JOURNAL OF VIROLOGY, May 1993, p. 2628–2636
0022-538X/93/052628-09$02.00/0

Vol. 67, No. 5

Comparison of DNA Binding and Integration Half-Site Selection by Avian Myeloblastosis Virus Integrase

DUANE P. GRANDGENETT,[1]* ROSS B. INMAN,[2] AJAYKUMAR C. VORA,[1] AND MICHAEL L. FITZGERALD[1]

Institute for Molecular Virology, St. Louis University Medical Center, 3681 Park Avenue, St. Louis, Missouri 63110,[1] and Institute for Molecular Virology, University of Wisconsin—Madison, Madison, Wisconsin 53706[2]

Received 7 October 1992/Accepted 2 February 1993

Insertion of the linear retrovirus DNA genome into the host DNA by the virus-encoded integrase (IN) is essential for efficient replication. We devised an efficient virus-like DNA plasmid integration assay which mimics the standard oligonucleotide assay for integration. It permitted us to study, by electron microscopy and sequence analysis, insertion of a single long terminal repeat terminus (LTR half-site) of one plasmid into another linearized plasmid. The reaction was catalyzed by purified avian myeloblastosis virus IN in the presence of Mg^{2+}. The recombinant molecules were easily visualized and quantitated by agarose gel electrophoresis. Agarose gel-purified recombinants could be genetically selected by transformation of ligated recombinants into *Escherichia coli* HB101 cells. Electron microscopy also permitted the identification and localization of IN-DNA complexes on the virus-like substrate in the absence of the joining reaction. Intramolecular and intermolecular DNA looping by IN was visualized. Although IN preferentially bound to AT-rich regions in the absence of the joining reaction, there was a bias towards GC-rich regions for the joining reaction. Alignment of 70 target site sequences 5′ of the LTR half-site insertions with 68 target sites previously identified for the concerted insertion of both LTR termini (LTR full-site reaction) indicated similar GC inflection patterns with both insertional events. Comparison of the data suggested that IN recognized only half of the target sequences necessary for integration with the LTR half-site reaction.

Efficient replication of retroviruses requires the integration of viral DNA into host DNA by the viral integrase (IN). Upon infection of cells and cytoplasmic synthesis of linear blunt-ended viral DNA by the reverse transcriptase in the virus particle, a 160S viral nucleoprotein complex is formed (1). IN is associated with the 160S complex (1, 7). Between the synthesis of linear blunt-ended DNA and the transport of the 160S nucleoprotein complex into the nucleus, IN specifically removes a dinucleotide from both DNA termini at their 3′ OH ends (11, 28), a prerequisite for integration. This complex is fully capable of integrating its trimmed viral DNA into exogenous DNA targets (2, 6, 23, 35).

Purified IN is also capable of 3′-end processing and strand transfer. The protein removes a dinucleotide from blunt-ended substrates which mimic the long terminal repeats (LTR) of viral DNA termini, exposing the conserved CA_{-OH} dinucleotide (1, 4, 8, 10, 19, 30, 37). Strand transfer, which requires a recessed 3′ OH LTR terminus, is mediated by IN with no exogenous energy source present (2–4, 6, 11, 18, 34). IN from various retroviruses produces a staggered cleavage, characteristic for that species, at the site of insertion on the host DNA (33). Recent evidence suggests that the joining of the recessed 3′ OH LTR terminus to the 5′ P end of target DNA at the site of cleavage occurs by a one-step transesterification mechanism (5, 36).

How does IN recognize its host target site for integration of viral DNA? Retroviruses seemingly integrate at random sites in the host DNA with no apparent sequence preference at the sites of insertion (33). However, studies have suggested that integration of retroviruses and retrotransposable elements is a nonrandom process (20, 29, 31). Concerted insertion of both recessed viral DNA termini (LTR full-site integration) by IN into lambda DNA demonstrated that integration occurs at multiple sites, but the limited number of units precluded analysis of sequence specificity (2, 18, 23). Recent studies using a similar lambda genetic assay employing purified avian myeloblastosis virus (AMV) IN followed by sequence analysis of the target site from 68 recombinants demonstrated the existence of "hot" and "cold" spots in the nonessential region of lambda-gtWES (9). Similarly, some target sequence trends were observed with the Moloney murine leukemia virus 160S integration complex by using naked DNA or minichromosomes as DNA targets (27). IN is also capable of inserting 3′ OH recessed double-stranded LTR oligonucleotides (LTR half-site integration) into other oligonucleotides (4, 18, 22, 30). This assay also indicates that not all of the potential target sites are equally utilized, but the small target substrate is restricted in the number of available sites.

By using purified AMV IN, it was shown that IN preferentially, although not exclusively, binds to certain regions on plasmid DNA, notably AT-rich regions and LTR DNA sequences contained in the same plasmid (14, 21, 25, 32). Preferential binding of IN to these sequences on plasmid DNA appears independent of supercoiling. In contrast, for efficient single-stranded cleavage of double-stranded DNA by AMV IN in the presence of Mg^{2+}, supercoiling of the DNA is necessary (13, 14).

We wanted to determine whether binding of IN to plasmid DNA sequences in vitro is equivalent to those sequences acting as target sites for integration. It is possible that IN complexed to an LTR terminus recognizes DNA sequences different from those recognized by free IN in solution. We used *Nde*I-linearized plasmid substrates capable of being integrated by AMV IN at high efficiency (25 to 40% of input

* Corresponding author.

substrate) into another linearized plasmid, mimicking the oligonucleotide LTR half-site reaction. Electron microscopy and partial denaturation of recombinants were used to map the frequency of IN binding to and LTR half-site strand transfer into pGEM-3 sequences. Individual recombinants could be selected as plasmids upon ligation and transformation into *Escherichia coli* HB101 cells. Mapping of donor-target site junctions was verified by restriction enzyme mapping and sequence analysis. We also compared the target sequences for this LTR half-site integration reaction with the target sequences for concerted insertion of both viral termini by AMV IN (9).

MATERIALS AND METHODS

Plasmids. A 65-bp double-stranded LTR synthetic fragment was cloned into the *Eco*RI site of pGEM-3 and is designated pFT-NdeI (8). In a separate plasmid, termed pFT-NdeI(SupF), the piAN7 *supF* gene was also cloned into pFT-NdeI at its *Hin*dIII site (9). Linearization of these plasmids with *Nde*I mimics the trimmed retrovirus DNA genome prior to integration (10). The LTR sequences at the termini of pFT-*Nde*I consisted of 24 and 36 bp of U3 and U5 sequences, respectively. All plasmids were purified by velocity sedimentation on neutral sucrose gradients.

AMV IN purification. AMV IN was purified to near homogeneity as previously described (14, 21).

