JOURNAL OF VIROLOGY, Oct. 1990, p. 4903–4906
0022-538X/90/104903-04$02.00/0

Vol. 64, No. 10

Sequence of the Circle Junction of Human Immunodeficiency Virus Type 1: Implications for Reverse Transcription and Integration

J. M. WHITCOMB,[1] R. KUMAR,[2] AND S. H. HUGHES[1]*

ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201,[1] and Department of Molecular Biology, Squibb Institute for Medical Research, Princeton, New Jersey 08543-4000[2]

Received 9 April 1990/Accepted 27 June 1990

The sequence of the LTR-LTR circle junction of human immunodeficiency virus type 1 (HIV-1) was determined. The circle junction sequences were amplified by the polymerase chain reaction and cloned into M13 sequencing vectors. The circle junction contains 4 base pairs that are not present in the integrated provirus. We show that reverse transcription in HIV-1 initiates with the addition of a dC to the tRNA primer, suggesting that the tRNA used to initiate reverse transcription ends with the consensus CCA triplet. This indicates that the source of one of the four bases in the circle junction is probably the terminal A of the tRNA primer used to initiate reverse transcription. We propose that, in HIV-1, removal of the tRNA primer by RNase H cleavage shows an unusual specificity such that cleavage occurs between the terminal rA and the adjacent rC of the tRNA primer. These data also imply that the HIV-1 integration protein removes two bases from each end of the linear viral DNA during integration as has been described for other well-studied retroviruses.

After the penetration of a susceptible host cell, the first stage in the retroviral life cycle is the reverse transcription of the single-stranded viral RNA genome into a linear double-stranded DNA (19, 21). This linear form, which is made in the cytoplasm, migrates to the nucleus, where it is the precursor to the integrated provirus and to two circular forms that contain either one or two long terminal repeats (LTRs) (1, 2, 6, 8–10, 12, 16, 19, 21). The sequence of the integrated provirus differs from the linear integrative DNA precursor at both ends. In general, 2 base pairs (bp) are removed from each end during the integration event (1, 2, 8–10, 14, 19, 21). The loss of these bases is characteristic of retroviral integration, and the mechanism of loss is intimately related to the actual process of integration (1, 2, 8, 9, 14). Circles with two LTRs appear to be formed by blunt-end ligation of the linear double-stranded DNA molecule. Whether this ligation event is catalyzed by cellular or viral enzymes is unclear; however, the bases that are lost from the ends of the linear viral DNA during integration are present at the LTR-LTR junction in the circles with two LTRs. Comparison of the circle junction sequence with the sequence at the boundaries of integrated proviral DNA allows a determination of the bases lost during integration. This information was instrumental in developing our current understanding of the integration process.

Previous analyses of integrated human immunodeficiency virus type 1 (HIV-1) proviruses predicted that either two or three extra bases would be present at the circle junction as opposed to the four extra bases for other well-studied retroviruses (1, 2, 8–10, 14, 19, 21). This prediction was primarily based on the definition of the right U5 boundary as the first base added onto the tRNA primer. All known tRNAs end with the triplet CCA, so the base following this triplet should define the right end of U5. In contrast to other retroviruses that have two bases between this position and the end of the integrated provirus, HIV-1 has only one base (13, 15, 20). This prediction suggested that the integration process for HIV-1 would differ from that described for other retroviruses. Either the integration protein would remove one base from each end of the linear DNA (instead of two), or it would act asymmetrically and remove one base from one end and two from the other.

The data we present here show that there are four bases present at the circle junction that are missing from the integrated provirus. While this suggests that integration of the HIV-1 linear DNA is quite similar to the integration of the genomes of other retroviruses, it raises the question of the origin of the fourth base. Examination of the HIV-1 sequence implies that the fourth base could be derived from the terminal A of the tRNA primer. There are two obvious possibilities: either the A is incorporated as a dA onto a tRNA primer that ends in CC or a specific RNase H cleavage leaves the terminal rA from the tRNA attached to the DNA strand.

