JOURNAL OF VIROLOGY, Dec. 1989, p. 5319–5327
0022-538X/89/125319-09$02.00/0

Vol. 63, No. 12

The Avian Retroviral Integration Protein Cleaves the Terminal Sequences of Linear Viral DNA at the In Vivo Sites of Integration

MICHAEL KATZMAN,[1,2]† RICHARD A. KATZ,[3] ANNA MARIE SKALKA,[3] AND JONATHAN LEIS[1]*

Departments of Biochemistry[1] and Medicine,[2] Case Western Reserve University, Cleveland, Ohio 44106, and Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111[3]

Received 13 April 1989/Accepted 8 September 1989

The purified integration protein (IN) of avian myeloblastosis virus is shown to nick double-stranded oligodeoxynucleotide substrates that mimic the ends of the linear form of viral DNA. In the presence of Mg^{2+}, nicks are created 2 nucleotides from the 3′ OH ends of both the U5 plus strand and the U3 minus strand. Similar cleavage is observed in the presence of Mn^{2+} but only when the extent of the reaction is limited. Neither the complementary strands nor sequences representing the termini of human immunodeficiency virus type 1 DNA were cleaved at analogous positions. Analysis of a series of substrates containing U5 base substitutions has defined the sequence requirements for site-selective nicking; nucleotides near the cleavage site are most critical for activity. The minimum substrate size required to demonstrate significant activity corresponds to the nearly perfect 15-base terminal inverted repeat. This in vitro activity of IN thus produces viral DNA ends that are joined to host DNA in vivo and corresponds to an expected early step in the integrative recombination reaction. These results provide the first enzymatic support using purified retroviral proteins for a linear DNA precursor to the integrated provirus.

Genetic evidence suggests that an activity of a virus-encoded protein is necessary for integration of a retroviral DNA intermediate into host cellular DNA. Mutations introduced into the 3′ portions of the *pol* genes of murine leukemia virus, spleen necrosis virus, and Rous sarcoma virus produce replication-defective viruses that retain reverse transcriptase (RT) activity but are unable to integrate viral DNA (4, 5, 25, 27, 32). The protein encoded by this region has been named IN, for its integration function (20).

Upon infection of cells, at least three species of double-stranded viral DNA transiently appear in the nucleus before the integration event occurs (17, 33). These include a linear form, flanked by long terminal repeats (LTRs) derived from the noncoding ends of the viral RNA, and two covalently closed circular forms, containing either one LTR or two tandem LTRs. Although recent evidence from Brown et al. (2) and Fujiwara and Mizuuchi (8), using murine leukemia virus, has shown that the outer ends of the LTRs in the major viral DNA substrate for integrative recombination are not covalently joined to one another, earlier results of Panganiban and Temin (26) with spleen necrosis virus had suggested that covalently closed molecules could serve as precursors for integration. After integration, the provirus is flanked by copies of the LTRs (18, 29) which are intact except for the loss of 2 base pairs from the outer ends at the points of attachment to cellular DNA (16, 37). This finding suggests a role for a nuclease capable of either removing 2 nucleotides (or base pairs) from each end of the linear viral DNA form or, alternatively, of making a 4-base-pair staggered cut across the LTR-LTR junction of the two LTR-containing circular forms (2, 8).

The *pol* gene of the avian sarcoma-leukosis viruses (ASLVs) encodes the 95-kilodalton β chain of RT. Proteolytic processing of the β chain to produce the shorter 63-kilodalton α chain, in vivo and in vitro, releases from the carboxyl end a 32-kilodalton phosphorylated polypeptide, pp32 (1, 9, 12, 28, 30, 31), which is encoded by the genetically defined integration locus. Both RT, which exists as an αβ heterodimer, and IN (pp32) have been purified to homogeneity (14, 22, 36), and each has in vitro DNA endonuclease activity (10, 11, 14, 21). Using a primer extension runoff assay to map cleavage sites, RT and IN have been shown to nick supercoiled DNA substrates that contain covalently linked tandem LTR sequences near the LTR-LTR junction (3, 7, 36). In the presence of the divalent cation Mn^{2+}, RT nicks each strand 3 nucleotides 5′ to the junction, whereas IN nicks each strand at sites 2 or 3 nucleotides 5′ to the junction. In contrast, in the presence of Mg^{2+}, RT has no detectable endonuclease activity, and IN shows less activity but apparently greater selectivity, nicking each strand precisely 2 nucleotides 5′ to the LTR-LTR junction (13, 15, 36). The latter sites correspond to those which are joined to host DNA upon integration. These observations have supported models in which the circular DNA intermediate could function as an immediate precursor to the integrated provirus, although participation of linear molecules was not excluded (35).

In this report, we show that the ASLV endonucleases have analogous and seemingly more efficient activities on oligodeoxynucleotide substrates that mimic the termini of the linear viral DNA intermediate. Thus, covalently closed circular molecules do not appear to be obligatory intermediates in the integration reaction. The use of oligodeoxynucleotides containing base substitutions has defined the limits of length and sequence required of active substrates for site-selective cleavage by IN.

MATERIALS AND METHODS

Reagents. [α-^{32}P]TTP (800 Ci/mmol) and [γ-^{32}P]ATP (6,000 Ci/mmol) were purchased from New England Nuclear Research Products, Boston, Mass. Snake venom (*Bothrops atrox*; no. V-5375) was obtained from Sigma Chemical Co.,

* Corresponding author.
† Present address: Department of Medicine, The Milton S. Hershey Medical Center, Pennsylvania State University, Hershey, PA 17033.

St. Louis, Mo. T4 polynucleotide kinase and *Asn*I were from Boehringer Mannheim Biochemicals, Indianapolis, Ind.

Viral enzymes. The αβ forms of avian myeloblastosis virus RT and IN (pp32) were purified as previously described (21, 36). RT was stored at −70°C at concentrations ranging from 7 to 19 RT units per μl. IN was stored at −70°C in 40% glycerol–0.4 to 0.6 M NaCl–33 mM Tris hydrochloride (pH 7.6)–1 mM dithioerythritol–0.1 mM EDTA–0.1% Triton X-100–2 mM β-glycerophosphate. The concentration of IN was 30 ng/μl (1 pmol/μl) as estimated by Coomassie brilliant blue staining of the pp32 doublet seen on a polyacrylamide-sodium dodecyl sulfate gel.

Oligodeoxynucleotides. Oligodeoxynucleotides used in these studies are shown in Fig. 1 and were prepared as previously described (7). Appropriate sequences were 5′ end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and gel purified as described previously (7). The specific activities of the radiolabeled oligodeoxynucleotides, as ascertained by the DE81 filter binding method (23), were 10^6 to 10^7 cpm/pmol.

