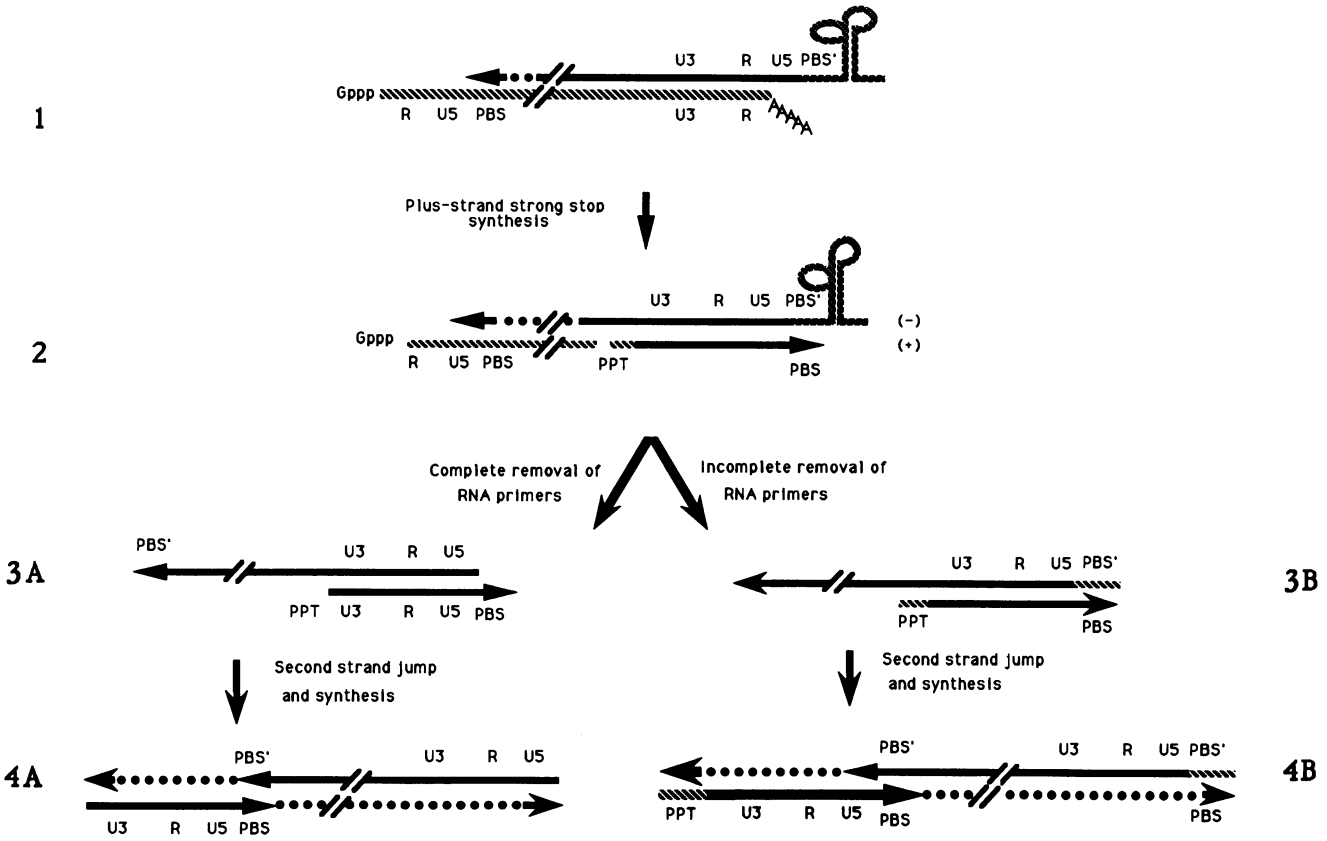

FIG. 4. Model for the incorporation of additional sequences at the LTR termini. ~~~, Viral RNA containing a cap (Gppp) and poly(A) tail; ▬, tRNA serving as the primer for minus-strand DNA synthesis. U3, R, and U5 sequences contributing to the LTR are indicated. PBS', Sequence complementary to the PBS. ■, minus- or plus-strand DNA synthesized by reverse transcriptase; ●●●, nascent DNA. Line 1, Synthesis of minus-strand DNA after the first jump, with the tRNA primer still present. Line 2, Plus-strand DNA synthesis is primed by the 3'-OH produced by a site-specific RNase H cleavage at the polypurine tract. Plus-strand DNA synthesis copies the first 18 nucleotides of the tRNA primer sequence. The second jump occurs by utilizing the base pairing between the PBS and the PBS'. Line 3A, RNase H activity should remove the tRNA primer sequence (PBS') and PPT RNA primers. Incomplete removal of the primers (line 3B) results in the duplication of these sequences upon completion of the minus- and/or plus-strand synthesis (line 4B).

