JOURNAL OF VIROLOGY, Nov. 1990, p. 5656–5659
0022-538X/90/115656-04$02.00/0

Vol. 64, No. 11

Removal of 3′-OH-Terminal Nucleotides from Blunt-Ended Long Terminal Repeat Termini by the Avian Retrovirus Integration Protein

AJAYKUMAR C. VORA, MICHAEL L. FITZGERALD, AND DUANE P. GRANDGENETT*

Institute for Molecular Virology, St. Louis University Medical Center, 3681 Park Avenue, St. Louis, Missouri 63110

Received 21 February 1990/Accepted 27 July 1990

The avian myeloblastosis virus integration protein (IN) was capable of removing a specific set of 3′-OH-terminal nucleotides from blunt-ended long terminal repeat (LTR) substrates which resembled linear viral DNA in vivo. The 3′-OH-recessed ends map to the in vivo site of integration on linear viral DNA. The linear DNA plasmid substrate was formed by the generation of a unique *Dra*I restriction enzyme site (TTT/AAA) at the circle junction of a 330-bp tandem LTR-LTR insert. IN preferentially released the three T nucleotides from the minus strand of the U3 LTR substrate compared with its ability to remove the three T nucleotides from the plus strand of the U5 LTR substrate. It was also observed that IN was capable of cleaving a non-LTR DNA substrate containing sequence homology to the U5 LTR terminus.

Retroviruses integrate their DNA by a nonhomologous recombination mechanism. This essential integration step for virus replication requires a DNA-binding endonuclease which is derived from the 3′-terminal region of the polymerase (*pol*) gene (7, 12). This proteolytically processed endonuclease, designated IN, has been demonstrated genetically to be essential for integration in a variety of retrovirus species (6, 19, 20, 22).

The precursor to the Moloney murine leukemia retrovirus integrated form of viral DNA appears to be linear (1, 2, 9, 10) and not a circular molecule containing a tandem long terminal repeat (LTR)-LTR junction (8, 16). Following synthesis of the Moloney murine leukemia retrovirus blunt-ended linear viral DNA, IN removes a set of two bases from both 3′-OH termini (3, 10, 21). Mutations in the Moloney murine leukemia retrovirus IN have demonstrated that the protein is necessary to generate these recessed ends (3, 21). The terminal nucleotides on each end of linear retrovirus DNA contain imperfect short inverted repeat sequences of approximately 15 nucleotides. The LTR inverted repeats are essential for integration (4, 5, 18). However, a single nucleotide deletion or an addition of two nucleotides at the blunt-ended U5 LTR terminus can be tolerated with no apparent effect on retrovirus integration (4, 21). The position of the 3′-OH-recessed ends generated by IN with wild-type LTR termini or with LTR terminal mutants described above always correlated directly with the site of integration at the 3′-OH-recessed ends of linear viral DNA (21).

In this study, a linear DNA substrate which closely resembled an in vivo linear avian retrovirus viral DNA molecule, the apparent precursor to the integrated form (Fig. 1), was created. By the addition of a single nucleotide at each of the termini, a unique blunt-ended *Dra*I restriction site was generated at the circle junction of two tandem LTRs. This plasmid (pAV-*Dra*I) was digested with *Eco*RI, dephosphorylated by calf intestine phosphatase, 5′-end labeled with [γ-^{32}P]ATP and T4 kinase (13). The *Eco*RI-labeled fragment was then digested with *Dra*I and either *Hin*fI or *Sph*I, allowing for the isolation of a 181-bp, U3 (left) LTR fragment and a 155-bp, U5 (right) LTR fragment, respectively (Fig. 1). The fragments were gel purified, and the specific activity of the labeled fragments was generally 3×10^4 to 4×10^4 cpm/ng of DNA (13). Purified avian myeloblastosis virus (AMV) IN (12, 15) was incubated with end-labeled DNA fragments under low- or high-salt conditions. For low-salt conditions, the reaction mixture (usually 50 μl or multiples thereof) contained 10 mM Tris hydrochloride (pH 7.5), 2 mM dithiothreitol (DTT), 25 mM NaCl, 0.1 mM EDTA, bovine serum albumin at 250 μg/ml, and 5 mM $MgCl_2$ or 1 mM $MnCl_2$. For high-salt conditions, the reaction mixture contained the components listed above, except that NaCl was 145 mM, $MgCl_2$ was 20 mM, and no bovine serum albumin was present. The endonuclease reactions containing the various DNAs were initiated at 37°C by the addition of enzyme. Normal times of incubation were 1 h for $MgCl_2$ and 15 min for $MnCl_2$. Reactions were terminated by adding 100 μl of a stop mixture containing 10 mM EDTA, 450 mM sodium acetate (pH 7.5), and 50 μg of yeast tRNA per ml. Samples were extracted once each with an equal volume of pH 7.0-adjusted phenol and chloroform-isoamyl alcohol (24:1). The extracted samples were processed for analysis on 6% polyacrylamide DNA sequencing gels as previously described (13, 17). The amount of end-labeled fragments per experiment varied only by approximately 10% in any lane as determined by Cerenkov counts per minute. The gels were exposed to X-ray film in the absence of an intensifying screen.