Preparation of oligodeoxynucleotide ladders. Markers for gel analysis were produced by the phosphodiesterase activity of snake venom on 5′ ^{32}P-labeled oligodeoxynucleotides. Substrates (0.2 to 1.0 pmol) were incubated with snake venom (final concentration of 0.1 to 50 μg/ml) in 50 mM Tris hydrochloride (pH 9.0)–10 mM $MgCl_2$ for 30 min at 37°C. Reactions were stopped by the addition of EDTA to 25 mM and loading buffer (95% formamide, 50 mM Tris hydrochloride [pH 8.4], 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol), followed by heating the mixture at 95°C for 5 min. The optimal sequence-specific nucleotide ladders were selected after gel analysis and autoradiography.

Annealing conditions. For reactions requiring double-stranded DNA substrates, 1 pmol of oligodeoxynucleotide (diluted to 10^5 cpm/pmol) was mixed with 3 pmol of unlabeled complementary oligodeoxynucleotide in a total volume of 6 μl of water in a 0.5-ml polypropylene microcentrifuge tube. Tubes were heated to 95°C for 5 min, spun in a tabletop centrifuge for 5 s, incubated at 37°C for 30 min, spun again, and then placed in ice. Annealing was confirmed by the change in migration of the labeled oligomers as analyzed by electrophoresis in nondenaturing polyacrylamide gels containing 190 mM glycine–25 mM Tris base. A threefold excess of the unlabeled complementary strand resulted in >95% duplex formation (data not shown). The 5′ ^{32}P-labeled oligodeoxynucleotides of length ≥14 bases were found to be double stranded by the above criteria (data not shown). A duplex containing U3[−]12, which could not be demonstrated by using these gel conditions, was indicated by the appearance of susceptibility to Mn^{2+}-dependent digestion by IN (discussed below; see Fig. 4).

Endonuclease reaction conditions. Standard 10-μl reaction mixtures contained 1 pmol of single-stranded or double-stranded oligodeoxynucleotide (specific activity, 10^5 cpm/pmol)–25 mM Tris hydrochloride (pH 8.3)–10 mM 2-mercaptoethanol–3 mM $MgCl_2$ or 2 mM $MnCl_2$–7 units of RT or 1 pmol of IN or 1 μl of the IN storage buffer. Reaction mixtures were incubated at 37°C for 90 min, and the reactions were stopped by the addition of 10 μl of loading buffer and heating to 95°C for 5 min. Terminated reaction mixtures were kept on ice or stored at −20°C. For gel analysis, aliquots were heated and loaded onto a 20% polyacrylamide–7 M urea denaturing gel (30 cm wide by 40 cm long by 0.4 mm thick), followed by electrophoresis at 1,800 V until the bromophenol blue dye had migrated 23 cm. Gels were autoradiographed with Kodak X-OMAT film for 12 to 24 h at −70°C, using a Dupont Cronex Lightning-Plus screen.

RESULTS

Endonuclease activity on double-stranded oligodeoxynucleotide substrates. The retroviral LTR consists of a region (U3) derived from the 3′ end of the RNA genome, a direct repeat (R) which exists at each end of the genome, and a region (U5) derived from the 5′ end of the RNA, in the order U3-R-U5. In the linear viral DNA intermediate, the terminus

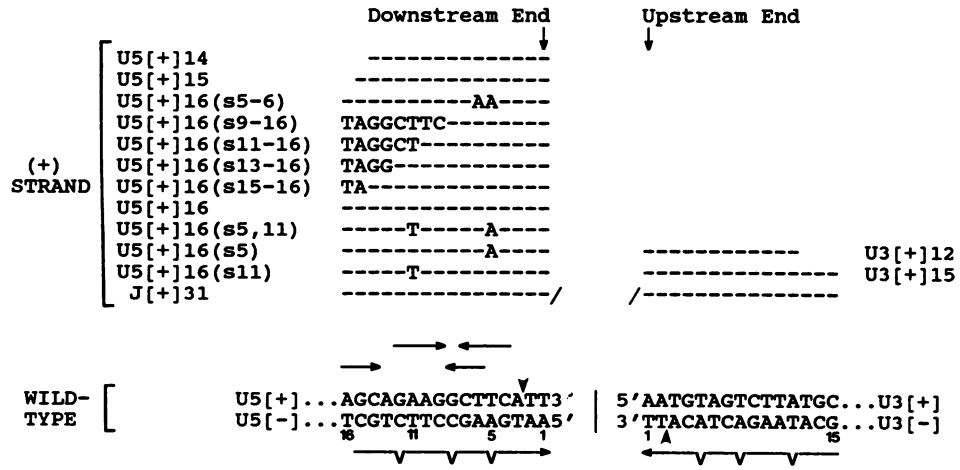

FIG. 1. Synthetic oligodeoxynucleotides. The wild-type ASLV DNA terminal sequences are shown at the bottom. U5[+] and U3[+] indicate the downstream or upstream end of the viral DNA plus strands, respectively. The vertical line marks the ends of the linear viral DNA which have been brought into juxtaposition; this would be the site of the LTR-LTR junction if these ends were covalently joined. Oligodeoxynucleotides derived from the plus strand are shown above the wild-type sequences, with homology indicated by dashes. The corresponding complementary minus-strand oligodeoxynucleotides are not shown. The nearly perfect 12-of-15-nucleotide inverted repeat at the viral DNA ends, as well as two inverted repeats in the U5 sequence, are indicated with horizontal arrows. Arrowheads denote the in vivo integration sites. Terminal oligodeoxynucleotides were named by the viral ends and DNA strands from which they derived, as well as by their lengths. Base substitutions (s) at specific nucleotide positions numbered from the viral termini are also indicated. J[+]31 and J[−]31 are 31-mers derived from the LTR-LTR junction of the plus or minus strand, respectively, and are continuous across the // symbols.

which is upstream (with respect to the genetic map) is formed by U3 sequences and the terminus which is downstream is formed by U5 sequences. These ends are shown adjacent to one another in Fig. 1, as they would be in a circular two-LTR molecule. In order to test the activity of the retroviral endonucleases on linear viral DNA termini, we used small linear double-stranded oligodeoxynucleotides which represent either of the terminal sequences as substrates for the viral proteins. One substrate (Fig. 1, bottom left) consisted of duplex 16-mers representing the U5 terminus of the downstream LTR in which only the 5′ end of the plus-strand oligodeoxynucleotide (U5[+]16) was labeled with ^{32}P. A second substrate (Fig. 1, bottom right) consisted of duplex 15-mers representing the U3 terminus of the upstream LTR with the radiolabel on the 5′ end of the minus-strand oligodeoxynucleotide (U3[−]15). Cleavage of these substrates at the 3′ ends of the labeled strands, as expected for the integration reaction, should produce labeled products shortened by 2 nucleotides (Fig. 1, bottom, arrowheads), which can be detected by autoradiography after separation of products by gel electrophoresis under denaturing conditions (7).