DNA binding. IN was routinely bound to supercoiled pFT-NdeI or pFT-NdeI(SupF) at an approximately 30:1 molar ratio of IN dimer to DNA. The binding medium contained 20 mM Tris-hydrochloride (pH 7.5), 40 mM NaCl, 3 mM dithiothreitol, and 0.1 mM EDTA, and the binding reaction was allowed to proceed for 10 min at 37°C.

Electron microscopy. (i) DNA-IN protein cross-linking. The reaction mixtures (50 μl) described above, containing 0.25 μg of DNA and 0.29 μg of IN, were incubated. The IN dimer/DNA molar ratio was 35:1. Formaldehyde (HCHO) was added to a final concentration of 1%, and incubation was continued for an additional 10 min. The solutions were dialyzed for 4 h by placing them on Millipore filters (type VM; 0.05-μm pore size) floating on 10 mM Tris-hydrochloride (pH 7.5). Portions of the dialyzed material were digested with *Sca*I under conditions suggested by the manufacturer (Promega), except that 10 U of *Sca*I per μg of DNA was added. Even with this high enzyme concentration, 20% of the sample remained circular, presumably because of the prior cross-linking. Samples of the circular or linearized DNA-IN complexes were then prepared for electron microscopy as described below.

Because of the relatively large excess of *Sca*I used in this experiment, a control experiment was carried out to be sure that this protein was not complicating or contributing to the effects attributed to IN. When DNA (not the DNA-IN complex) was digested with the same amount of *Sca*I and prepared for electron microscopy as discussed below, no significant amount of bound protein could be observed.

(ii) Preparation of samples for electron microscopy. Undenatured DNA samples or the cross-linked DNA-IN complexes were prepared for electron microscopy by dilution in buffer to final concentrations of 20 mM sodium bicarbonate, 6 mM EDTA, 14 mM NaCl, and 9% HCHO. The solutions were incubated for 5 min at 25°C before dilution with an equal volume of formamide. The solutions were then adjusted to 0.01% cytochrome *c* and spread by the drop method (15). Samples were prepared at various degrees of partial denaturation by the procedures described above, except that the buffer also contained 12 mM NaOH and the solution was heated to temperatures between 35 and 40°C for 15 min before cooling and addition of formamide.

(iii) Measurements from electron microscopy. Electron micrographs were recorded on 70-mm film, and measurements were made by the semiautomatic method already described (24).

DNA strand transfer. Optimized integration conditions employing *Nde*I-linearized pFT-NdeI(SupF) and lambda-gtWES as the donor and target, respectively, and AMV IN were described previously (9). In the present study, only pFT-NdeI with or without *supF* was included in the reaction mixture. The reaction mixture included 20 mM Tris-hydrochloride (pH 7.5), 100 mM NaCl, 10 mM dithiothreitol, 5 mM $MgCl_2$, 10% dimethyl sulfoxide, 5% polyethylene glycol (PEG), and 0.05% Nonidet P-40. The protein-to-DNA molar ratio was generally fixed between 40:1 and 80:1, with 0.5 μg of DNA in a total volume of 50 μl. IN was preincubated with DNA for 10 min on ice, PEG was added, and the reaction was initiated by the addition of $MgCl_2$. Incubation took place for 1 h at 37°C unless otherwise indicated. The reactions were terminated by the addition of sodium dodecyl sulfate (SDS), EDTA, and proteinase K at final concentrations of 1%, 17 mM, and 1 mg/ml, respectively. The samples were further incubated at 37°C for 1 h and then subjected to two phenol-chloroform (1:1) and three ether extractions. The DNA was precipitated by ethanol. The samples were subjected to electrophoresis on 0.6 or 1% agarose gels or analyzed directly by electron microscopy as described above. Linear DNA size markers were obtained from Bethesda Research Laboratories.

5′-end labeling of plasmids. *Nde*I-linearized pFT-NdeI was dephosphorylated with calf intestine phosphatase and 5′ end labeled with [γ-^{32}P]ATP and T4 kinase (25). The labeled DNA was purified by DEAE-cellulose chromatography and ethanol precipitated.

Identification and isolation of pFT-NdeI recombinants. Dimeric Y-type recombinant DNA was isolated from agarose gel slices by treatment with Gelase (Epicentre Technologies) as recommended by the manufacturer or by extraction with phenol. The agarose gel slice was cut into multiple pieces, pH-adjusted phenol (500 μl) was added, and the sample was vortexed gently. The sample was incubated at 37°C for 2 h and then microcentrifuged for 10 min. The supernatant was removed, 10 μg of tRNA was added, and the sample was precipitated by ethanol. The two DNA extraction procedures gave similar transformation efficiencies with HB101.

DNA sequencing. fmol DNA sequencing was performed according to the manufacturer's (Promega) protocol. A list of the donor-target junction sequences is available upon request.

RESULTS

Binding of IN to pFT-NdeI(SupF). We wanted to determine whether IN bound preferentially to specific regions on supercoiled pFT-NdeI(SupF). We selected conditions which were not conducive for integration, i.e., lack of U3 or U5 DNA termini and the absence of Mg^{2+}. With 0.1 mM EDTA present, IN was bound to supercoiled pFT-NdeI(SupF), cross-linked with 1% HCHO, dialyzed, and digested with *Sca*I. Samples were prepared for electron microscopy under nondenaturing conditions (see Materials and Methods). Examination of 280 molecules showed that 80% had reacted with protein. The following structures were observed in

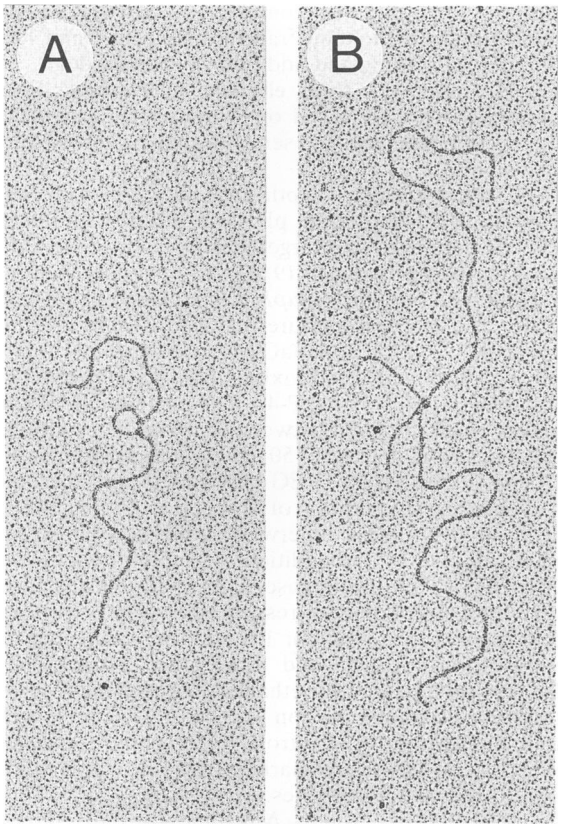

FIG. 1. Visualization of AMV IN-DNA complexes. AMV IN was bound to supercoiled pFT-NdeI(SupF) and processed as described in Materials and Methods for electron microscopy. The DNA was linearized by *Sca*I digestion after HCHO cross-linking of the DNA-protein complex. (A) Example of an interaction leading to a looped structure; (B) a less frequent interaction between two molecules.

