Virology **286,** 304–316 (2001)
doi:10.1006/viro.2001.0978, available online at http://www.idealibrary.com on IDEAL®

Paired DNA Three-Way Junctions as Scaffolds for Assembling Integrase Complexes

Erik P. Johnson and Frederic D. Bushman[1]

Infectious Disease Laboratory, The Salk Institute, La Jolla, California 92037

Received February 16, 2001; returned to author for revision April 17, 2001; accepted April 21, 2001

Early steps of retroviral replication involve reverse transcription of the viral RNA genome and integration of the resulting cDNA copy into a chromosome of the host cell. The initial DNA breaking and joining steps of integration are carried out by the virus-encoded integrase enzyme. Integrases bind specifically to the ends of the unintegrated viral cDNA but nonspecifically to target DNA. Conventional assays *in vitro* reveal primarily the nonspecific DNA binding mode, complicating studies of integrase–DNA complexes. Here, we report an investigation of unconventional DNA structures useful for positioning integrase at predetermined sites. We find that paired DNA three-way junctions can be used to mimic branched DNAs normally formed as reaction intermediates. The three-way junctions differ from authentic intermediates in the connectivity of the DNAs, which, in contrast to the authentic intermediate, allow formation of stable DNA structures under physiological conditions. Assays *in vitro* showed that integrase can direct hydrolysis at sequences resembling the viral cDNA ends within the three-way junction, but not on junctions with mutant sequences. Changing the spacing between the paired three-way junctions disrupted the cleavage pattern, emphasizing the importance of the correct DNA scaffold. DNase I footprinting studies revealed protection of specific bases at the terminus of the LTR in the three-way junction complex, but not on control linear DNA, specifying the locations of tight interactions between integrase and DNA. Paired DNA three-way junctions are attractive reagents for structural studies of integrase–DNA complexes.

Key Words: HIV; integrase; Rous sarcoma virus; DNA; retrovirus.

INTRODUCTION

A critical step in the replication of a retrovirus is the integration of the viral genome into the host DNA. This function is carried out by the virus-encoded integrase enzyme (for reviews, see Brown, 1997; Esposito and Craigie, 1999; Hansen *et al.,* 1998). Integration has been the focus of intense study partly as a model of enzymatic DNA breakage and joining reactions, and partly because it is the only enzyme encoded by the human immunodeficiency virus (HIV) for which there are not yet clinically useful inhibitors (Bushman *et al.,* 1998; Hansen *et al.,* 1998; Hazuda *et al.,* 2000; Mathe, 2000; Pommier and Neamati, 1999; Thomas and Brady, 1997). In this study, we have chosen to examine the integrase of Rous sarcoma virus (RSV), which is relatively soluble compared to many integrases and for which a structure of a dimer containing two of the three protein domains has recently become available (Bujacz *et al.,* 1995; Yang *et al.,* 2000).

Upon entry into the cell, the retroviral RNA genome is reverse transcribed to generate a double-stranded linear cDNA copy of the viral RNA genome. Each end of the linear viral cDNA contains a long terminal repeat sequence (LTR), which is composed of U3, R, and U5 regions. *In vivo,* integrase enzyme assembles onto the LTR cDNA, along with several other proteins, to form the preintegration complex (PIC) (Brown, 1997; Hansen *et al.,* 1998).

Integration begins with removal of a short oligonucleotide from each 3′ end of the viral cDNA, a dinucleotide in the RSV case, leaving a two-base 5′ overhang (Fig. 1A). The recessed 3′ ends are then joined to 5′ ends in the target DNA (strand transfer reaction; Fig. 1B) (Brown *et al.,* 1989; Fujiwara and Mizuuchi, 1988). Purified integrase protein can carry out the terminal cleavage and strand transfer reactions *in vitro* (Bushman *et al.,* 1990; Craigie *et al.,* 1990; Katz *et al.,* 1990; Katzman *et al.,* 1989; Sherman and Fyfe, 1990). Insertion of the DNA ends occurs on either side of a DNA major groove (Pruss *et al.,* 1994; Pryciak *et al.,* 1992). The gaps resulting from integration are likely to be repaired by host DNA repair enzymes (Yoder and Bushman, 2000). This leaves a characteristic 6-bp duplication on either end of the inserted provirus for RSV, while for HIV the duplication is 5 bp. Target site selection appears to be only loosely sequence-dependent (Bor *et al.,* 1996; Carteau *et al.,* 1998; Fitzgerald and Grandgenett, 1994; Pryciak *et al.,* 1992a,b; Pryciak and Varmus, 1992). However, integration is favored *in vitro* at sites of DNA distortion, often associated with curvature of the target DNA (Bor *et al.,* 1995; Pruss *et al.,* 1994; Pryciak *et al.,* 1992a,b; Pryciak and Varmus, 1992).