In the presence of Mg^{2+}, AMV IN released the terminal three T nucleotides from both the plus strand of the R-LTR terminal fragments (Fig. 2A and B, top dots) and the minus strand of the L-LTR terminal fragments (Fig. 2C). Thus, these data confirm previous in vitro (14) and in vivo (3, 10, 21) observations of IN cleaving the terminal sequences of LTR DNA at the in vivo sites of integration (Fig. 1). The prominent 5′-end-labeled cleaved fragment produced by IN with either LTR substrate comigrated with the third T residue from the LTR terminus derived by the Maxam and

* Corresponding author.

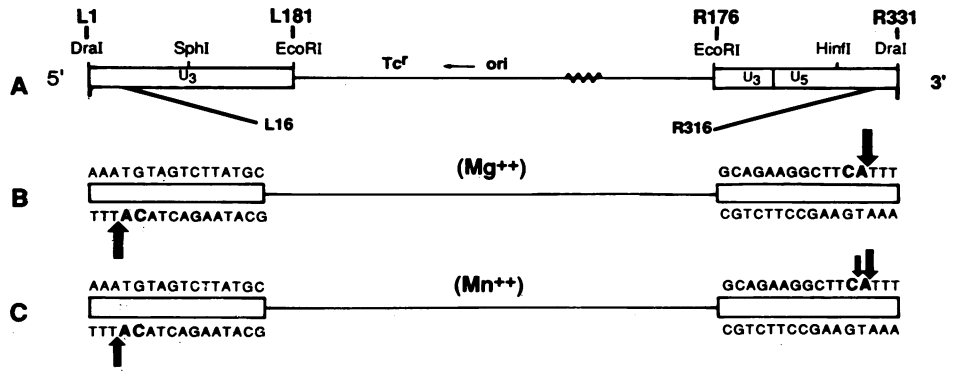

FIG. 1. Schema of the linear plasmid (pAV-*Dra*I) containing blunt-ended LTR termini. (A) A unique *Dra*I restriction site was generated by site-directed mutagenesis between two adjacently cloned LTRs, resulting in the addition of one nucleotide per terminus. L1 to L181, terminal repeat sequences adjacent to the *gag* region on the left end of the linear map; R176 to R331, terminal repeat sequences adjacent to the *src* region on the right end of the linear map of unintegrated avian sarcoma retrovirus DNA (13). The unique LTR 3'-terminal (U3) and 5'-terminal (U5) sequences are identified. Tc^r, Tetracycline resistance gene; ori, origin of replication. (B) The short inverted repeats at the ends of the LTR DNA are shown. Arrows indicate the major sites of cleavage by AMV IN in the presence of 20 mM Mg^{2+} and high salt levels. (C) Arrows indicate sites of cleavage by AMV IN in the presence of 1 mM Mn^{2+} and high salt levels.

Gilbert chemical cleavage reaction (17). The comigration of these DNA fragments demonstrated that the nick mapped immediately 3' to the CA sequence on the plus viral strand of the R-LTR terminus and also 3' to the AC sequence on the minus viral strand of the L-LTR terminus (Fig. 1B) (11–14, 23).