The wild-type duplex substrates were incubated with purified RT or IN in the presence of the divalent cations Mg^{2+} or Mn^{2+} and analyzed as above. RT was found to be inactive on either duplex substrate in the presence of Mg^{2+} (Fig. 2A and B, lanes b) but had nicking activity when provided with Mn^{2+} (Fig. 2A and B, lanes e). With the labeled U5 duplex as substrate, the predominant new band observed was a 13-mer, reflecting a preferential nick 3 nucleotides from the 3′ end of the plus strand (Fig. 2A, lane e, small arrowhead). An analogous preferred site was seen with the labeled U3 duplex (Fig. 2B, lane e, small arrowhead). These sites of cleavage, 1 nucleotide 5′ to those expected for the integration reaction, are identical to the sites observed earlier by using supercoiled substrates with covalently joined termini and a primer extension assay (3, 7, 36).

When IN was assayed with these wild-type substrates, using Mn^{2+} as the divalent cation, greater than 50% of the starting material (as estimated by densitometric tracings of the autoradiograms) was converted to shorter oligodeoxynucleotides. Under these conditions, the lengths of the labeled products revealed preferential cleavage at sites 2 or 3 nucleotides from each 3′ end (Fig. 2A and B, lanes f, large and small arrowheads) as well as cleavages at other sites. In contrast, in the presence of Mg^{2+} only 10% of the substrate was cleaved by the same amount of IN, and the major

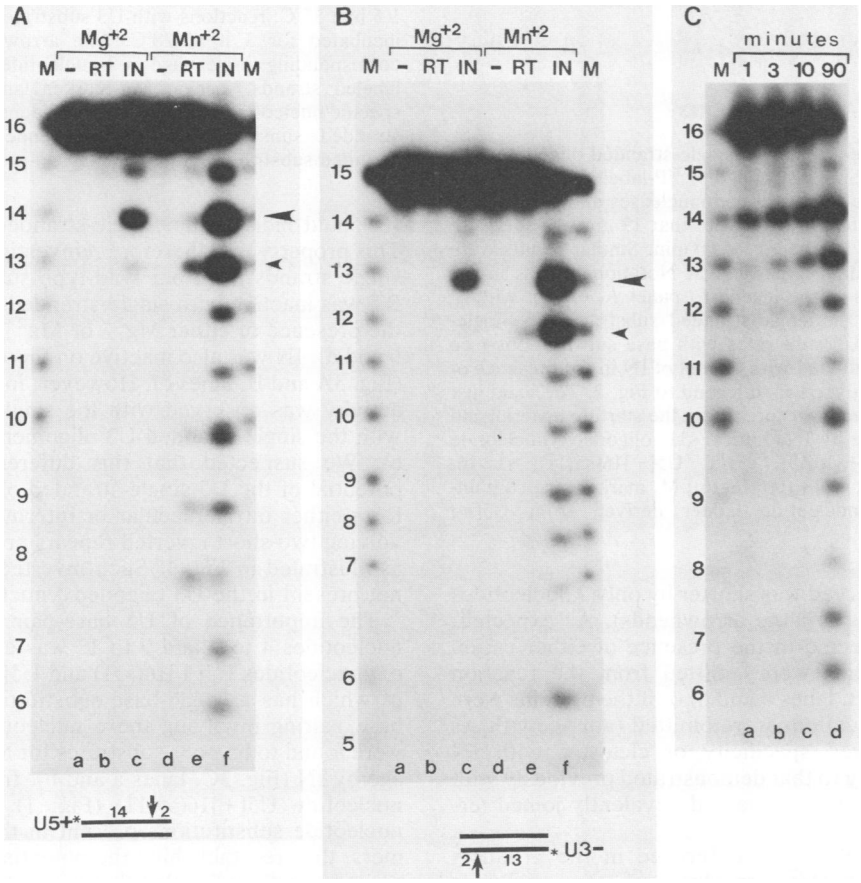

FIG. 2. Endonuclease activity on double-stranded oligodeoxynucleotide substrates. (A) The 5′ ^{32}P-labeled U5[+]16 was annealed with its unlabeled complementary sequence and incubated without enzyme (− lanes), with 7 units of RT, or with 1 pmol of IN in the presence of Mg^{2+} or Mn^{2+} and analyzed as described in Materials and Methods. As shown schematically at the bottom, nicking (arrow) 2 nucleotides from the 3′ end of the 5′ labeled strand (*) yields a product 2 nucleotides shorter than the starting material (large arrowhead). A product shortened by 3 nucleotides is indicated by the small arrowhead. M, Marker lane containing sequence-specific nucleotide ladder. (B) As in the legend to panel A, except with 5′ ^{32}P-labeled U3[−]15 as the duplex substrate. (C) The double-stranded ^{32}P-labeled U5 substrate was incubated with 1 pmol of IN in the presence of Mn^{2+}, as described in the legend to panel A, lane f. After 1, 3, 10, and 90 min of incubation, aliquots were removed, added to loading buffer, heated to stop the reaction, and analyzed as above.