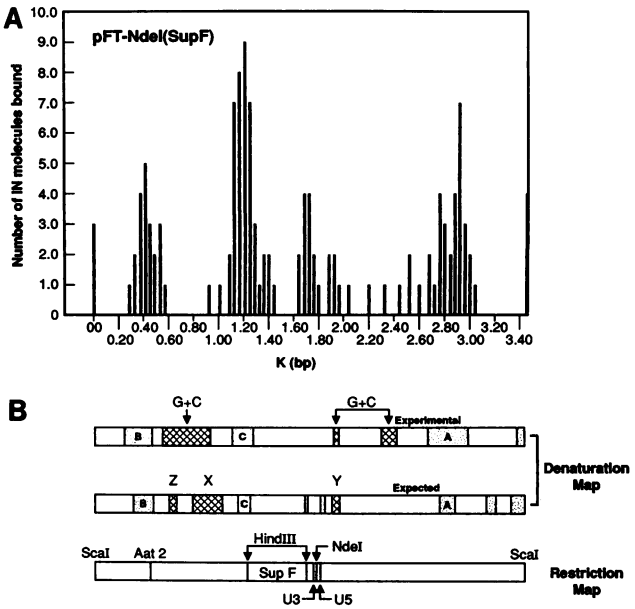

FIG. 2. Mapping the location and frequency of IN bound to pFT-NdeI(SupF). (A) Supercoiled pFT-NdeI(SupF) complexed with AMV IN was cut with *Sca*I and prepared for electron microscopy (see Materials and Methods). The position of the protein structures was measured from a DNA end. Two types of complex were measured: those involving one or two protein structures and those involving a DNA loop caused by protein interaction at two positions on the DNA. The histogram shows the distribution of single and double interacting positions (Fig. 1A). (B) The experimental denaturation map (15) was determined by partial denaturation of pFT-NdeI(SupF) cut with *Sca*I. Measurements were made with 61 and 24 molecules at average degrees of denaturation of 17 and 74%, respectively. These data resulted in the average denaturation pattern shown; the stippled areas denote the most AT-rich regions (according to the data, the percent A+T content decreases in the order A ≥ B > C). The expected denaturation pattern was determined from the known base sequences by the method of Funnel and Inman (12) with an averaging segment length of 100 bp. The expected denaturation map relies on the DNA sequence, and therefore its left end is known. The hatched regions denote GC-rich regions in the order X > Y = Z. The experimental denaturation map has been aligned to match the expected map. It can be seen that panel A, showing the distribution of IN binding sites, can be aligned so that the distribution closely matches the most AT-rich regions of the molecule. All data have been normalized to correspond to a total length of 3.401 kbp. The restriction map shows the positions and polarities of various sites on the molecule.

these 225 DNA-protein complexes. (i) DNA molecules containing a single protein structure represented 11% of the complexes (the protein was usually located at some nonterminal position and less frequently was bound at an end). (ii) About 20% of the complexes involved protein interaction at two positions within the same molecule to form a looped structure (Fig. 1A) (in a minor fraction of these molecules, the loop was located at an end). (iii) Complexes involving two DNA molecules amounted to 6%, and an example is shown in Fig. 1B. (iv) The remaining molecules (63%) were more complex and involved interaction at three or more positions within a single molecule to produce multiple loops.

Figure 2A shows the distribution of IN binding positions on single molecules with one or two points of interaction. The distribution shows at least three significant regions of preferred binding. These regions appear to correspond with the most AT-rich regions in the DNA, as determined by experimental denaturation and by sequence alignment (Fig. 2). Although there is a good correlation between the binding positions and the three most AT-rich regions within this molecule, the relationship is not absolute because binding appears to be greatest within region C, whereas this region is less AT rich than regions A and B. We do not know which DNA features are responsible for IN binding to these AT-rich regions. There is also an indication that a slight preference for binding around the U3-to-U5 segment might exist. Similar frequencies of IN binding to regions A, B, and C were observed with the plasmid without the *supF* segment (data not shown).

What is the minimal unit of IN bound to DNA? AMV IN is a dimer in solution as determined by velocity sedimentation (14) or by chemical cross-linking studies with disuccinimidyl suberate (DSS) and as determined by denaturing SDS-polyacrylamide gel electrophoresis. AMV IN exists as two polypeptides upon reducing SDS-polyacrylamide gel electrophoresis: one minor full-length polypeptide (286 amino acids) and one major slightly truncated polypeptide (approximately 282 amino acids) at its carboxyl terminus (26). Treatment of IN in the range of 0.32 to 2.2 μM with DSS (2 mM) or varying the DSS concentration (0.01 to 2 mM) with IN at 2.2 μM resulted in the formation of two

separate cross-linked polypeptide chains with molecular weights of 62,000 and 64,000 upon reducing SDS-polyacrylamide gel electrophoresis (38). Therefore, each IN-DNA complex observed in Fig. 1 is probably a dimer, although larger oligomers may be present. It has recently been shown that bacterially expressed Rous sarcoma virus IN functions as a multimer in solution (16).