In the presence of Mn^{2+}, AMV IN preferentially nicks supercoiled DNA, on one strand or the other, either two or three nucleotides from the LTR circle junction (11, 13, 23). With one added nucleotide per terminus, the AMV IN enzyme also removed three or four nucleotides from the 3'-OH end of the R-LTR terminus of pAV-*Dra*I in the presence of Mn^{2+} (Fig. 1C and 2A and B). The hydrolysis of this third or fourth phosphodiester bond from the terminus by IN varied depending on the protein-DNA ratio, salt conditions, and time of incubation. However, as with Mg^{2+}, AMV IN in the presence of Mn^{2+} removed only the three T nucleotides from the L-LTR terminus (Fig. 2C) mapping 3' to the AC sequence on the minus viral strand (Fig. 1C). As demonstrated previously, AMV IN, using Mn^{2+} as divalent metal ion, also nicked DNA at a variety of sites (Fig. 2).

Conditions which permitted the most efficient removal of nucleotides from the R- or L-LTR terminal substrates were established (data not shown). In comparison with low-salt conditions, the rate of hydrolysis of the three T nucleotides by IN was increased (twofold) at the optimum NaCl concentration of 145 mM. The nuclease activity of IN was inhibited at higher NaCl concentrations. At 145 mM NaCl, the optimum concentration for $MgCl_2$ was between 15 and 20 mM and the optimum concentration for $MnCl_2$ was 1 mM. In comparison with results obtained with 5 mM $MgCl_2$, IN cleavage of the three T nucleotides from the L-LTR DNA fragment was increased approximately fourfold at the high-salt concentration with 20 mM $MgCl_2$ and with excess DNA.

Equal quantities of the R- and L-LTR DNA fragments were incubated simultaneously in the same reaction mixture with IN under high-salt conditions, either with Mg^{2+} (Fig. 3) or Mn^{2+} (data not shown). With Mg^{2+}, the L-LTR fragment was cleaved by IN approximately two to three times more efficiently than the R-LTR fragment at the optimal protein-to-DNA ratios (Fig. 3, lanes 4 to 6). With Mn^{2+}, the rate of cleavage by IN of the L-LTR fragment was again higher than the R-LTR (data not shown). The ability of IN to cleave either LTR DNA fragment was inhibited at the higher protein concentrations (Fig. 3, lanes 7 to 10). These data demonstrated that the L-LTR DNA fragment was preferentially cleaved by IN in the presence of an equal concentration of the R-LTR fragment.

Several linear blunt-ended DNA substrates possessing no or partial sequence relationship to the LTR termini were incubated with AMV IN. IN did not remove any detectable 3'-OH-terminal nucleotides smaller than 7 nucleotides from a blunt-ended *Eco*RV fragment from pBR322 in the presence of either Mg^{2+} or Mn^{2+} at 145 mM NaCl (Fig. 4). Several minor DNA fragments were evident upon incubation of IN with this *Eco*RV DNA fragment in the presence of Mn^{2+} (Fig. 4, lane 3). Under identical conditions, IN removed the usual three T nucleotides from the minus strand of the L-LTR DNA substrate derived from pAV-*Dra*I (Fig. 4). In contrast, AMV IN with either divalent cation removed two sets of nucleotides from the 3' strand of a blunt-ended *Ssp*I fragment (Fig. 4). Comigration experiments of the cleaved fragments with chemical sequence markers demonstrated that IN effectively removed two nucleotides from the *Ssp*I terminus (Fig. 4, top dot). Likewise, at a reduced rate, a set of five nucleotides which map adjacent to the CA sequence was removed (Fig. 4, bottom dot). The *Ssp*I fragment contained 8 bp which were nearly identical to the short inverted repeat sequences of U5 (Fig. 4, underlined sequences).