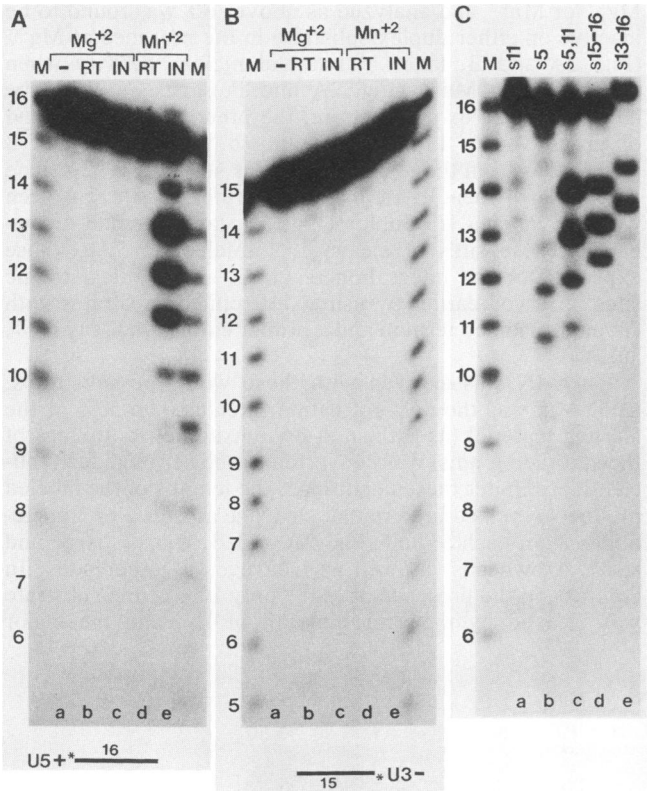

FIG. 3. Endonuclease activity on single-stranded oligodeoxynucleotide substrates. (A) Single-stranded 5′ ^{32}P-labeled U5[+]16 was used as substrate for the retroviral endonucleases and analyzed as described in the legend to Fig. 2, except that 19 units of RT were used and reactions were incubated for 60 min. Similar results were obtained in six independent experiments. Notations are as in the legend to Fig. 2. (B) As in the legend to panel A, except with 5′ ^{32}P-labeled U3[−]15 as the single-stranded substrate. (C) Single-stranded 5′ ^{32}P-labeled U5 substrates with base substitutions (see Fig. 1 and text) were incubated with 1 pmol of IN in the presence of Mn^{2+} for 90 min and analyzed as in legend to Fig. 2. The band just beneath the 16-mer in lane b was present in the starting material and does not reflect cleavage by IN. Lanes: s11, oligodeoxynucleotide U5[+]16(s11); s5, U5[+]16(s5); s5,11, U5[+]16(s5,11); s15-16, U5[+]16(s15-16); s13-16, U5[+]16(s13-16); M, marker lane containing sequence-specific nucleotide ladder derived from U5[+]16(s5,11).

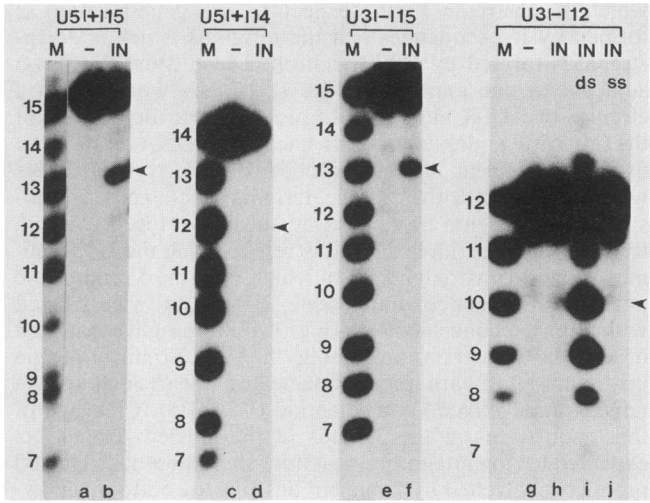

FIG. 4. Endonuclease activity on duplex substrates of various lengths. The indicated oligodeoxynucleotides were used as double-stranded (all lanes except lane j) or single-stranded (lane j) substrates and incubated without enzyme (− lanes) or with 1 pmol of IN in the presence of Mg^{2+} (lanes a through h) or Mn^{2+} (lanes i and j). Reactions with U5 substrates (lanes a through d) were incubated for 1.5 h at 37°C; reactions with U3 substrates (lanes e through j) were incubated for 3 h at 28°C. The arrowheads indicate positions corresponding to cleavage of 2 nucleotides from the 3′ ends of the labeled strands. Lanes: M, marker lanes containing sequence-specific nucleotide ladders; ss, Mn^{2+}-dependent reaction on single-stranded substrate; ds, Mn^{2+}-dependent reaction on double-stranded substrate.

cleavage product observed was shorter by only 2 nucleotides (Fig. 2A and B, lanes c, large arrowheads). As expected, cleavage was not detected in the presence of either cation when the viral proteins were omitted from the reaction mixture (Fig. 2A and B, lanes a and d) or if the proteins were included but divalent cations were omitted (not shown). As with RT, the observed specificity of cleavage with IN corresponded precisely to that demonstrated previously with supercoiled substrates that contained covalently joined terminal sequences.

We questioned whether the difference in the products observed for IN using Mg^{2+} or Mn^{2+} reflects a different specificity or was due to the greater extent of cleavage evident with Mn^{2+}. When Mn^{2+}-dependent digestion was limited to that observed typically with Mg^{2+}, the cleavage patterns were indistinguishable (compare Fig. 2A, lane c, with Fig. 2C, lane a).

Endonuclease activity on single-stranded oligodeoxynucleotide substrates. Earlier studies had shown that both RT and IN could nick certain single-stranded DNA substrates (7). This property was therefore reinvestigated, using the labeled single strands from our wild-type substrates. As expected, RT was inactive with single-stranded U5 or U3 oligomers in the presence of either Mg^{2+} or Mn^{2+} (Fig. 3A and B, lanes b and d). IN was also inactive on these substrates with Mg^{2+} (Fig. 3A and B, lanes c). However, in the presence of Mn^{2+}, activity was observed with the single-stranded U5 but not with the single-stranded U3 oligomer (Fig. 3A and B, lanes e). We suspected that this difference might reflect the potential of the U5 single-stranded oligodeoxynucleotide to form either intramolecular or intermolecular base pairs involving two short inverted repeats present in this sequence, as illustrated in Fig. 1. Such inverted repeat sequences are not present in the U3 oligodeoxynucleotide.

The importance of U5 base-pair interactions involving nucleotides 4 to 7 and 9 to 12 was evaluated with oligodeoxynucleotides U5[+]16(s11) and U5[+]16(s5) (Fig. 1), each of which has a single-base substitution that would disrupt base pairing involving these nucleotides. These oligomers were found to be poor substrates for Mn^{2+}-dependent cleavage by IN (Fig. 3C, lanes a and b). In contrast, oligodeoxynucleotide U5[+]16(s5,11) (Fig. 1), which contains both nucleotide substitutions present in the previous two oligomers, thus reestablishing the potential for base pairing, was a good substrate for cleavage by IN (Fig. 3C, lane c). Thus, interactions involving these nucleotides appear to be required for the observed susceptibility of the wild-type single-stranded U5 substrate to Mn^{2+}-dependent cleavage by IN.