Formation of pFT-NdeI recombinants. Is there a correlation between selection of DNA binding sites on pFT-NdeI by AMV IN and target site recognition by IN for strand transfer? AMV IN was incubated with *Nde*I-linearized pFT-NdeI under optimal strand transfer conditions for concerted pairwise insertion of pFT-NdeI(SupF) into lambda-gtWES (9). We wanted to quantitate the formation of pFT-NdeI recombinants catalyzed by IN without lambda present. The substrate was 5′ end labeled at the *Nde*I site before strand transfer. Recombinants larger than unit-length pFT-NdeI migrated more slowly on 1% agarose gels (Fig. 3A). IN was preincubated with pFT-NdeI on ice for 10 min before strand transfer, which was initiated by the addition of Mg^{2+} and incubation at 37°C. Preincubation was done to permit binding of IN to internal DNA sequences as well as to the U3 and U5 termini. Two major recombinant species (1 and 2) were observed (Fig. 3A) just below and above the 6,557-bp lambda restriction size marker, respectively, and numerous other minor species were observed between them (Fig. 3A). The major more slowly migrating species was evident much earlier in the reaction (Fig. 3A, lane 3), with the faster-migrating species arising later in the reaction (Fig. 3A, lanes 7 and 8). Gel slices encompassing the entire region between both major ^{32}P-labeled species and gel slices of the unused linear substrate were excised, and the Cerenkov radiation was counted. A total of 14, 19, 22, 26, and 27% of the input substrate was integrated by IN after 5, 10, 15, 30, and 60 min, respectively (Fig. 3A, lanes 3 to 7). No major radioactivity (<0.2% of input substrate) was observed in the corresponding control lanes, 1 and 2, at the DNA recombinant positions. Other more slowly migrating DNA recombinants were observed between the 9,416- and 23,130-bp lambda markers (Fig. 3A, lanes 7 and 8). An additional 18% of the input DNA was observed as recombinants in this region. These larger, more slowly migrating species are more evident in Fig. 3B.

To address the substrate specificity of these DNA recombinants produced by IN, pGEM-3 was linearized with *Hin*dIII. Digestion of this plasmid by *Hin*dIII exposes a 3′-OH CA dinucleotide recessed by 4 nucleotides on both termini. After a strand transfer reaction with AMV IN, the DNAs were purified and subjected to electrophoresis. Careful examination of Fig. 3B, lanes 1 to 3, showed that a few minor recombinant products were observed with *Hin*dIII-linearized DNA, while both pFT-NdeI (lanes 4 and 6) and pFT-NdeI(SupF) (lanes 7 to 9) digested with *Nde*I were capable of forming the normal quantity of DNA recombinants. The efficient formation of these pFT-NdeI recombinants was not dependent on the presence of dimethyl sulfoxide (Fig. 3B, compare lane 5 with lane 6 or lane 8 with lane 9) or the presence of PEG in the strand transfer reactions (data not shown).

Identification of recombinant structures by electron microscopy. To visualize the structures of the pFT-NdeI recombinants, several series of reactions were performed, and the DNA products were purified and subsequently processed for spreading and visualization by electron microscopy. The quantity of DNA recombinants was sufficiently high to directly count the recombinants without separation from the input substrate. Several examples of the major species are illustrated in Fig. 4, and all of the recombinants were quantitated (Table 1).

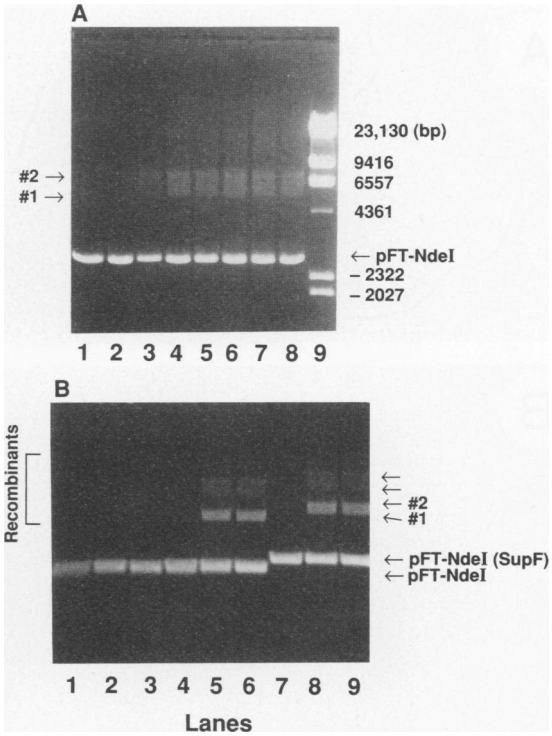

FIG. 3. Formation of covalently linked recombinants by AMV IN with several DNA substrates. (A) *Nde*I-linearized pFT-NdeI was 5′ end labeled with [γ-^{32}P]ATP. The DNA (0.25 μg, 2 × 10^5 cpm) was subjected to standard strand transfer conditions and purified as described in Materials and Methods. Lane 1 contained no IN; lane 2 contained IN, but the reaction was immediately stopped by adding SDS and EDTA. The samples were incubated for various times (see Results) with IN (lanes 3 to 7, with lane 8 being a duplicate of lane 7). Each sample was subjected to electrophoresis, the appropriate regions were excised, and radioactivity was counted. Recombinant species 1 and 2 are indicated. The numbers on the right indicate the sizes of the linear DNA markers in lane 9. (B) Standard strand transfer conditions were used. The samples were incubated for 60 min at 37°C with Mg^{2+}. *Hin*dIII-linearized pGEM-3 (0.45 μg) was incubated without IN (lane 1) or at a 70:1 ratio of IN dimers to DNA (lanes 2 and 3). No dimethyl sulfoxide was present in lanes 3, 6, and 9. The same protocol and gel loading were used for *Nde*I-linearized pFT-NdeI (0.52 μg) (lanes 4 to 6) and *Nde*I-digested pFT-NdeI-(SupF) (0.52 μg) (lanes 7 to 9). The entire sample for each reaction was subjected to electrophoresis on a 0.6% agarose gel.

By measuring the length of both molecules involved in pFT-NdeI recombinants and knowing that either terminus is capable of strand transfer, it was possible in most cases to discern which DNA molecules were the donor molecules and which were the target molecules. For example, Fig. 4B demonstrates a typical dimeric Y-type structure observed with pFT-NdeI; similar structures were observed with pFT-NdeI(SupF) (data not shown). The position of the insertion site on the target molecule is variable (Fig. 5). Approximately 10% of the recombinants were of this dimeric Y type with either substrate (Table 1). Both the target and donor molecules of both plasmids were full-length molecules. By using ϕX174 DNA as a standard (5.386 kbp), the 2.932-kbp pFT-NdeI monomeric unit was found to correspond to 2.85