The integrase molecule is composed of three protein domains. The amino-terminal domain binds zinc, the central domain contains the active site, and the carboxyl-

[1] To whom reprint requests should be addressed. Fax: (858) 554-0341. E-mail: bushman@salk.edu.

0042-6822/01 $35.00

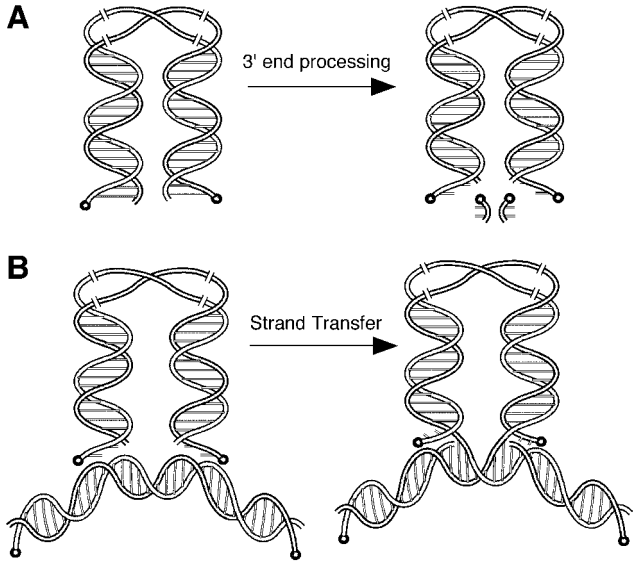

FIG. 1. Integration of cDNA by Rous sarcoma virus integrase. The 5′ ends of the DNA strands are represented by spheres. After assembly of the viral cDNA and the integrase protein into the preintegration complex, integrase removes a dinucleotide from the 3′ end of each end of the viral cDNA (A). The modified preintegration complex then binds to the host target DNA. A strand transfer reaction then takes place in which the 3′ ends of the viral cDNA are joined to the target DNA, simultaneously producing a nick in the target strand (B).

terminal domain contributes to DNA binding (Bujacz *et al.,* 1995; Cai *et al.,* 1997; Dyda *et al.,* 1994; Eijkelenboom *et al.,* 1997; Yang and Steitz, 1995). Each of the three domains has been shown to dimerize, and the full-length protein has been shown to function as a multimer (Eijkelenboom *et al.,* 1995; Engelman *et al.,* 1993; Lodi *et al.,* 1995; van Gent *et al.,* 1993). Recent structures of two domain integrase derivatives, containing the catalytic and C-domains of RSV, SIV, and HIV, provide further evidence for dimerization and allow higher order complexes to be modeled (Chen *et al.,* 2000a,b; Yang *et al.,* 2000).

While catalysis is dependent on the sequences at the ends of the viral cDNA, retroviral integrase binds DNA largely nonspecifically, showing at most slight preference for binding to the LTR ends *in vitro* (Engelman *et al.,* 1994; Hazuda *et al.,* 1994; Pemburton *et al.,* 1996; Schauer and Billich, 1992; van Gent *et al.,* 1991; Vink and Plasterk, 1993; Woerner *et al.,* 1992; Yi *et al.,* 1999; Yoshinaga *et al.,* 1994). Integrase likely has multiple DNA binding modes, a specific mode for binding the LTR ends and a nonspecific binding mode for capturing target DNA. This has complicated efforts to study DNA–integrase complexes, since nonspecific binding obstructs efforts to analyze specific complexes.

In an effort to constrain the position of integrase on DNA, we have developed branched substrate molecules designed to form scaffolds that restrict integrase binding. Figure 2A illustrates the DNA junction expected in an integration intermediate. The upper and lower arms model the viral cDNA ends, while the other two arms model target DNA. Several groups have studied the action of HIV integrase on such substrates (named "crossbone disintegration" substrates) (Chow and Brown, 1994b; Mazumder *et al.,* 1994; van den Ent *et al.,* 1994). The two halves of the substrate are joined by only the 5 target base pairs between the LTR arms, forming the crossbone structure. The 5-bp target region is not expected to be stably base paired under physiological or experimental conditions, so in solution the crossbone substrate will be present as two halves. Integrase was found nevertheless to carry out reactions in which a 3′ end in one half attacked DNA sites in the other, indicat-