Analysis of the susceptibility of other oligodeoxynucleotides suggested that interactions involving nucleotides 6 to 8 and 14 to 16 were not important. For instance, oligodeoxynucleotides U5[+]16(s15-16) and U5[+]16(s13-16),

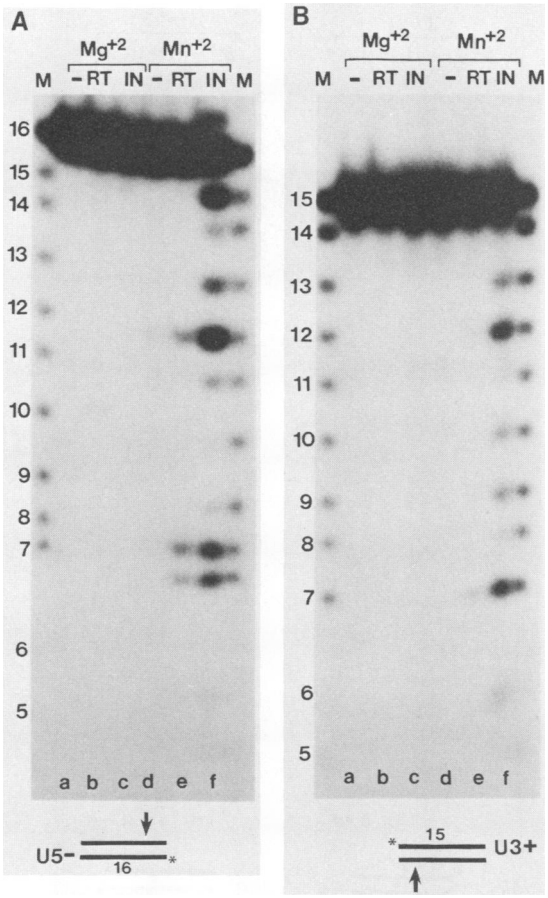

FIG. 5. Lack of specific endonuclease activity on complementary strands. (A) The 5′ ^{32}P-labeled U5[−]16 was annealed with its unlabeled complementary sequence and analyzed as in the legend to Fig. 2. The arrow in the schematic at the bottom indicates the position of a nick on the unlabeled strand expected from the Mg^{2+}-dependent activity of IN (as in the legend to Fig. 2). No bands were evident in reaction lanes below the positions of 5-mers. (B) As in the legend to panel A, except with 5′ ^{32}P-labeled U3[+]15 as the duplex substrate.

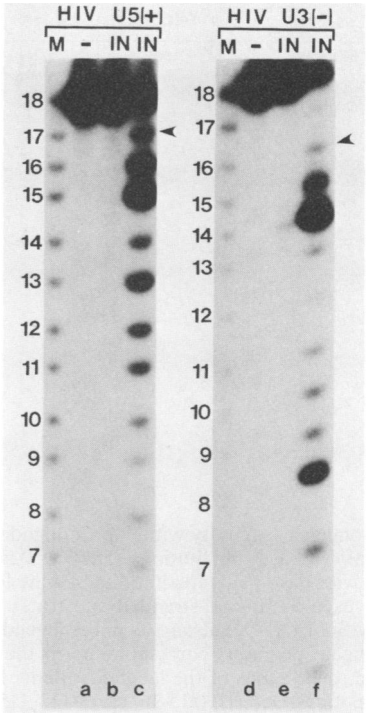

FIG. 6. Lack of specific endonuclease activity by ASLV IN on HIV-1 oligomer substrates. The 5′ ^{32}P-labeled oligodeoxynucleotides, representing the plus strand at the U5 end (HIV U5[+], 5′GTGGAAAATCTCTAGCAG3′) or the minus strand at the U3 end (HIV U3[−], 5′TGAATTAGCCCTTCCAGT3′) of the LTRs of HIV-1, were annealed to their unlabeled complementary sequences and tested for susceptibility to the ASLV IN. Duplex substrates were incubated without enzyme (− lanes) or with 1.5 pmol of the ASLV IN in the presence of Mn^{2+} (lanes a, c, d, and f) or Mg^{2+} (lanes b and e), as described in Materials and Methods. Arrowheads indicate the presumed in vivo cleavage sites that would yield the integrated HIV provirus (34). M, Marker lanes containing sequence-specific nucleotide ladders.

which have base substitutions that selectively interfere with such interactions but leave the other inverted repeat intact (Fig. 1), were also found to be good substrates for cleavage by IN (Fig. 3C, lanes d and e). Note that the cleavage products observed in this gel migrated differently in each lane because of sequence differences of the substrates. Taken together, these results indicate that IN requires that a substrate contain a duplex structure although the cleavage site itself need not be base paired.

What is the minimum length required of duplex substrates for site-selective Mg^{2+}-dependent cleavage? To investigate this question, we prepared double-stranded U5 terminal oligodeoxynucleotides 14 or 15 base pairs in length (Fig. 1). The 15-mer duplex, but not the 14-mer duplex, was susceptible to IN in the presence of Mg^{2+} at 37°C (Fig. 4, lanes b and d). This incubation temperature is below the estimated melting temperature of 40°C for the duplex 14-mers (24), and analysis by nondenaturing gel electrophoresis indicated that the 14-mer duplex was present under these conditions (data not shown).

We also prepared and tested 12-mers that represented the U3 terminus (Fig. 1). Because the estimated melting temperature for this sequence is only 30°C (24), reactions were conducted at 28°C. At this temperature, U3 duplex 15-mers were readily nicked by IN in the presence of Mg^{2+}, whereas minimal or no nicking was repeatedly observed with the 12-mers (Fig. 4, lanes f and h). This was not due to a problem with annealing the 12-mers, as indicated by the appearance of susceptibility to Mn^{2+}-dependent nicking by IN under these conditions (Fig. 4, compare lanes i and j). We thus conclude that the minimal length of a duplex substrate for the site-selective Mg^{2+}-dependent activity of IN is 15 base pairs from the U5 terminus or 13 to 15 base pairs from the U3 terminus.

Sequence specificity of the endonuclease reaction. Our assay measures nicking of one particular strand in each duplex. To verify the specificity of this reaction, the 5′ ends of the U5[−] and U3[+] strands in the wild-type duplexes were ^{32}P-labeled (shown schematically in Fig. 5). When these substrates were incubated with either RT or IN in the presence of Mg^{2+}, no cleavage (14-mers or 13-mers, respectively) was detected on the labeled strands (Fig. 5, lanes b and c). Thus, the presence of duplex DNA ends is not sufficient for the cleavage reaction. In addition, because labeled dinucleotides were not detected on the lower portions of the gels (no bands were seen in the reaction lanes below those shown in Fig. 5), the products demonstrated in