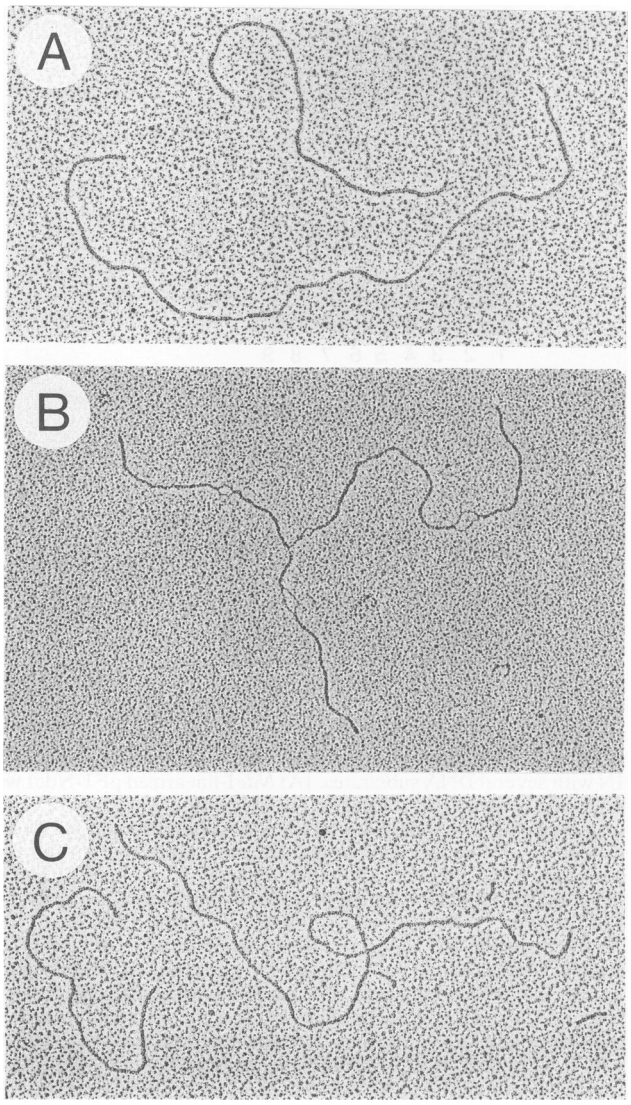

FIG. 4. Determining sizes and structures of pFT-NdeI recombinants. *Nde*I-digested pFT-NdeI was subjected to standard strand transfer conditions at 37°C for 1 h. The DNA was purified and spread for visualization by electron microscopy. The percentages of the different recombinant structures identified are indicated in Table 1. (A) Examples of monomeric linear and dimeric linear molecules (Table 1). Measurements indicate that the dimeric type may be shorter than expected for a true dimer. In two experiments, the lengths were found to be 450 ± 100 bp and 410 ± 100 bp shorter than the expected dimeric length of 5,864 bp. (B) Example of a Y-type molecule (Table 1). In this particular example, the upper right arm corresponds to monomeric-length DNA, while the other two arms together represent a second monomer. These molecules were partially denatured in order to map junction positions. (C) An extreme Y-type structure in which the junction is close to the end of the target molecule (see text). The left arm is monomeric, while the other two arms together represent a second monomer.

± 0.07 kbp by electron microscopy measurement. Multi-Y-type molecules (Table 1; 2 to 5% of the recombinants), which consisted of two donors and one target or one donor and two targets, also were full-length molecules when either plasmid was used as a substrate. Branched circular molecules (Table 1; up to 3% of the recombinants) were unit-length molecules

TABLE 1. pFT-NdeI recombinant species[a]

Species	% of total recombinants			
	pFT-Nde(SupF)		pFT-NdeI	
	With PEG (n = 478)	Without PEG (n = 755)	With PEG (n = 248)	Without PEG (n = 605)
Monomeric linear[b]	78.0	85.0	80.0	78.0
Dimeric linear[c]	0.4	0.0	10.0	2.0
Dimeric Y type[d]	15.0	10.0	7.0	13.0
Multi-Y type[e]	4.0	2.0	2.0	5.0
Branched circle[f]	3.0	3.0	0.8	2.0

[a] In these experiments, three molecules were omitted from the count: residual circular substrates, X-type molecules (assumed to be accidental overlaps of two linear monomer units), and several complex types. In control experiments with either substrate, incubation was carried out in the absence of IN, and only monomer linear types were observed in 300 observations. n, number of molecules counted.

[b] Unreacted substrate. See example in Fig. 4A.

[c] These molecules were slightly shorter than expected for a true end-to-end dimer (see text). An example is shown in Fig. 4A.

[d] Y-shaped structures in which one arm is of monomeric length and the other two arms together correspond to a second monomeric unit. Examples are shown in Fig. 4B and C.

[e] Complex Y types, usually involving three monomers, but sometimes involving four or more units.

[f] Monomeric units in which one end has been joined to an internal region of the molecule.

which contained one terminus integrated internally in the same molecule (autointegration).

The location of each structurally different recombinant species on the agarose gel has not been defined (Fig. 3). Nonetheless, it appears that DNA recombinant species 1 and 2 and the species between them consist of various extremes of the dimeric Y types and that their migration is dependent on the location of the donor molecule on the target. The branched circle could also migrate to this region. The multi-Y types and other complex structures probably migrate more slowly on agarose gels than recombinant species 2 (Fig. 3B).

A second major recombinant species was observed with pFT-NdeI but not with pFT-NdeI(SupF) (Table 1; 2 to 10% of the recombinants), although the yield was variable. AMV IN was capable of forming a significantly high percentage of dimeric linear molecules approximately 410 to 450 bp shorter than that expected for two linear molecules simply ligated together (Fig. 4A). This species may migrate slightly faster than the 6,557-bp size marker (Fig. 3A). An expected full-length dimeric linear molecule would be 5.8 ± 0.13 kbp long, but the experimentally determined length was 5.41 ± 0.11 kbp. Few linear dimeric molecules (0.3%) were observed in the control samples containing the input substrate, which was incubated under strand transfer conditions in the absence of IN.

Preference of U3 termini over U5 termini acting as donor substrates. Experimentally determined denaturation maps of DNA together with the known sequence permit the identification of termini on linear molecules (12) (Fig. 1). pFT-NdeI(SupF) linearized by *Nde*I was subjected to several different degrees of denaturation. The experimentally derived denaturation map was aligned with the expected map similar to that shown in Fig. 2. Thus, we were able to align the denatured monomeric linear map of pFT-NdeI(SupF) with similarly denatured dimeric Y types. This alignment to the donor molecule of Y types showed that of 140 molecules scored (Fig. 5A), 66% of the inserted termini were U3 while the rest were U5. The U3 terminus of autointegrated mole-