We present data showing that reverse transcription in HIV-1 is initiated with the addition of dC to the tRNA primer. We predict that the removal of the tRNA primer by RNase H will show an unusual specificity and that the terminal rA of the tRNA primer will be the base found at the right end of the linear double-stranded DNA integrative precursor.

MATERIALS AND METHODS

Viral DNA and RNA. Cell lysates of HIV-1 (strain h5151/Rf)-infected H9 cells were kindly provided to us by Mel Campbell, George Pavlakis, and Barbara Felber. Low-molecular-weight DNA was enriched by Hirt extraction. HIV-1 RNA was prepared from rapid harvest virus provided to us by Julian Bess, Jr., and Larry Arthur. The virus was lysed in TENS (10 mM Tris chloride [pH 7.4], 100 mM NaCl, 5 mM EDTA, 0.5% sodium dodecyl sulfate), extracted twice with an equal volume of phenol-chloroform (1:1), and ethanol precipitated overnight at −20°C. The pellet was dried and suspended in a small volume of diethyl pyrocarbonate-treated sterile water. HIV-1 RNA was further enriched by two cycles of annealing and elution on oligo(dT) spin columns (Pharmacia, Inc., Piscataway, N.J.). The viral RNAs

* Corresponding author.

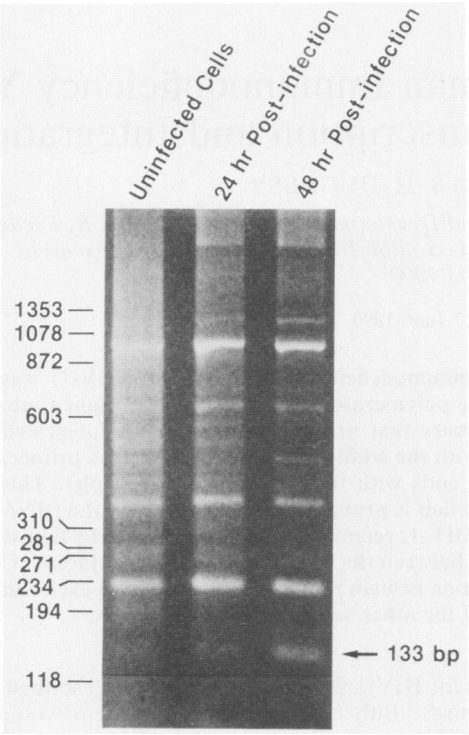

FIG. 1. PCR amplification of the circle junction region of HIV-1. The products of the amplification reaction were fractionated on a 2% low-melting-temperature agarose gel and visualized by staining with ethidium bromide. The large arrow indicates the 133-bp band containing the circle junction. Numbers on left show size in base pairs.

were then applied to a sucrose step gradient (0, 5, 10, 15, and 20%) and sedimented for 2 h at 35,000 rpm in an SW40 rotor. The optical density at 260 and 280 nm of 0.5-ml fractions was determined, and peak fractions were ethanol precipitated and suspended in 50 μl of diethyl pyrocarbonate-treated sterile water.

RCAS is a replication-competent retroviral vector derived from the SR-A strain of Rous sarcoma virus by deletion of the *src* gene (11). Sequences of the LTRs and flanking sequences are identical to those of Rous sarcoma virus. RNA was prepared from rapid harvest virus and purified by sedimentation on a sucrose step gradient in parallel with the HIV-1 RNA.