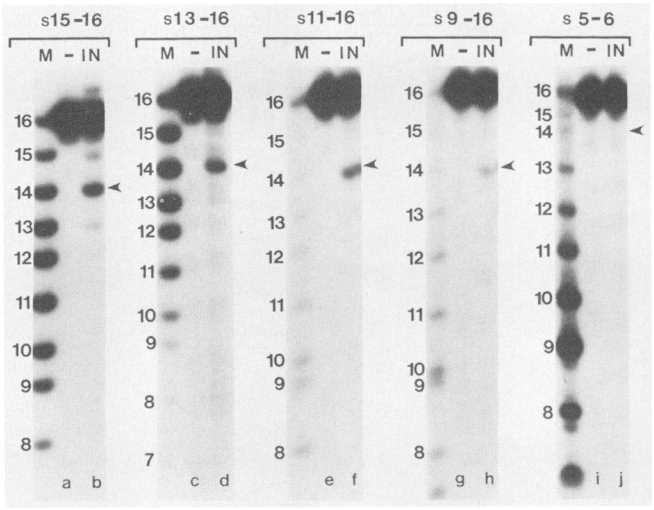

FIG. 7. Endonuclease activity with duplex oligodeoxynucleotide substrates containing base substitutions. The 5′ ^{32}P-labeled oligodeoxynucleotides with base-pair substitutions (see the legend to Fig. 1 and text) were used as double-stranded substrates for the Mg^{2+}-dependent activity of IN. Details are as in the legend to Fig. 2. The arrowheads indicate positions corresponding to the cleavage of 2 nucleotides from the 3′ ends of the labeled strands. Panels: s15-16, oligodeoxynucleotide U5[+]16(s15-16); s13-15, U5[+]16(s13-16); s11-16, U5[+]16(s11-16); s9-16, U5[+]16(s9-16); s5-6, U5[+]16(s5-6).

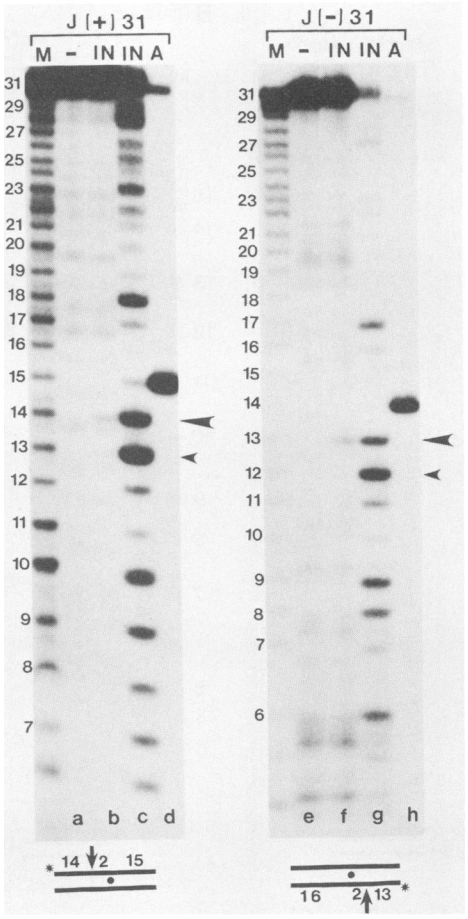

FIG. 8. Endonuclease activity on linear duplex junction sequences. Double-stranded 5′ ^{32}P-labeled oligodeoxynucleotides J[+]31 (lanes a through d) or J[−]31 (lanes e through h) were incubated without enzyme (− lanes), with 1.5 pmol of IN, or with 8 units of the restriction enzyme *Ase*I (A) in the presence of Mg^{2+} (lanes a, b, d, e, f, and h) or Mn^{2+} (lanes c and g), as described in Materials and Methods. Lanes d and h were loaded with only 10% as many counts per minute of labeled DNA as were the other lanes. As represented schematically at the bottom, nicking (arrows) at sites 2 nucleotides 5′ to the LTR-LTR junction (dots) produces 5′ labeled (*) 14-mers from J[+]31 or 13-mers from J[−]31 (large arrowheads). Nicking at sites 3 nucleotides 5′ to the junction yields 5′ labeled 13-mers or 12-mers, respectively (small arrowheads). M, Marker lanes containing sequence-specific nucleotide ladders (positions 29, 30, and 31 on the autoradiograms were unambiguously assigned by using a shorter exposure than that shown).

Fig. 2 must represent site-selective nicks that are restricted to one strand of duplex substrates (see diagrams at the bottom of Fig. 5).

In contrast to the results with Mg^{2+}, IN did cut the complementary strands of the duplex U5 and U3 substrates in the presence of Mn^{2+} (Fig. 5, lanes f). However, this activity was relatively nonselective. Nonselective cleavage of these substrates was also displayed by RT in the presence of Mn^{2+} (most evident in Fig. 5A, lane e).

In a separate series of experiments, we tested the ability of the ASLV IN to cleave double-stranded 18-mers representing the deduced U5 and U3 ends (34) of the LTRs of the human immunodeficiency virus type 1 (HIV-1). The ASLV IN had little or no activity on these substrates in the presence of Mg^{2+} (Fig. 6, lanes b and e). However, significant activity was observed with Mn^{2+} (Fig. 6, lanes c and f). Because integrated HIV-1 provirus appears to be missing only 1 U5 but 2 U3 terminal base pairs (34), cleavage 1 nucleotide from the 3′ end of the U5 plus strand or 2 nucleotides from the 3′ end of the U3 minus strand would be expected for an integration-specific reaction. However, these sites (Fig. 6, arrowheads) were not selectively nicked. We conclude that with Mg^{2+} the ASLV IN requires authentic ASLV terminal sequences for site-selective cleavage, but with Mn^{2+} the reaction is much less specific.

Which nucleotides in the terminal sequences are required for site-selective cleavage? Some variability in substrate sequence must be tolerated by IN with Mg^{2+}. This was suggested by the ability of IN to cleave both the U5 and U3 ends which have 3 of 15 nucleotide mismatches (Fig. 1). To determine which nucleotides in the wild-type U5 substrate were most important, we made a series of base-pair substitutions that extended for increasing distances toward the cleavage site (Fig. 1). These substrates were labeled to equivalent specific activities and incubated with IN in the presence of Mg^{2+}. The results showed that base substitutions of more than 10 nucleotides upstream of the cleaved end were tolerated; such duplexes (s15-16, s13-16, and s11-16) were nicked at the expected sites. The extent of the reaction gradually decreased compared with that of the wild-type duplex (Fig. 7, lanes b, d, and f) as the extent of the base changes became greater. With the largest substitution (s9-16), markedly diminished susceptibility was observed (Fig. 7, lane h). Thus, nucleotides closer to the cleavage site are more critical for susceptibility to IN. This result was confirmed with the U5 duplex s5-6 which has only 2 base-pair substitutions (Fig. 1) but at positions just 2 nucleotides upstream of the cleavage site. This duplex was not nicked by IN (Fig. 7, lane j).