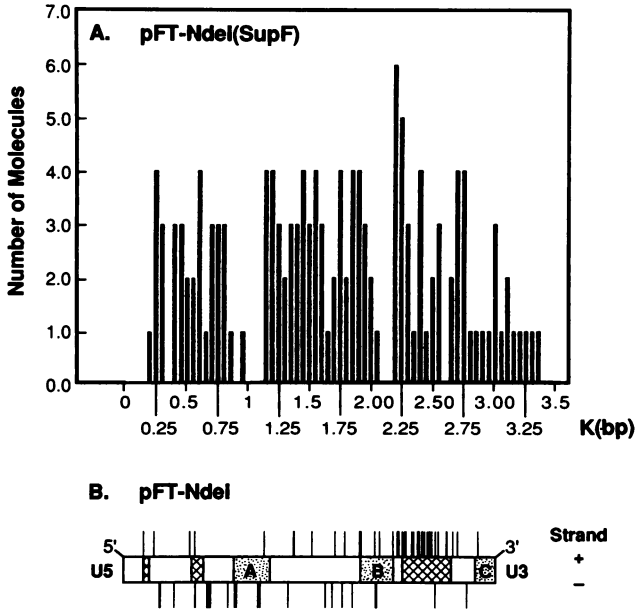

FIG. 5. Mapping the frequency and location of the donor molecule on pFT-NdeI(SupF). (A) As analyzed by electron microscopy, the distribution and frequency of junction positions in the target molecule are indicated. A total of 140 junctions were mapped. Both the U3 and U5 donor termini are included in the scoring. (B) Experimental denaturation map obtained for the experiment in A, permitting alignment of the U5 (left) and U3 (right) termini. Note that the *supF* gene is not present but would map immediately to the right of region C (Fig. 2). Sequenced recombinants depicting the donor-target junctions of pFT-NdeI recombinants were plotted on this denaturation map. Each vertical line represents one insertion on either the plus strand (above map) or the minus strand (below map) of pFT-NdeI. The experimentally derived AT (▨)- and GC (▩)-rich regions are also shown (see legend to Fig. 2 for further explanation).

cules (branched circles) or trimeric multi-Y types was also the preferred (2:1) donor terminus.

Mapping of LTR half-site integration events on pFT-NdeI (SupF) by electron microscopy. The frequency and distribution of the insertion positions on target pFT-NdeI(SupF) molecules were scored (Fig. 5A). Only dimeric Y types were analyzed by using monomeric length substrate denaturation maps as a polarity guide to map target sites. The distribution of donor-target junctions (140 scored) appeared to be nearly random. Integration did not occur at a higher rate in AT-rich region A, B, or C, even though IN bound frequently to these regions (Fig. 2). Note that few target sites near either the U3 or U5 termini were evident and that integration into GC-rich regions (Y and X + Z) was not restricted, although IN bound infrequently to these regions (Fig. 2).

There are several possible reasons why IN is not able to efficiently integrate DNA into region A, B, or C. First, these regions may be saturated with IN (Fig. 2), thereby physically blocking IN complexed to an LTR terminus which is capable of strand transfer. Second, donor molecules in these regions could be unstable because of partial denaturation of the recombinants upon examination by electron microscopy to map junction positions. This instability may be due to the lack of covalent integrity in this region, because other IN molecules not involved in integration may introduce single-stranded cleavages in these AT-rich regions (14, 32). This appears not to be the case because even when 80% of the DNA was denatured, we observed 10% of the sample as Y-type structures (200 molecules counted).

Mapping donor-target junctions by restriction enzyme analysis and sequencing. The studies described above permitted a global evaluation of the abilities of AMV IN to bind DNA sequences and to integrate within the same sequences. We wanted to produce a more refined map for target sites on pFT-NdeI and to verify that AT-rich regions of pGEM-3 do not serve as more efficient integration sites than GC-rich regions. Our analyses were limited to recombinant species 1 and 2 and recombinants migrating between them on agarose gels (Fig. 3) because we wanted to avoid the even more complex multi-Y types and other complex structures (Fig. 3B, most slowly migrating bands, and Table 1).

We took advantage of the altered migration of recombinants from unused *Nde*I-linearized substrate on agarose gels. Standard strand transfer reactions were performed, and the recombinants were subjected to electrophoresis on 0.6 or 1% agarose gels (Fig. 3A). The extracted gel slices encompassed both species 1 and 2. The DNA was eluted and ligated under dilute conditions. The ligated DNA was transformed into HB101 with ampicillin selection. Several hundred colonies were obtainable from each DNA preparation. The colonies were screened by DNA plasmid size by comparison with supercoiled pFT-NdeI on agarose gels. Plasmids larger than pFT-NdeI were subjected to digestion with *Hin*dIII, *Pvu*II, *Eco*RI, *Nde*I, or a combination of these enzymes to map donor-target junctions. Sequence analysis was performed by the fmol sequencing technique and by 5'-^{32}P-labeled primers which were hybridized to isolated *Hin*dIII or *Pvu*II fragments containing the donor-target site. U3 donor-target sites were analyzed with a primer just 3' of the pGEM-3 *Hin*dIII site, and U5 donor-target sites were analyzed with the T7 promoter primer.

Approximately 25% of the colonies contained supercoiled plasmids larger than pFT-NdeI, while the rest appeared to be unit length. Only one of the recombinants was slightly smaller (because of autointegration) than pFT-NdeI. The two DNA extraction methods used for isolation of recombinants produced similar yields of these larger (dimeric Y types) recombinants (data not shown). These data suggest that the ligation of the *Nde*I termini of a molecule serving as a target in a recombinant molecule (Fig. 4B) was more efficient (75%) than the ligation of target-donor *Nde*I termini in a recombinant molecule upon selection in HB101. All recombinants had the 65-bp *Eco*RI fragment originally inserted into pFT-NdeI.

Sequence analysis of 70 pFT-NdeI recombinants confirmed that the U3 terminus is preferred to the U5 terminus, as previously shown (Fig. 5A). A total of 63% of the recombinants had U3 insertions, while 37% had U5 insertions. Apparently because of pGEM-3 genome polarity, only recombinants which had the plus-strand U3 terminus inserted into plus-strand sequences (Fig. 5B) or had the minus-strand U5 terminus inserted into minus-strand sequences (Fig. 5B) were able to survive as replicating plasmids. All of the inserted viral DNA termini ended with the conserved CA_{OH} dinucleotide.

The distribution of these donor-target junctions derived from biologically selected pFT-NdeI recombinants is shown in Fig. 5B and is tabulated in Table 2. There is a correlation between what was observed by electron microscopy and what was observed by sequencing for mapping donor-target junctions. For example, from the left end between nucleotides 1 and 853 of pFT-NdeI, the number of recombinants mapping there was approximately 24% for each method

TABLE 2. Analysis of integration and DNA binding sites on pFT-NdeI

Nucleotides	% of insertions/section of pFT-NdeI[a]		No. of bp/integration[b]	% of total bound IN molecules[c]	% G+C[d]
	Electron microscopy	Sequenced			
1–853	22.8	25.7	17.4	11.6	55.1
854–1144 (region A)	4.2	5.7	26.4	26.7	37.8
1145–1904	30.0	17.1	14.0	8.9	50.0
1905–2168 (region B)	7.1	8.5	17.5	13.3	39.8
2169–2781	25.7	41.4	9.3	8.9	59.4
2782–2932 (region C)	4.2	1.4	21.4	30.3	42.7

[a] A total of 140 recombinant molecules (Fig. 5A) were counted and the percentage of insertions for each section of pFN-NdeI was calculated. The *supF* region of pFT-NdeI(SupF) was not tabulated. A similar calculation was performed for the sequenced pFT-NdeI recombinants (70 total scored) (Fig. 5B).