PCR. Polymerase chain reaction (PCR) was performed as follows. Final reaction conditions were 50 mM Tris chloride (pH 8.3), 50 mM KCl, 7 mM $MgCl_2$, 170 μg of bovine serum albumin per ml, 1 mM each dATP, dCTP, dGTP, and dTTP, 1 μM each primers 1605 and 1607, 8% dimethyl sulfoxide, 1 μg of DNA, and 2 U of *Taq* polymerase (U.S. Biochemical Corp., Cleveland, Ohio) in a total volume of 25 μl. Thirty cycles were performed in a Perkin-Elmer Cetus DNA Thermal Cycler by using a step-cycle file programmed for 89°C, 40 s, 62°C, 1 min 20 s. Primer sequences were as follows: 1605 (U3), 5′-TCAGGGAAGTAGCCTTGTGT-3′; 1607 (U5), 5′-TAACTAGAGATCCCTCAGAC-3′.

Cloning and sequencing. The HIV-1-specific band at approximately 130 bp was excised and extracted from a 2% low-melting-temperature agarose gel. After digestion with *Mbo*I, the 85-base fragment containing the circle junction (see Fig. 2b) was isolated on low-melting-temperature agarose and ligated to M13mp18 and M13mp19 cut with *Bam*HI. The ligated DNA was used to transform competent *Escherichia coli* DH5α (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). M13 clones with inserts were processed and sequenced according to the Sequenase DNA Sequencing Kit protocol (U.S. Biochemical Corp.). The amplified *Mbo*I fragment was also digested with *Sca*I or *Eco*RV in a 20-μl volume. The resulting fragments were radioactively labeled by filling in the *Mbo*I overhangs with Klenow fragment in the presence of [^{32}P]dATP and [^{32}P]dGTP and then chasing with unlabeled deoxynucleoside triphosphates.

Primer tRNA labeling. Viral RNA (1.5 μl of 50-μl sample) was added to a reverse transcription reaction mixture containing 25 mM Tris chloride (pH 8.0), 50 mM KCl, 10 mM $MgCl_2$, 2 mM dithiothreitol, 1 μg of purified HIV-1 reverse transcriptase (3), and 0.25 μCi of [^{32}P]dATP, [^{32}P]dCTP, [^{32}P]dGTP, or [^{32}P]dTTP in four separate reactions. The reaction mixes were incubated at 45°C for 15 min. One-half of the reaction was treated with a mixture of RNases A and T_1. The samples were heated to 100°C in 95% formamide, separated on a 6% polyacrylamide sequencing gel, and visualized by autoradiography.

RESULTS AND DISCUSSION

We used the PCR to amplify the junction sequences from the two LTR circles present in a Hirt extract of HIV-1-infected H9 cells. We used rapid temperature cycling and stringent annealing temperatures to increase the specificity of the PCR. In these experiments, we were unable to define the conditions so that only one fragment was amplified. However, a specific band of the predicted size was amplified, in addition to numerous nonspecific bands. The specific band of ca. 130 bp was not amplified when DNA from uninfected cells was used but appeared in PCR amplifications of DNA samples prepared from cells 24 to 48 h after infection with HIV-1 (Fig. 1). This band contained approximately 50 bp of U5 and 80 bp of U3 (Fig. 2b). We used the

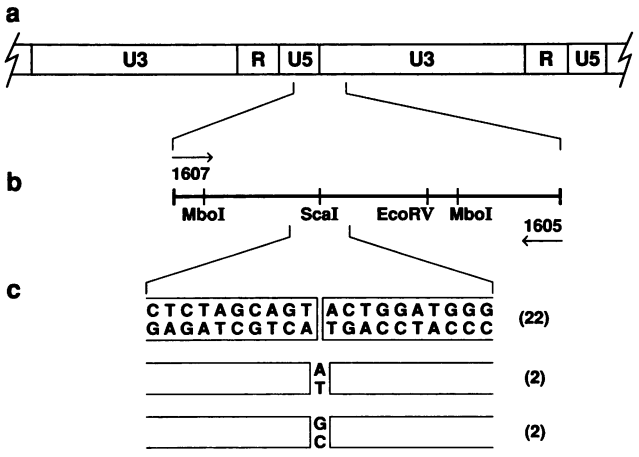

FIG. 2. Schematic representation of the circle junction region of HIV-1. (a) Predicted orientation of the LTRs at the circle junction. (b) Restriction map of the 133-bp PCR-generated fragment. The *Mbo*I sites used for cloning are indicated. PCR primer positions are indicated by arrows. (c) Sequences of the circle junction clones. The number in parentheses is the number of clones seen with each of the indicated sequences.