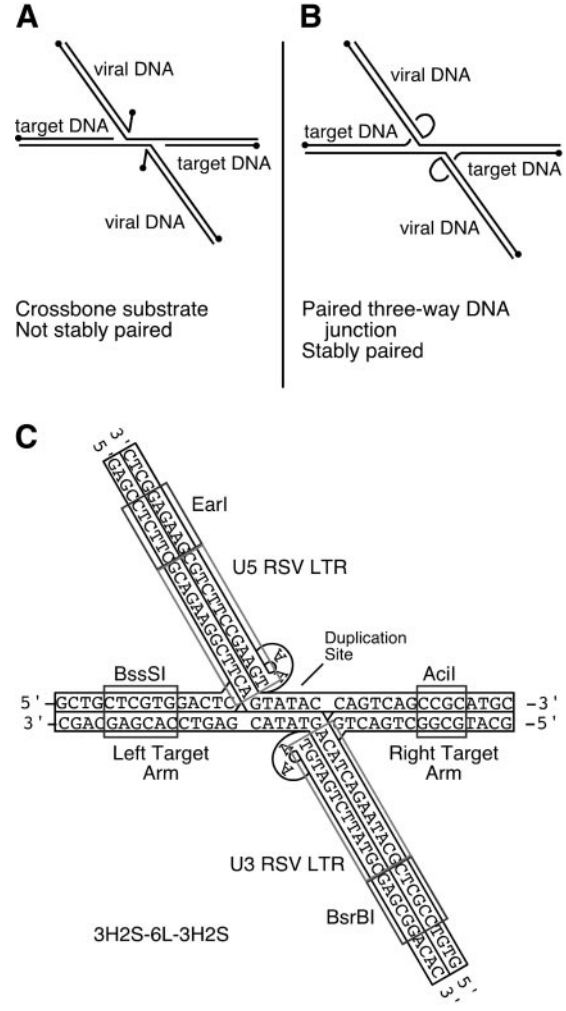

FIG. 2. Artificial integration intermediates. The two viral ends are joined to different target strands, but on the same side of the double helix, across the major groove. For RSV integrase, integrated viral strands are separated by 6 bases of host DNA. 5′ DNA strand ends are represented by a filled circle. (A) Artificial integration intermediate modeling integrase reaction product. (B) Paired three-way junction modeling an integration product. The 5′ overhanging dinucleotide is joined to the 3′ end of the adjacent target strand. (C) Paired three-way junction, 3H2S-6L-3H2S, modeling the integration intermediate. The light gray boxes indicate RSV LTR sequence. The dark gray boxes indicate restriction enzyme cleavage sites for use in characterizing the substrate.

ing that integrase protein–protein interactions can appose the two halves of this substrate to permit covalent chemistry.

In this study, we have constructed related DNA structures that are stable in the absence of integrase, with the goal of using such molecules to direct integrase binding to specific locations. To form stable DNA scaffolds, the 5′ overhanging nucleotides are joined to the 3′ ends of the target arms, forming a pair of three-way DNA junctions (Fig. 2B). We find that integrase binds specifically to correct three-way junction substrates and carries out hydrolysis at the junction between the viral cDNA and target DNA, allowing us to study the integrase–DNA complexes formed.

RESULTS

Design and nomenclature of paired DNA three-way junctions

We have constructed branched DNA molecules, double three-way junctions, in which two arms model the viral cDNA ends, and the other two model target DNA (Fig. 2B, sequence presented in Fig. 2C). The junction points are separated by 6 bp, the normal spacing in the target DNA between the points of joining of the RSV cDNA ends. In our model DNAs, the three-way junctions are formed (conceptually) by joining the 5′ overhanging dinucleotide in the viral DNA end to the 3′ end of the adjacent target strand. This has the advantage of stabilizing the fully base-paired DNA in solution, though we note it has the potential to distort the substrate out of the optimal conformation for binding or catalysis by integrase. We first discuss construction and assay of these substrates and then studies probing the optimal conformation.

We have devised a nomenclature for double three-way junction molecules that extends the naming system used previously for single three-way junctions (Welch *et al.,* 1995). The double three-way junction shown in Fig. 2C is termed 3H2S-6L-3H2S. "3H" refers to the three helices in a junction, and "2S" refers to the two-base single-stranded bulge as in Welch *et al.* (1995). The "6L" refers to the 6-bp linker between three-way junctions.

The 3H2S-6L-3H2S sequence includes 13 bp of the RSV-LTR on each of the two LTR arms, one matching the U5 LTR sequence, and the other matching U3 LTR sequence, as expected for the RSV integration intermediate (Katz *et al.,* 1982) (Fig. 2C). The sequence between the LTR arms was chosen to match the weak consensus derived from studies of avian integrase target sites (Fitzgerald and Grandgenett, 1994). Sequences at the tips of the four DNA arms contain restriction enzyme recognition sites for analysis of DNA duplex formation.