[b] The number of base pairs per integration was calculated by taking the actual number of integrations per region from both methods of counting this number and dividing it into the number of base pairs per region.

[c] The number of IN molecules (112 total counted) bound per region was derived from Fig. 2. The *supF* region of pFT-NdeI(SupF) was not tabulated for bound IN.

[d] The overall G+C content of pFT-NdeI is 50.9%.

(Table 2). For AT-rich regions A, B, and C, there were similar numbers of recombinants with either method, suggesting that the observed frequency of IN-mediated integration of either U3 or U5 substrates into these AT-rich regions is correct (Table 2 and Fig. 5). Integration appears to be nearly random, except that there appears to be a bias toward a GC-rich region (29 insertions between nucleotides 2169 and 2781) with nearly double the density of integrations per base pair relative to the other regions (Table 2). All of the insertions were at individual sites, except for three identical recombinants from two individual strand transfer reactions. The insertion site was at nucleotide 580 (pGEM-3 sequence) on the minus strand of pFT-NdeI (Fig. 5B). Table 2 also shows the frequency of bound IN per region as shown in Fig. 2.

We compared the sequence composition of the target sites of the LTR half-site reaction obtained by using pFT-NdeI (Fig. 5B) with that of the target sites of the LTR full-site reaction obtained by using lambda DNA as a target (9). With the LTR full-site reaction, a pattern of high and low G+C content at positions symmetrically related about the target site duplication center (6 bp) was observed (9) (Fig. 6). This pattern of sequence composition appeared to be limited to 14 bp around the duplication center (Fig. 6, positions −7 to +7). Alignment of the 70 LTR half-site recombinants relative to the site of insertion indicated a target site sequence bias similar to that found for the LTR full-site recombinants. However, the sequence pattern for the LTR half-site products was confined to the positions at the immediate LTR-target joint and positions 5′ of the joint (Fig. 6, positions −3 to −7). The LTR half-site set did not have the dyad symmetry relations observed for the LTR full-site set (Fig. 6, positions −7 to +7) (9).

DISCUSSION

We have compared the abilities of AMV IN to utilize specific plasmid sequences for binding and to utilize such sequences for the LTR half-site integration reaction. AMV IN prefers to bind AT-rich regions in the absence of strand transfer conditions but does not favor these regions for integration. The U3 termini are preferred approximately 2 to 1 over the U5 termini for both the LTR half-site integration reaction (this report) and for the LTR full-site integration reaction (9). IN is also capable of forming both inter- and intra-looped DNA structures via apparent protein-protein contacts between IN molecules bound at different locations on virus-like DNA plasmids.

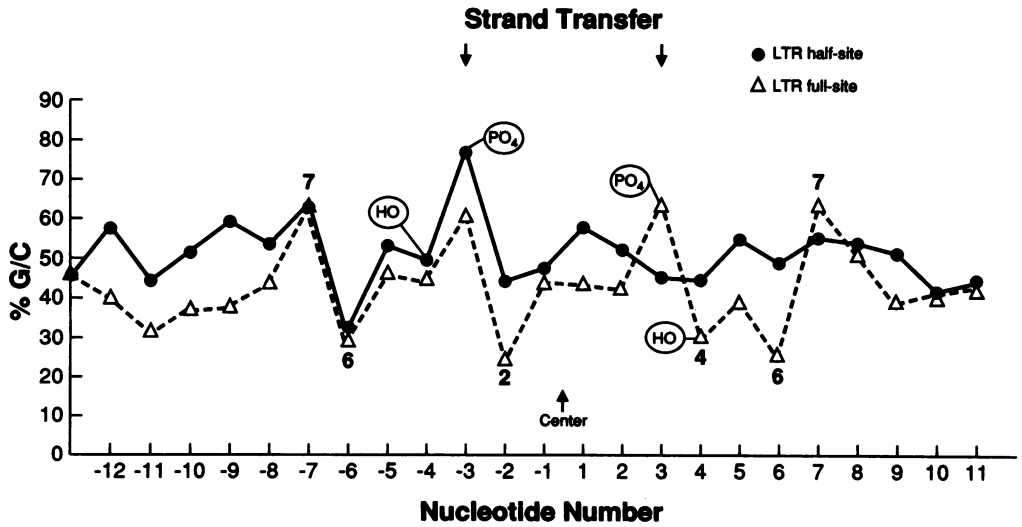

FIG. 6. Comparison of LTR half-site and full-site strand transfer reactions. Seventy pFT-NdeI recombinants were sequenced, and the target site sequences were aligned. The percent G+C composition for each position encompassing the LTR half-site insertion is graphed (●). The top left arrow indicates attachment of the viral CA_{-OH} to nucleotide −3 for the pFT-NdeI LTR half-site reaction (this report). The PO_4 and HO symbols facilitate visualization of the strand transfer for the LTR half-site or full-site (△) reaction (9). The bottom arrow indicates the center of the 6-bp duplication center observed with LTR full-site insertion. There is no duplication of the target site upon LTR half-site strand transfer.

Recognition of DNA sequences by AMV IN for LTR half-site integration and DNA binding in the absence of strand transfer appear not to be equivalent. Comparison of data for these two events (Fig. 2 and 5 and Table 2) suggests that IN prefers AT-rich regions for binding but not integration. It is possible that IN in solution has an available binding site different from the target binding site on IN already interacting with recessed LTR termini, which is primed for strand transfer. Some evidence suggests that there are two different DNA binding regions on IN (17, 39), one of which prefers single-stranded DNA over double-stranded DNA (26). Although IN requires specific U3 or U5 LTR terminal sequences for 3′ OH processing and strand transfer, it is unclear how IN complexed to these termini binds to and simultaneously recognizes target DNA sequences for the integration reaction (see below).

A 160S nucleoprotein complex identified in the cytoplasm of virus-infected cells is fully capable for 3′-end processing and subsequent strand transfer of the recessed termini. The location and quantity of IN molecules in these 160S complexes are unknown (7), although there appear to be numerous polymerase gene molecules (reverse transcriptase and IN) per virion. We have demonstrated that AMV IN is capable of forming looped DNA structures (Fig. 1) as well as complex interactions involving three or more positions within a single DNA molecule. These data suggest that IN in the 160S complex may also be capable of protein-protein contacts, thereby helping to maintain both its structural integrity and its integration potential. Since AMV or Rous sarcoma virus IN appears to be multimeric in solution (this report and references 14 and 16), the possibility that a dimer or a tetramer is responsible for the LTR full-site integration reaction exists.