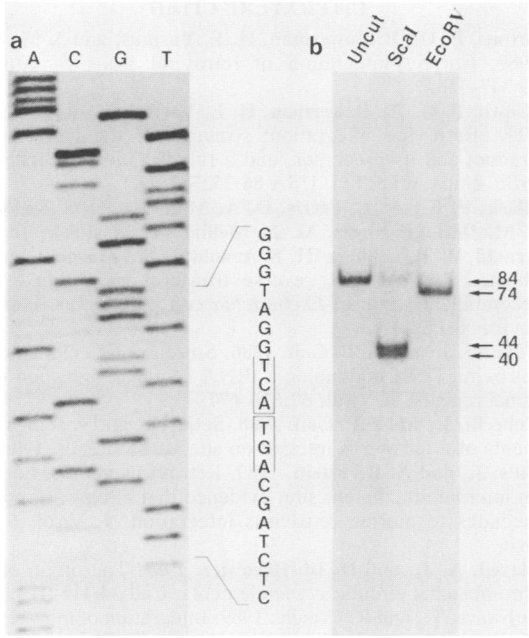

FIG. 3. (a) Sequence of HIV-1 circle junction clone. Four bases, GTAC, are found between the conserved CA and TG dinucleotides that bound the integrated provirus. (b) Restriction analysis of the PCR-generated fragment containing the circle junction sequences. The preparation of the labeled, amplified segment is described in Materials and Methods. Aliquots of the amplified segment were digested with the indicated restriction enzyme (*Sca*I or *Eco*RV). The digestion products were fractionated on a 15% acrylamide gel and visualized by autoradiography. Incomplete *Sca*I digestion is consistent with a subpopulation of the amplified segment having variant sequences at the circle junction (see text).

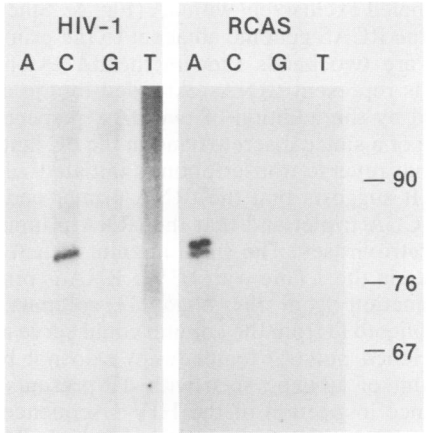

FIG. 4. Single DNA base extensions of the tRNA primers of HIV-1 and RCAS. The conditions used in the extension reaction are given in Materials and Methods. Each lane contains the reaction products obtained by extending the tRNA primer with purified HIV-1 transcriptase (3) in the presence of the indicated [α-^{32}P] deoxynucleoside triphosphate. The reaction products were denatured by heating to 100°C in formamide and fractionated on a 6% sequencing gel. The position of migration of DNA size markers is shown at the right of the figure. RCAS (a derivative of Rous sarcoma virus) initiates reverse transcription with dA, while HIV-1 initiates with dC. There are two bands in the RCAS dA lane because the first two bases that are incorporated are both dA and there are tRNAs extended by the addition of one and two bases.

*Mbo*I sites flanking the circle junction to clone the fragment into M13 vectors for sequencing. Twenty-six clones were sequenced through the circle junction, and a consensus sequence was determined (Fig. 2c). Four bases (GATC) were present between the conserved CA and TG dinucleotides that are found at the boundaries of the integrated HIV-1 provirus. These four bases were present in all 26 clones sequenced. The clones had single-base changes at sites outside the circle junction, indicating that they derive from several independent amplification and cloning events.