Katzman et al. (14) previously demonstrated that AMV IN was capable of releasing two nucleotides from double-stranded oligonucleotide LTR substrates, mimicking the ends of linear avian retrovirus DNA. The minimum length required for IN nicking was a 15-bp duplex containing the entire short invert repeat located at each LTR terminus. The R- and L-LTR DNA fragments derived from pAV-*Dra*I used in this study were at least 10 times longer than this 15-bp oligonucleotide LTR substrate. Comparison of the results obtained by Katzman et al. (14) with results in this report suggests little direct participation of LTR sequences, outside the short inverted repeats, which are necessary for

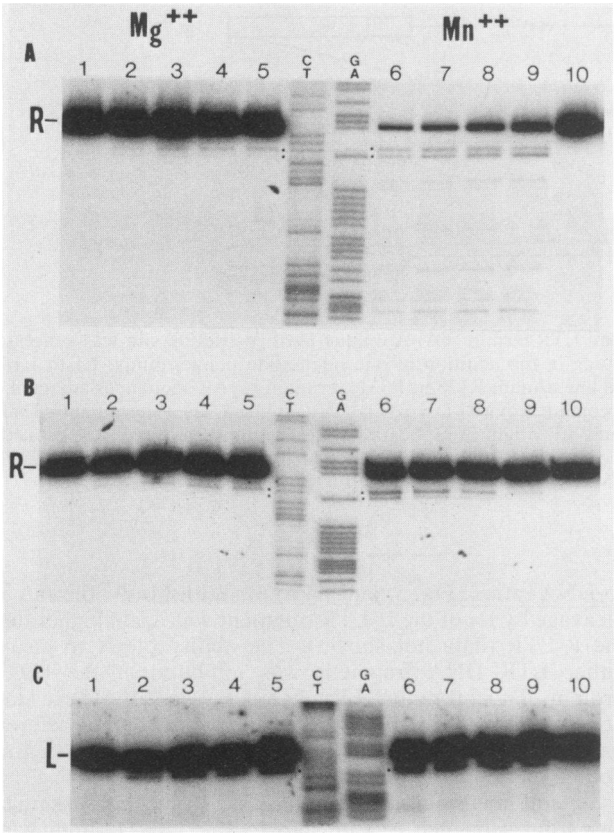

FIG. 2. Specificity of AMV IN nuclease activities on the R- and L-LTR termini of pAV-*Dra*I. (A) The R-LTR terminus (11 ng, 155 bp) was incubated with AMV IN at 9 μg/ml under low-salt conditions with either 5 mM Mg^{2+} or 1 mM Mn^{2+}. The molar mass ratio of protein to DNA was 360. Samples were removed at the indicated times, processed, and analyzed on a polyacrylamide DNA sequencing gel. Nearly equal amounts of each sample per experiment, as defined by Cerenkov counts per minute, were analyzed in each lane. Lane 1, Control DNA incubated for 180 min without IN in the presence of Mg^{2+}; lanes 2 to 5, DNA incubated with IN for 30, 60, 120, and 180 min, respectively; lanes $\frac{C}{T}$ and $\frac{G}{A}$, C+T and G+A chemical markers, respectively; lane 10, control DNA incubated for 90 min without IN in the presence of Mn^{2+}, lanes 9 to 6, DNA incubated with IN for 20, 40, 60, and 90 min, respectively. Dots to the left and right of the chemical markers in each panel identify relevant cleaved fragments (Fig. 1). (B) The R-LTR terminus (25 ng) was incubated with AMV IN at 4.7 μg/ml under high-salt conditions with either 20 mM Mg^{2+} or 1 mM Mn^{2+}. The protein-to-DNA ratio was 75. Analysis was performed as described for panel A. Lane 1, Control DNA incubated for 120 min without IN; lanes 2 to 5, DNA incubated with IN for 30, 60, 90, and 120 min, respectively; lanes $\frac{C}{T}$ and $\frac{G}{A}$, C+T and G+A chemical markers, respectively; lane 10, control DNA incubated for 120 min without IN; lanes 9 to 6, DNA incubated with IN for 30, 60, 90, and 120 min, respectively. (C) The L-LTR terminus (28 ng) was incubated with AMV IN at 4.7 μg/ml under high-salt conditions with either 20 mM Mg^{2+} or 1 mM Mn^{2+}. The protein-to-DNA ration was 76. Analysis was performed as indicated for panel A. Lane 1, Control DNA; lanes 2 to 5, DNA incubated with IN at 30, 60, 90, and 120 min, respectively; lanes $\frac{C}{T}$ and $\frac{G}{A}$, C+T and G+A chemical markers respectively; lane 10, control DNA; lanes 9 to 6, DNA incubated with IN for 15, 30, 60, and 120 min, respectively. The chemical cleavage between the TA nucleotides at the *Dra*I cleavage site was not always efficient. The exposure times of gels in panels A, B, and C were different.