Sequence analysis of the circle junctions. Twenty-eight plasmids containing the *Hin*dIII-*Sac*I insert were analyzed by sequencing (20) on both strands (Fig. 3). None of the junctions contained the sequence CAGCTG. Twenty-six of the clones had the U3 terminus $\frac{ACTG}{TGAC}\cdots$. This finding implies the reverse transcriptase maintains high specificity for the RNase H cleavage producing the plus-strand origin and for removal of the RNA primer after DNA synthesis. This cleavage occurs after the string of six G residues within the PPT (Fig. 2, line 2).

Surprisingly, only 5 of the 28 isolates contained the alternate predicted circle junction CAGACTG (Fig. 3, isolates 1 to 3). The most abundant circle junction, found in 16 of 28 examined, contained the sequence ···CAGTACTG··· (Fig. 3, isolates 6 to 11). One additional isolate also contained four bases between the CA-TG sequence used in integration (isolate 12), with a G→C base substitution within the junction. Analysis of the sequences generating the U5 LTR terminus indicates that the first nucleotide within the tRNA PBS encodes a T residue. Isolates 15 and 16 contained much larger sequences unambiguously encoded by the PBS. In isolate 15, the TGGC sequence is the first four nucleotides of the PBS. Even more convincing is isolate 16, in which the entire 18 nucleotides of the PBS have been included in the circle junction (see Fig. 2 for PBS sequence).

Heterogeneity at the U5 LTR terminus can be explained by either of two models. In the first model, the tRNA primer is incompletely removed during reverse transcription (Fig. 4). The PBS sequence is initially copied during plus-strand strong-stop synthesis, providing the sequence duplication required for the second jump. If the tRNA molecule is incompletely removed by RNase H, the remaining tRNA sequence can be copied again, during completion of the plus-strand synthesis. This model predicts that some linear viral DNA molecules could contain RNA moieties at their 5' termini. For Moloney murine leukemia virus, incomplete removal of the tRNA was frequently detected when the U5 LTR termini were altered by either deletion or insertion mutations (4, 5). The HIV-1 virus may have adapted to accommodate the loss of one nucleotide between the PBS and the site of integration through an alteration in RNase H cleavage. The results for HIV-1 and murine leukemia virus indicate that the specificity for RNase H cleavage may not simply be the junction between the DNA and RNA, but rather some sequence or spatial recognition.

An alternative model for the insertions is that the tRNA primer had deletions at its 3' terminus. Processing of the tRNA is found in copia, the *Drosophila* retroviruslike particle (11). However, isolate 16 could not exclusively result from this model because removal of all of the PBS' leaves no

region of homology for hybridization of the tRNA to viral RNA.

Isolates 13 and 14, containing the sequences CAGTT ACTG and CAGTAACTG, respectively, cannot be explained entirely by either model. The additional T could be encoded by the PBS; however, the additional A · T base pair cannot. Moreover, the A · T base pair cannot result from incomplete removal of the plus-strand RNA primer from the PPT, which would insert a G · C base pair. Analysis of circle junctions from the Schmidt-Ruppin A strain of Rous sarcoma virus has similarly identified insertions of a nontemplate-directed A · T base pair (14). A cellular enzyme with terminal transferase activity or reverse transcriptase itself (Bradley Preston, personal communication) may be adding nucleotides to the accumulating linear molecules. Alternatively, reverse transcriptase may add the nucleotides while attempting to hairpin loopback second-strand synthesis before the jumping reaction.

Circle junctions with small deletions were detected. Isolates 4 and 5 are both missing the terminal bases removed in the course of integration. If linear species with recessed 3′ termini are a detectable intermediate in HIV infection, then nuclease digestion of the single-strand overhang followed by intramolecular ligation would produce these species. Isolate 16 also lacks the terminal nucleotide from the U3 terminus.