The sequence we determined predicts that a novel *Sca*I site (AGTACT) would be generated when the circle junction is formed. Figure 3b shows that *Sca*I digestion of the amplified circle junction DNA yielded fragments of the expected size. It is interesting that the digestion of the amplified circle junction by *Sca*I is not complete. Of the 26 clones sequenced, 4 had an additional base present precisely at the circle junction (Fig. 2c). The origin of these extra bases is not immediately apparent. The G · C base pair could have been incorporated from the polypurine tract adjacent to U3 or from the tRNA terminus adjacent to U5. The A · T insertion, however, could not arise from the HIV-1 genome by using a simple model based on reverse transcription. We suggest that during the formation of the circle junction, extra bases can occasionally be incorporated with no apparent specificity. These were the only insertions into the sequence we observed, which suggests that they were not PCR artifacts. Most PCR artifacts are reported to be either substitutions or deletions, and the frequency of insertions is believed to be very low (17). Rare single-base insertions of this type have been seen previously at the circle junctions of other retroviruses (5, 18) and of *copia* (7). Larger inserts have also been seen (4). Since we size selected the amplified segment before cloning, any larger insertions would probably have been missed by our procedure.

Digestion of the amplified circle junction with *Sca*I provides a strong argument that the sequence determination is correct. It is also fortuitous for research on the mechanism of HIV integration. Although the circles are not the precursors of the integrated proviral DNA, digestion of the circle junction with *Sca*I generates what we believe to be the exact sequence at the linear ends of unintegrated HIV DNA, which should be an appropriate substrate for HIV-1 integration machinery (2, 8, 9, 14). The availability of large quantities of substrate should facilitate study of the integration reaction. We predict that the HIV-1 integration protein will remove two bases from each of the 3′ ends of this DNA molecule.

How then does HIV-1 generate the right end of the linear DNA? The right end of the linear DNA molecule is defined by two discrete events. First, viral DNA synthesis is primed from a cellular tRNA, and second, RNase H removes the tRNA primer from the newly synthesized DNA molecule. We determined which base HIV-1 uses to initiate reverse transcription. Intact viral RNA was purified from virions. In retroviral virions, the RNA genome is present as a dimer with the cellular tRNA primers already annealed. The viral RNAs were introduced into a series of in vitro reverse transcription reactions in which one of the four individual deoxynucleoside triphosphates was present in each of four separate reactions with purified HIV-1 reverse transcriptase (3).

We used RCAS, a vector derived from Rous sarcoma virus (11), as a control. As expected, the tRNA primer of

RCAS is labeled exclusively with dA (Fig. 4). Since there are two Us in the RCAS genome adjacent to the primer-binding site, there are two bands seen in the dA extension lane. These bands represent tRNAs extended by the addition of one dA and by the addition of two dAs, respectively. The appearance of a single discrete band in the dC lane of HIV-1 indicates that reverse transcription is initiated with a dC in this virus. It suggests that the tRNA primer ends with the conserved CCA triplet and that the tRNA priming event is typical of retroviruses. The small amount of heterogeneous priming seen in the T lane with HIV-1 RNA is probably due to contamination from the oligo(dT) columns. A small amount of oligo(dT) from the column could serve as a primer for the enriched poly(A) template. We do not believe the small amount of labeling seen with dT presents a serious problem since inspection of the HIV-1 sequences predicts that either dC or dA will be incorporated. The labeled tRNAs from the RCAS and HIV-1 reactions are, as expected, degraded by RNase or by treatment with base (data not shown).

The fact that HIV-1 initiates reverse transcription with a dC, as predicted from the sequences of cloned HIV-1 proviruses, means that the A predicted to be at the right end of the linear double-stranded DNA molecule must be an rA. We suggest that the presence of this rA does not adversely affect processing of the ends of the linear molecule by the integration protein.