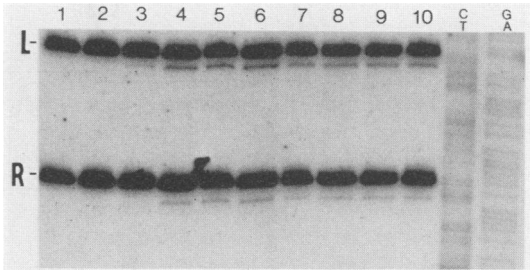

FIG. 3. Selective cleavage of the L-LTR DNA terminal substrate by AMV IN. AMV IN was incubated simultaneously with both the L-LTR fragment (7.5 ng) and the R-LTR fragment (7.5 ng) derived from pAV-*Dra*I with 20 mM Mg^{2+} under high-salt conditions for 1 h. The samples were analyzed as described for Fig. 2. Lane 1, incubated control DNA; lanes 2 to 10, IN at concentrations of 0.12, 0.24, 0.48, 0.71, 0.94, 1.18, 1.97, 2.95, and 5.9 μg per ml, respectively. The protein-to-DNA ratios varied from 0.73 in lane 2 to 36 in lane 10. The C+T and G+A chemical markers are also shown. An equivalent enzyme storage buffer was utilized instead of the enzyme to maintain identical assay conditions in each sample.

AMV IN to recognize and nick these LTR substrates in vitro.

AMV IN nicks the minus strand on the L-LTR DNA terminus two to three times more efficiently than the plus strand on the R-LTR terminus (Fig. 3). The short inverted repeat sequences on the 3′-OH strands of both LTR termini contain three nucleotide differences (Fig. 1), suggesting that these changes may play a role in controlling IN nicking properties. Modifications of the 5′ region of the short inverted repeat contained in the R-LTR double-stranded oligonucleotide (14) did affect the rate at which IN nicked at the CA sequence on the plus strand. There was no apparent preference for nicking of the L- or R-LTR oligonucleotide substrates by IN with either Mg^{2+} or Mn^{2+} (14). In our study, the ability of AMV IN to specifically nick the L-LTR DNA substrate at the in vivo site of integration in the presence of Mn^{2+} as well as Mg^{2+} suggests that these L-LTR sequences and R-LTR sequences are recognized differently by IN (Fig. 2). It is possible that the preference of IN for the L-LTR substrate is due to the recognition of LTR sequences by IN outside the short inverted repeat or the length of the DNA substrate itself.

AMV IN was not capable of releasing nucleotides from the blunt-ended *Eco*RV fragment (Fig. 4), from a blunt-ended *Eco*RI restriction site labeled with [^{32}P]TTP and [^{32}P]dATP (data not shown) (15), or from double-stranded oligonucleotide substrates specific for the human immunodeficiency virus LTR DNA termini (14). However, AMV IN nicked the *Ssp*I fragment at either 2 or 5 nucleotides in from the 3′-OH terminus (Fig. 4). Examination of the sequence 5′ terminal to the *Ssp*I restriction site revealed that there were 8 bp which are identical to the U5 LTR terminus, except for the substitution of an A for a T (Fig. 4, underlined sequence). The fortuitous 8-bp short inverted repeat sequence on the *Ssp*I fragment apparently promotes nicking by IN at one or both sites. That the majority of nicking by IN was not directly 3′ to the CA sequence of the *Ssp*I fragment raises the question of what sequence information determines the site of cleavage. In vivo experiments (21) indicate that the CA moiety and internal LTR sequences play a dominant role, but distance from the terminus may also contribute information, as indicated by the *Ssp*I fragment results.