Endonuclease activity on oligodeoxynucleotide substrates with LTR-LTR junction sequences. Previous experiments

using a primer extension runoff protocol demonstrated that RT and IN preferentially nicked near the LTR-LTR junction in supercoiled substrates but at low efficiency (3, 7). A linear double-stranded oligodeoxynucleotide substrate which includes LTR-LTR junction sequences was therefore tested as a substrate in the current assay system. Oligodeoxynucleotides were constructed which included the 16 terminal U5 nucleotides joined to the 15 terminal U3 nucleotides. The resulting 31-mers represent the sequences of the plus strand (J[+]31) and the minus strand (J[−]31) at the LTR-LTR junctions present in the two LTR-containing circular viral DNA forms (Fig. 1). The duplex 31-mer junction sequence was shown to be accessible to endonucleolytic cleavage by use of the restriction enzyme *Ase*I which cleaves each strand 1 nucleotide 5′ to the LTR-LTR junction; this also provided an internal size marker (Fig. 8, lanes d and h). In the presence of Mg^{2+}, IN was able to nick each strand of the duplex 31-mers at the biologically relevant site 2 nucleotides 5′ to the LTR-LTR junction (Fig. 8, lanes b and f, large arrowheads). However, the extent of the reaction was reproducibly lower than that observed with oligodeoxynucleotide substrates that represented the U5 or U3 termini alone (Fig. 2), but it was comparable to the extent of the reaction previously observed with supercoiled substrates containing LTR-LTR junctions (7). Mixing experiments demonstrated that the 31-mers did not contain inhibitors of the Mg^{2+}-dependent activity of IN, and the duplex natures of the 31-mer substrates were confirmed by native gel electrophoresis (data not shown).

In the presence of Mn^{2+}, IN exhibited a high level of activity on these same substrates. Among the many cleavages observed, preferential nicking occurred at sites 2 or 3 nucleotides 5′ to the junction (Fig. 8, lanes c and g, large and small arrowheads), with greater nicking at the latter site. This site had previously been reported by Grandgenett and Vora (13), using larger linear duplex DNA substrates. With the duplex 31-mers, unlike the wild-type terminal U5 duplex, limiting the incubation time with Mn^{2+} did not increase the specificity for the site 2 nucleotides from the junction (data not shown).

DISCUSSION

The retroviral integration reaction is incompletely understood, due in large part to the absence of an in vitro system which can reconstitute this reaction with purified viral proteins. Currently the ASLVs provide the only retroviral system from which enzymatically active *pol* gene-associated endonucleases have been purified. This makes possible an in vitro biochemical approach to analysis of the cleavage reaction which is a presumed early step in the integration process.

Purified ASLV RT and IN have previously been reported to have in vitro site-selective endonuclease activity on supercoiled DNA substrates containing covalently linked tandem LTR sequences. We have now described comparable activity on short duplex oligodeoxynucleotide substrates that represent the ends of the linear viral DNA intermediate. The ASLV RT has endonuclease activity on such substrates only in the presence of Mn^{2+}, nicking the U5 plus and the U3 minus strands 3 nucleotides from the ends which represent the viral termini. IN also demonstrates strong Mn^{2+}-dependent endonuclease activity on these substrates, preferentially nicking each strand 2 or 3 nucleotides from the end, with relatively more of the former at early times in the reaction. In the presence of Mg^{2+}, however, IN preferentially nicks these substrates 2 nucleotides from the end, at sites which match the in vivo integration sites (Fig. 2). Although it is probable that other factors influence the integration reaction in vivo, these and previously reported results support the Mg^{2+}-dependent endonuclease of IN as the more biologically relevant activity with respect to cleavage of viral DNA (3, 6, 7, 13, 15, 36). However, during integration, both viral and host DNAs must be cleaved before joining. Analyses of host DNA target sites have revealed little sequence specificity with regard to this reaction. Thus, if the retroviral *pol* gene products are involved in cutting host DNA, one might expect to see both selective and nonselective endonuclease activities. It may be that the less-selective Mn^{2+}-dependent activity is a manifestation of the capacity for IN or RT to make the apparently sequence-independent cleavages of host DNA that occur during integration (34).

Since selective cleavage is seen in vitro both on substrates that mimic the LTR termini of linear viral DNA and on substrates that represent the LTR-LTR junction of circular viral DNA, our experiments can not exclude either DNA form as a possible precursor to ASLV integration. However, the data in this report show that covalent joining of the two viral DNA ends is not mandatory for site-selective cleavage and provide the first enzymatic support using purified retroviral proteins for a linear precursor to the integrated provirus. Whether IN recognizes the noncovalently joined ends independently of one another or in juxtaposition, thus mimicking the LTR-LTR junction, is currently unknown.

Monitoring the endonuclease activity of IN with small oligodeoxynucleotide substrates represents a significant improvement over the previously used primer extension runoff assay both in simplicity and in increased yield of product. This has made it possible to examine kinetic parameters of the cleavage reaction. Our preliminary results suggest that the K_m for Mg^{2+}-dependent cleavage by IN is approximately 10^{-6} M (data not shown), a value close to that for the utilization of DNA primer-template by the DNA-directed DNA polymerase activity of RT (19). In addition, the amount of product formed in 90 min was found to approximate the amount of enzyme present (data not shown), suggesting a stoichiometric relationship between IN and substrate. This possibility is currently being investigated.

Previously, site-selective endonuclease activity by RT was demonstrated on single-stranded tandem LTR junction sequences cloned into M13 DNA (6). Demonstration of the appropriate cleavage sites on these circular substrates required that 22 nucleotides of the plus strand or 44 nucleotides of the minus strand be present (7). Examination of these required sequences revealed that the limits defined two different inverted repeats that potentially could form base pairs (D. Cobrinik, Ph.D. thesis, Case Western Reserve University, Cleveland, Ohio, 1988). When unannealed oligodeoxynucleotides were used as substrates in our new assay, Mn^{2+}-dependent activity was also detected only with sequences that had the capacity to form base-paired structures. Mg^{2+}-dependent activity had an absolute requirement for double-stranded substrates. We thus conclude that substrate recognition and cleavage by both IN and RT require duplex DNA near the cleavage site, as is presumably present in vivo. This suggests that agents which could selectively denature the viral termini might potentially interfere with the cleavage reaction and block integration.