What causes formation of dimeric linear molecules with pFT-NdeI but not with pFT-NdeI(SupF) by AMV IN is unclear (Table 1 and Fig. 4A). These slightly shortened (ca. 410-bp) dimers could be due to the formation of a special precursor class of extreme Y-type recombinants, as shown in Fig. 4C (i.e., an integration hot spot located approximately 400 bp from a terminus). The approximately 400-bp tail on the extreme Y-type molecule could then act as a donor terminus (Fig. 4C). The local concentration of DNA at this new potential donor-target junction would be greater than that with molecules free in solution. Examination of 145 pFT-NdeI dimeric Y-type molecules by electron microscopy failed to demonstrate preferential formation of such an extreme Y-type molecule (data not shown). Rather, the distribution of donor-target junctions was very similar to that obtained with pFT-NdeI(SupF) (Fig. 5), except that there was a gap between 250 and 325 bp from an end which lacked donor-target junctions. The apparent lack of numerous Y-type recombinants mapping between 250 and 325 bp from a terminus may be due to the conversion of these extreme Y types into the linear dimers (Fig. 4C).

Comparison of the target site sequence composition indicated sequence bias for both the LTR half- and full-site reactions. However, the positional bias of the LTR half-site reaction was confined to sequences at the immediate insertion site and sequences 5′ to the insertion site. To the 3′ of the LTR joint, not including the first nucleotide, there was no apparent coincidence of G+C content with the LTR full-site sequences. It is interesting that the sequence bias for the LTR half-site reaction was 5′ to the insertion site. This suggests that the complex which integrated the single LTR end recognized target sequences 5′ to the insertion site. This observation gains significance in that the recognition of sequences 5′ to LTR termini must also be necessary for the demonstrated specificity of the LTR 3′-end processing reaction (4, 8, 19, 30). It may be that recognition of the LTR ends and recognition of the target site share the same fundamental mechanism, as has been previously suggested (5).

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (CA16312 to D.P.G. and GM14711 to R.B.I.).

We thank C. Hall for typing the manuscript.

REFERENCES

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. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein *in vitro*. Science **249**:1555–1558.
4. **Craigie, R., T. Fujiwara, and F. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. Cell **62**:829–837.
5. **Engelman, A., K. Mizuuchi, and R. Craigie.** 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell **67**:1211–1221.
6. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 DNA *in vitro*. Proc. Natl. Acad. Sci. USA **87**:4164–4168.
7. **Farnet, C. M., and W. A. Haseltine.** 1991. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J. Virol. **65**:1910–1915.
8. **Fitzgerald, M. L., A. C. Vora, and D. P. Grandgenett.** 1991. Development of an acid-soluble assay for measuring retrovirus integrase 3′-OH terminal nuclease activity. Anal. Biochem. **196**:19–23.
9. **Fitzgerald, M. L., A. C. Vora, and D. P. Grandgenett.** 1992. Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. J. Virol. **66**:6257–6263.
10. **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.
11. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
12. **Funnel, B. E., and R. B. Inman.** 1979. Comparison of partial denaturation maps with the known sequence of SV40 and ϕX174 DNA. J. Mol. Biol. **131**:331–340.
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. **Inman, R. B., and M. S. Schnos.** 1970. Partial denaturation of thymine- and BU-containing lambda DNA in alkali. J. Mol. Biol. **49**:93–98.
16. **Jones, K. S., J. Coleman, G. W. Merkel, T. M. Laue, and A. M. Skalka.** 1992. Retroviral integrase functions as a multimer and can turn over catalytically. J. Biol. Chem. **287**:16037–16040.
17. **Kahn, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1991. Retrovirus integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19**:851–860.
18. **Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. Cell **63**:87–95.
19. **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.
20. **Kitamura, Y., Y. M. H. Lee, and J. M. Coffin.** 1992. Nonrandom integration of retroviral DNA *in vitro*: effect of CpG methylation. Proc. Natl. Acad. Sci. USA **89**:5532–5536.
21. **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.
22. **Leavitt, A. D., R. B. Rose, and H. E. Varmus.** 1992. Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. J. Virol. **66**:2359–2368.
23. **Lee, Y. M. H., and J. M. Coffin.** 1990. Efficient autointegration of avian retrovirus DNA in vitro. J. Virol. **64**:5958–5965.
24. **Littlewood, R. K., and R. B. Inman.** 1982. Computer-assisted DNA length measurements from electron micrographs with special reference to partial denaturation mapping. Nucleic Acids Res. **10**:1691–1706.
25. **Misra, T. K., D. P. Grandgenett, and J. T. Parsons.** 1982. Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. J. Virol. **44**:330–343.
26. **Mumm, S. R., R. Horton, and D. P. Grandgenett.** 1992. v-Src enhances phosphorylation at Ser-282 of the Rous sarcoma virus integrase. J. Virol. **66**:1995–1999.
27. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes *in vitro*. EMBO J. **11**:291–303.
28. **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.
29. **Sandmeyer, S. B., V. W. Bilanchone, D. J. Clark, P. Morcos, G. F. Carle, and G. M. Brodeur.** 1988. Sigma elements are position-specific for many different yeast tRNA genes. Nucleic Acids Res. **16**:1499–1507.
30. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expressed in Escherichia coli possesses selective DNA cleaving activity. Proc. Natl. Acad. Sci. USA **87**:5119–5123.
31. **Shih, C. C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.
32. **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.
33. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
34. **Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. Plasterk.** 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. J. Virol. **65**:4636–4644.
35. **Vink, C., D. C. van Gent, and R. H. A. Plasterk.** 1990. Integration of human immunodeficiency virus types 1 and 2 DNA in vitro by cytoplasmic extracts of Moloney murine leukemia virus-infected mouse NIH 3T3 cells. J. Virol. **64**:5219–5222.
36. **Vink, C., E. Yeheskiely, G. A. van de Marel, J. H. van Boom, and R. H. A. Plasterk.** 1991. Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. Nucleic Acids Res. **19**:6691–6698.
37. **Vora, A. C., M. L. Fitzgerald, and D. P. Grandgenett.** 1990. Removal of 3′-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. J. Virol. **64**:5656–5659.
38. **Vora, A. C., and D. P. Grandgenett.** Unpublished data.
39. **Woerner, A. M., M. Klutch, J. G. Levin, and C. J. Marcus-Sekura.** 1992. Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. AIDS Res. Hum. Retroviruses **8**:297–304.