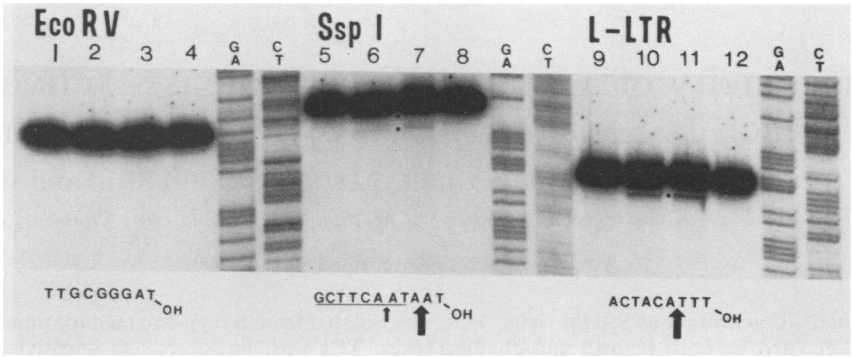

FIG. 4. Digestion of non-LTR blunt-ended DNA substrates by AMV IN. AMV IN (6 μg/ml) was incubated with an *Eco*RV DNA fragment (6 ng, 185 bp) (lanes 1 to 4), an *Ssp*I fragment (193 bp) (lanes 5 to 8), or the L-LTR terminal fragment (8 ng, 181 bp) (lanes 9 to 12). Lanes 1, 4, 5, 8, 9, and 12, control DNA without IN; lanes 2, 6, and 10, DNA with 20 mM Mg^{2+}; lanes 3, 7, and 11, DNA with 1 mM Mn^{2+}. High-salt conditions were employed in all lanes. Dots between lanes 6 and 7 and lanes 10 and 11 identify relevant cleaved fragments. Terminal sequences and cleavage sites (arrows) are indicated below. Incubation was for 1 h for Mg^{2+} and 15 min for Mn^{2+}. The first G+A and C+T markers were derived from the *Eco*RV-labeled strand, and the next two G+A and C+T chemical markers were different sets for the L-LTR DNA fragment of pAV-*Dra*I. The first C+T lane of the L-LTR fragment was displaced upward by 1 bp because of unremoved salt.

We thank Charlene Hall for typing the manuscript.

This work was supported in part by Public Health Service grant CA16312 from the National Cancer Institute.

LITERATURE CITED

1. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469–478.
2. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA *in vitro*. 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.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
5. **Colicelli, J., and S. P. Goff.** 1988. Isolation of an integrated provirus of Moloney murine leukemia virus with long terminal repeats in inverted orientation: integration utilizing two U3 sequences. J. Virol. **62**:633–636.
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. **Eisenman, R. N., W. S. Mason, and M. Linial.** 1980. Synthesis and processing of polymerase proteins of wild-type and mutant avian retroviruses. J. Virol. **36**:62–78.
8. **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.
9. **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.
10. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
11. **Grandgenett, D. P., and A. C. Vora.** 1985. Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. Nucleic Acids Res. **13**:6205–6221.
12. **Grandgenett, D. P., A. C. Vora, and R. D. Schiff.** 1978. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. Virology **89**:119–132.
13. **Grandgenett, D. P., A. C. Vora, R. Swanstrom, and J. C. Olsen.** 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. J. Virol. **58**:970–974.
14. **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.
15. **Knaus, R. J., P. J. Hippenmeyer, T. K. Misra, D. P. Grandgenett, V. R. Muller, and W. M. Fitch.** 1984. Avian retrovirus pp32 DNA binding protein: preferential binding to the promoter region of long terminal repeat DNA. Biochemistry **23**:350–359.
16. **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.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **Panganiban, A. T., and H. M. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature (London) **306**:155–160.
19. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.
20. **Quinn, T. P., and D. P. Grandgenett.** 1988. Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. **62**:2307–2312.
21. **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.
22. **Schwartzberg, P. J., 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.
23. **Terry, R., D. A. Soltis, M. Katzman, D. Cobrinik, J. Leis, and A. M. Skalka.** 1988. Properties of avian sarcoma-leukosis virus pp32-related *pol*-endonucleases produced in *Escherichia coli*. J. Virol. **62**:2358–2365.