Assaying the conformation of 3H2S-6L-3H2S

The four DNA strands comprising 3H2S-6L-3H2S were synthesized as oligonucleotides and annealed by heating to 95°C and slow cooling to 4°C. A crucial point for further studies was that of whether the resulting products were homogenous double three-way junctions or heterogeneous partially single-stranded forms. Annealed DNA products were analyzed on native polyacrylamide gels and found to yield single predominant labeled bands (Fig. 3A), consistent with correct assembly. To test the extent of annealing more rigorously, the 3H2S-6L-3H2S substrate was end-labeled on each of the two target DNA arms and probed by cleavage with restriction enzymes (Figs. 3B and 3C). After cleavage, samples were denatured and separated by electrophoresis on DNA sequencing-type gels. For this experiment, the left target arm is one base shorter than the right target arm, allowing analysis of each in the same reaction mixture. The uncut sample yielded two closely spaced bands corresponding to the two labeled strands (compare Fig. 3C, lanes 1 and 2 with 3). Cleavage with restriction enzymes that cut in each of the four DNA arms yielded nearly complete digestion, indicative of correct duplex formation (Fig. 3C, lanes 4–7). Control experiments showed that the restriction enzymes used only cleave double-stranded DNA efficiently and not single-stranded DNA (data not shown). These studies indicated that the annealing procedure resulted in efficient assembly of each arm in the three-way junction.

These data do not address the conformation of the DNA at the three-way junction, which may well contain one or several melted base pairs. In addition, possible base pairing of the six complementary bases in the linker region was not directly assayed, and these bases may well not be stably base paired (discussed below). A further unknown is the DNA stacking interactions at the three-way junction, which will dictate the directions in which the DNA arms project from the junction region. Integrase is known to distort its DNA substrates upon binding, and DNA distortion in model substrates promotes catalysis by integrase. We therefore expect that the binding of integrase will result in local unpairing and potentially other types of DNA distortion at the three-way junction, so that effects of added integrase will largely dictate the DNA conformation.

Integrase-catalyzed hydrolysis on 3H2S-6L-3H2S

Integrase-catalyzed hydrolysis was tested as a first step in characterizing the integrase–DNA complexes formed on the paired three-way junctions. The 3H2S-6L-3H2S substrate was 5′ labeled on the U5 LTR end and incubated with RSV integrase. The reactivity of 3H2S-6L-3H2S was first compared to the previously reported crossbone integration substrate (Fig. 4A). The extent of cleavage, as measured by the release of the labeled LTR