We are now attempting to determine the specificity with which the RNase H of HIV-1 reverse transcriptase removes the tRNA primer sequences. There have been numerous reports in the literature of circle junction clones containing bases derived from the tRNA primer (5, 18) and one report of a clone containing a nearly complete copy of a tRNA at this position (4). We suggest that HIV has evolved to routinely incorporate this rA into the linear DNA that serves as the precursor to the integrated provirus. This finding implies that there is a strong selection to maintain symmetry at the ends of the linear integrative DNA precursor. The 4-bp spacing seen at the circle junction between the conserved CA and TG dinucleotides reflects the preference of the integrative machinery for symmetry at the ends of the linear DNA precursor.

ACKNOWLEDGMENTS

We are grateful to Mel Campbell, Barbara Felber, and George Pavlakis for the HIV-1-infected cells, to Julian Bess, Jr., and Larry Arthur for the gift of rapid harvest HIV-1 virions, to Marilyn Powers for the oligonucleotides, and to Hilda Marusiodis for typing the manuscript.

This research was sponsored by the National Cancer Institute under contract no. N01-C0-74101 with ABL.

ADDENDUM IN PROOF

We used HIV-1 (h515/Rf) for these studies. The sequence of this strain of HIV-1 differs significantly from the sequence of the HXB_2 strain which was used to design PCR primers. We have repeated the experiments with HXB_2-infected H9 cells and have found the same four bases present at the circle junction. In addition, the PCR amplification of Hirt DNA from HXB_2-infected H9 cells yielded a single band of the expected size as opposed to the large number of bands seen in Fig. 1 of this report.

LITERATURE CITED

1. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA *in vitro*. Cell **49**:347–356.
2. **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.
3. **Clark, P. K., A. L. Ferris, D. A. Miller, A. Hizi, K.-W. Kim, S. M. Deringer-Boyer, M. L. Mellini, A. D. Clark, Jr., G. F. Arnold, W. B. Lebherz III, E. Arnold, G. M. Muschik, and S. H. Hughes.** 1990. HIV-1 reverse transcriptase purified from a recombinant strain of *Escherichia coli*. AIDS Res. Hum. Retrovirus **6**:753–764.
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. **Ellis, J., and A. Bernstein.** 1989. Retrovirus vectors containing an internal attachment site: evidence that circles are not intermediates to murine retrovirus integration. J. Virol. **63**:2844–2846.
7. **Flavell, A. J., and D. Ish-Horowicz.** 1983. The origin of extrachromosomal circular *copia* elements. Cell **34**:415–419.
8. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86**:3065–3069.
9. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
10. **Goff, S. P.** 1986. Retroviral integration and insertional mutagenesis, p. 325–361. *In* R. Kucherlapati (ed.), Gene transfer. Plenum Publishing Corp., New York.
11. **Hughes, S. H., J. J. Greenhouse, C. J. Petropoulos, and P. Sutrave.** 1987. Adaptor plasmids simplify the insertion of foreign DNA into helper-independent retroviral vectors. J. Virol. **61**: 3004–3012.
12. **Lobel, L. I., J. E. Murphy, and S. P. Goff.** 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. J. Virol. **63**:2629–2637.
13. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starich, S. F. 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**:277–294.
14. **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 on IN function and terminal DNA sequence. Cell **58**:47–54.
15. **Sanchez-Pescador, R., M. D. Power, P. J. Barr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, A. Renard, A. Randolph, J. A. Levy, D. Dina, and P. A. Luciw.** 1985. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science **227**:484–492.
16. **Shank, P. R., and H. E. Varmus.** 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. J. Virol. **25**:104–114.
17. **Tindall, K. R., and T. A. Kunkel.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry **27**:6008–6013.
18. **Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. M. Verma.** 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell **27**:97–108.
19. **Varmus, H. E.** 1988. Retroviruses. Science **240**:1427–1435.
20. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40**:9–17.
21. **Weiss, R., N. Teich, H. E. Varmus, and J. Coffin (ed.).** 1985. RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.