Ten base substitutions and one nucleotide insertion previously unidentified were isolated which could have been caused by *Taq* polymerase or reverse transcriptase. Both enzymes lack proofreading functions and have high levels of misincorporation (1, 23). The U3 LTR encodes part of the *nef* gene. Within this region, isolate 14 (C→A) has been previously identified in HXB2. Isolates 9 and 10 maintain the amino acid which is read. Isolate 11 results in the substitution of Leu (CTA) with Pro (CCA).

Even if the amplification protocol did not proportionally increase the input virus pool directly, the total number of unique isolates (eight) suggests that the termini are not homogeneous. Three of the eight unique isolates result from a mechanism suggesting heterogeneity of the U5 sequence. Two additional isolates (13 and 14) could in part be explained by heterogeneity at the U5 terminus. Direct analysis of the linear viral DNA is an alternate approach to analysis of the termini of the viral DNA. This approach, however, has proven more difficult than has been the case for murine leukemia virus (19), possibly because of the heterogeneity of the termini.

It is not known whether any or all of these circle junctions in their linear precursor form (2, 7) would function during integration. Linear molecules resulting from incomplete removal of the tRNA could contain a ribonucleotide at the 5′ terminus and may not be recognized by the IN protein. This might explain the accumulation of high levels of unintegrated DNA observed with HIV encephalitis (15). Integration occurs site specifically 3′ to the CA dinucleotide and results in the loss of the terminal sequences, including the additional nucleotides identified. When the most abundant circle junction with the sequence CAGTACTG is in a linear form, the IN protein (3, 6, 10, 16, 19, 21) should remove two bases symmetrically from each LTR terminus. This mechanism of removing two bases from each end of the LTR has been generally found for other retroviruses studied (22).

This work was aided by grant MV-51430 from the American Cancer Society. Partial support was provided by the General Research Support of the Robert Wood Johnson Medical School. M.J.R. is a recipient of a Leukemia Society of America, Inc. Special Fellow award.

LITERATURE CITED

1. **Bebenek, K., J. Abbotts, J. D. Roberts, S. H. Wilson, and T. A. Kunkel.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J. Biol. Chem. **264**:16948–16956.
2. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA. Cell. **49**:347–356.
3. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86**:2525–2529.
4. **Colicelli, J., and S. P. Goff.** 1986. Structure of a cloned circular retroviral DNA containing a tRNA sequence between the terminal repeats. J. Virol. **57**:674–677.
5. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
6. **Donehower, L. A., and H. E. Varmus.** 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA **81**:6461–6465.
7. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
8. **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.
9. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–371.
10. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63**:5319–5327.
11. **Kikuchi, Y., Y. Ando, and T. Shiba.** 1986. Unusual priming mechanism of RNA-directed DNA synthesis in copia retrovirus-like particles of Drosophila. Nature (London) **323**:824–826.
12. **Kim, S., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J. Virol. **63**:3708–3713.
13. **Mead, D. A., E. Szczesha-Skorupa, and B. Kemper.** 1986. Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng. **1**:67–74.
14. **Olsen, J. C., and R. Swanstrom.** 1985. A new pathway in the generation of defective retrovirus DNA. J. Virol. **56**:779–789.
15. **Pang, S., Y. Koyanagi, S. Miles, C. Wiley, H. V. Vinters, and I. S. Y. Chen.** 1989. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. Nature (London) **343**: 85–89.
16. **Quinn, T. P., and D. P. Grandgenett.** 1988. Genetic evidence of the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. **62**:2307–2312.
17. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. J. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
18. **Rattray, A. J., and J. J. Champoux.** 1989. Plus-strand priming by Moloney murine leukemia virus: the sequence feature important for cleavage by RNase H. J. Mol. Biol. **208**:445–456.
19. **Roth, M. J., P. L. Schwartzberg, and S. P. Goff.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence of IN function and terminal DNA sequence. Cell **58**:47–54.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
22. **Skalka, A. M.** 1988. Integrative recombination in retroviruses, p. 701–724. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic

recombination. American Society for Microbiology, Washington, D.C.
23. **Tindall, K. R., and T. A. Kunkel.** 1988. Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry **27**:6008–6013.
24. **Van Beveren, C., J. Coffin, and S. Hughes.** 1985. Appendixes B, p. 1106–1123. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring, N.Y.
25. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* M. H. Howe and D. E. Berg (ed.), Mobile DNA. American Society for Microbiology, Washington D.C.
26. **Varmus, H., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring, N.Y.
27. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses, p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring, N.Y.
28. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.