Detailed analysis of the site-selective Mg^{2+}-dependent reaction on duplex substrates revealed some of the features required for substrate susceptibility to IN. The minimal

length (Fig. 4) of substrates required to see activity corresponds to the length of the conserved terminal inverted repeat of the ASLVs, which is 15 nucleotides (Fig. 1). Although unrelated sequences of sufficient length were poor substrates (Fig. 5 and 6), some sequence variation in substrates was tolerated. For example, the three mismatches between the U5 and U3 termini, as well as base substitutions which were introduced into U5 more than 8 nucleotides from the biologically relevant cleavage site, did not significantly affect the efficiency of nicking (Fig. 7). Nucleotide substitutions extending closer to the cleavage site, however, diminished the amount of product formed. A 2-base-pair substitution just 2 nucleotides from the cleavage site completely inactivated the substrate (Fig. 7). The effect of this mutation on viral DNA integration in vivo is currently being investigated. These results indicate that sequences adjacent to the conserved 5'CATT3' found near the ends of most retroviral DNA (34) are required for specific recognition and cleavage, whereas sequences more remote from these ends are not. A comparison of LTR sequences of various strains of ASLV LTRs shows a conservation of nucleotides in the last 8 to 12 base pairs at the U3 end. This is consistent with the above U5 substrate mutation studies and supports the notion that the *cis*-activating sequences required for DNA integration include approximately 8 to 12 base pairs of the LTR termini.

The development of the in vitro assay system described in this report will make possible further definition of the sequence and substrate requirements for site-selective cleavage by retroviral endonucleases. This assay may also allow for the development and screening of a new class of antiviral agents capable of interfering with the endonuclease activity necessary for retroviral integration.

ACKNOWLEDGMENTS

The authors thank David Cobrinik for helpful suggestions and Stuart Le Grice and W. Bannwarth of Hoffmann-La Roche Co., Basal, Switzerland, for the generous gift of several oligodeoxynucleotides used in these studies.

This work was supported in part by Public Health Service grant CA 38046 (J.L.) from the National Cancer Institute; by Cancer Research Center grant P30 CA 43703; by Public Health Service grants CA 49042 (A.M.S.), CA 48703 (A.M.S.), CA 06927 (A.M.S.), and RR 05539 (A.M.S.) from the National Institutes of Health; by a grant from the Pew Charitable Trust; and also by an appropriation from the Commonwealth of Pennsylvania. M.K. was supported by a Physician's Research Training Fellowship (PRTF-77) from the American Cancer Society. R.A.K. was supported by the W.W. Smith Charitable Trust.

LITERATURE CITED

1. **Alexander, F., J. Leis, D. A. Soltis, R. M. Crowl, W. Danho, M. S. Poonian, Y.-C. E. Pan, and A. M. Skalka.** 1987. Proteolytic processing of avian sarcoma and leukosis viruses *pol-endo* recombinant proteins reveals another *pol* gene domain. J. Virol. **61**:534–542.
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. **Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis.** 1987. Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. J. Virol. **61**:1999–2008.
4. **Donehower, L. A.** 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3' region of *pol*. J. Virol. **62**:3958–3964.
5. **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.
6. **Duyk, G., J. Leis, M. Longiaru, and A. M. Skalka.** 1983. Selective cleavage in the avian retroviral long terminal repeat sequence by the endonuclease associated with the αβ form of avian reverse transcriptase. Proc. Natl. Acad. Sci. USA **80**: 6745–6749.
7. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. DeHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. J. Virol. **56**:589–599.
8. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
9. **Gibson, W., and I. M. Verma.** 1974. Studies on the reverse transcriptase of RNA tumor viruses. Structural relatedness of two subunits of avian RNA tumor viruses. Proc. Natl. Acad. Sci. USA **71**:4991–4994.
10. **Golomb, M., and D. P. Grandgenett.** 1979. Endonuclease activity of purified RNA-directed DNA polymerase from avian myeloblastosis virus. J. Biol. Chem. **254**:1606–1613.
11. **Golomb, M., D. P. Grandgenett, and W. Mason.** 1981. Virus-coded DNA endonuclease from avian retrovirus. J. Virol. **38**:548–555.
12. **Grandgenett, D. P., R. J. Knaus, and P. J. Hippenmeyer.** 1983. Antibodies against a synthetic peptide of the avian retrovirus pp32 protein and the β DNA polymerase subunit. Virology **130**:257–262.
13. **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.
14. **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.
15. **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.
16. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus ev1: integration creates a 6 bp repeat of host DNA. Cell **23**:155–164.
17. **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.
18. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, and H. E. Varmus.** 1978. Proviruses of avian sarcoma virus are terminally redundant, coextensive with unintegrated linear DNA and integrated at many sites. Cell **15**:1397–1410.
19. **Hurwitz, J., and J. P. Leis.** 1972. RNA-dependent DNA polymerase activity of RNA tumor viruses. I. Directing influence of DNA in the reaction. J. Virol. **9**:116–129.
20. **Leis, J., D. Baltimore, J. M. Bishop, J. Coffin, E. Fleissner, S. P. Goff, S. Oroszlan, H. Robinson, A. M. Skalka, H. M. Temin, and V. Vogt.** 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol. **62**:1808–1809.
21. **Leis, J., G. Duyk, S. Johnson, M. Longiaru, and A. M. Skalka.** 1983. Mechanism of action of the endonuclease associated with the αβ and ββ forms of avian RNA tumor virus reverse transcriptase. J. Virol. **45**:727–739.
22. **Leis, J. P.** 1976. RNA-dependent DNA polymerase activity of RNA tumor virus. VI. Processive mode of action of avian myeloblastosis virus polymerase. J. Virol. **19**:932–939.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 473. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Meinkoth, J., and G. Wahl.** 1984. Hybridization of nucleic acids immobilized on solid supports. Analyt. Biochem. **138**:267–284.
25. **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.

26. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
27. **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.
28. **Rho, H. M., D. P. Grandgenett, and M. Green.** 1975. Sequence relatedness between the subunits of avian myeloblastosis virus reverse transcriptase. J. Biol. Chem. **250**:5278–5280.
29. **Sabran, J. L., T. W. Hsu, C. Yeater, A. Kaji, W. S. Mason, and J. M. Taylor.** 1979. Analysis of integrated avian RNA tumor virus DNA in transformed chicken, duck, and quail fibroblasts. J. Virol. **29**:170–178.
30. **Schiff, R. D., and D. P. Grandgenett.** 1978. Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the β polypeptide of the avian retrovirus DNA polymerase. J. Virol. **28**:279–291.
31. **Schiff, R. D., and D. P. Grandgenett.** 1980. Partial phosphorylation *in vivo* of the avian retrovirus pp32 DNA endonuclease. J. Virol. **36**:889–893.
32. **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.
33. **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.
34. **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.
35. **Skalka, A. M., and J. Leis.** 1984. Retroviral DNA integration. Bioessays **1**:206–210.
36. **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.
37. **Varmus, H. E.** 1983. Retroviruses, p. 411–503. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.