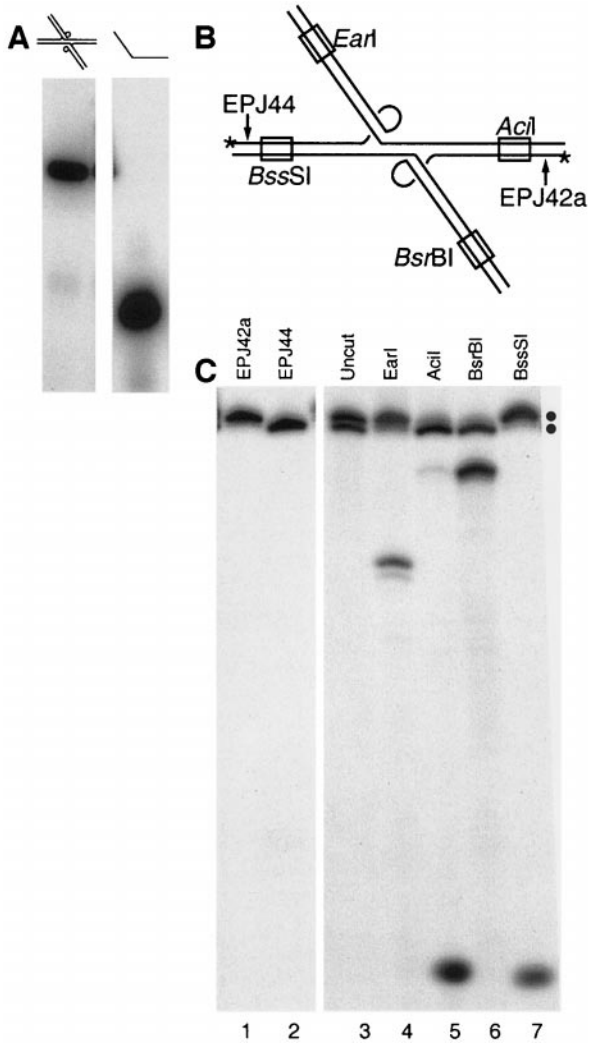

FIG. 3. Characterization of 3H2S-6L-3H2S substrate. (A) The 3H2S-6L-3H2S substrate was electrophoresed under nondenaturing conditions (lane 1). Single-stranded EPJ41 is also shown for comparison (lane 2). Diagrams of the substrate in each lane is shown above. (B) The 3H2S-6L-3H2S substrate used for restriction enzyme characterization. Oligonucleotides EPJ42a and EPJ44 are indicated by arrows, and are 5′ end labeled in this experiment. Positions of ^{32}P labeling are indicated by an asterisk. The positions of restriction enzyme cleavage sites are indicated by open boxes. (C) The 3H2S-6L-3H2S substrate was undigested (lane 3) or treated with *Ear*I (lane 4), *Aci*I (lane 5), *Bsr*BI (lane 6), or *Bss*SI (lane 7). Labeled oligonucleotides EPJ42a and EPJ44 were also loaded individually for comparison (lanes 1 and 2, respectively). The running positions of EPJ42a and EPJ44 are marked with closed circles. The minor band in lane 4 is likely due to the presence of −1 forms of the oligonucleotide. The minor band in lane 5 is probably a partial digestion product. The *Bsr*BI restriction site contains an *Aci*I site within it, so cleavage by *Aci*I at this site coupled with inefficient cleavage at the *Aci*I site on the right target arm will produce a fragment of the length seen for the extra band in lane 5. All restriction enzymes used are inactive on single-stranded DNA substrates.

fragment, was comparable between the two substrates (about 7% cleavage). This supports the idea that integrase is assembling on the 3H2S-6L-3H2S substrate with an efficiency similar to the crossbone substrate.

The activities in Mg^{2+} and Mn^{2+} were compared, since RSV integrase is known to be active in both (Bushman and Wang, 1994; Katz *et al.,* 1990; Katzman *et al.,* 1989). In the presence of Mn^{2+}, a product was generated of the size expected for integrase-mediated cleavage adjacent to the conserved CA dinucleotide (Fig. 4B). This is the site of hydrolysis by the normal terminal cleavage activity of integrase, though, in this case, taking place in the DNA three-way junction. Cleavage also occurred in the presence of Mg^{2+}, the metal cofactor expected to be employed *in vivo,* though to a lesser extent (Fig. 4B, lanes 6–8). In addition, hydrolysis was also carried out to a similar extent by RSV IN 49–286 (summarized in Fig. 5).

The correct LTR sequence is required for integrase-mediated hydrolysis

Several mutant versions of the 3H2S-6L-3H2S substrate were constructed to probe the features necessary for hydrolysis. Autoradiograms illustrating cleavage of substrates 5′ labeled on the U5 ends are shown in Fig. 4. Data for cleavage at all points in the DNA substrates (involving independent labeling of each DNA strand in the substrate) are summarized in Fig. 5. A mutant version of the substrate (mut3H2S-6L-3H2S) was constructed in which the last 6 bases of each viral DNA strand were replaced with nonspecific bases (Fig. 4C). This allows the importance of the correct terminal LTR sequences to be evaluated. No significant activity was seen, indicating that the cleavage activity is strictly dependent on the presence of the correct LTR sequence.

Correct spacing between three-way junctions is required for optimal cleavage

Another mutant was also constructed, 3H2S-11L-3H2S, in which 5 extra bases were added to the 6-bp linker region (Fig. 4D). The separation of the viral arms by a half-helical turn places them on opposite sides of the target DNA helix, allowing the importance of spacing to be assessed. Cleavage was observed in Mn^{2+} at the correct positions at the ends of the LTR DNA arms, though at least threefold weaker at each position than with the correct 3H2S-6L-3H2S cleavage. Additional abnormal cleavages were observed one turn downstream, opposite the other LTR DNA arm. Evidently, the addition of a half-helical turn between the two viral DNA ends disrupted the cleavage reaction, reducing the efficiency and leading to hydrolysis at ectopic sites. As a control, a substrate with altered spacing and mutated LTR sequences was also constructed (mut3H2S-11L-3H2S; Figs. 5G and 5H). For the mut3H2S-11L-3H2S substrate, normal cleavage was abolished by the mutation of the LTR DNA ends, but the aberrant downstream cleavage persisted, indicating that downstream cleavage is not dependent on the LTR sequence.

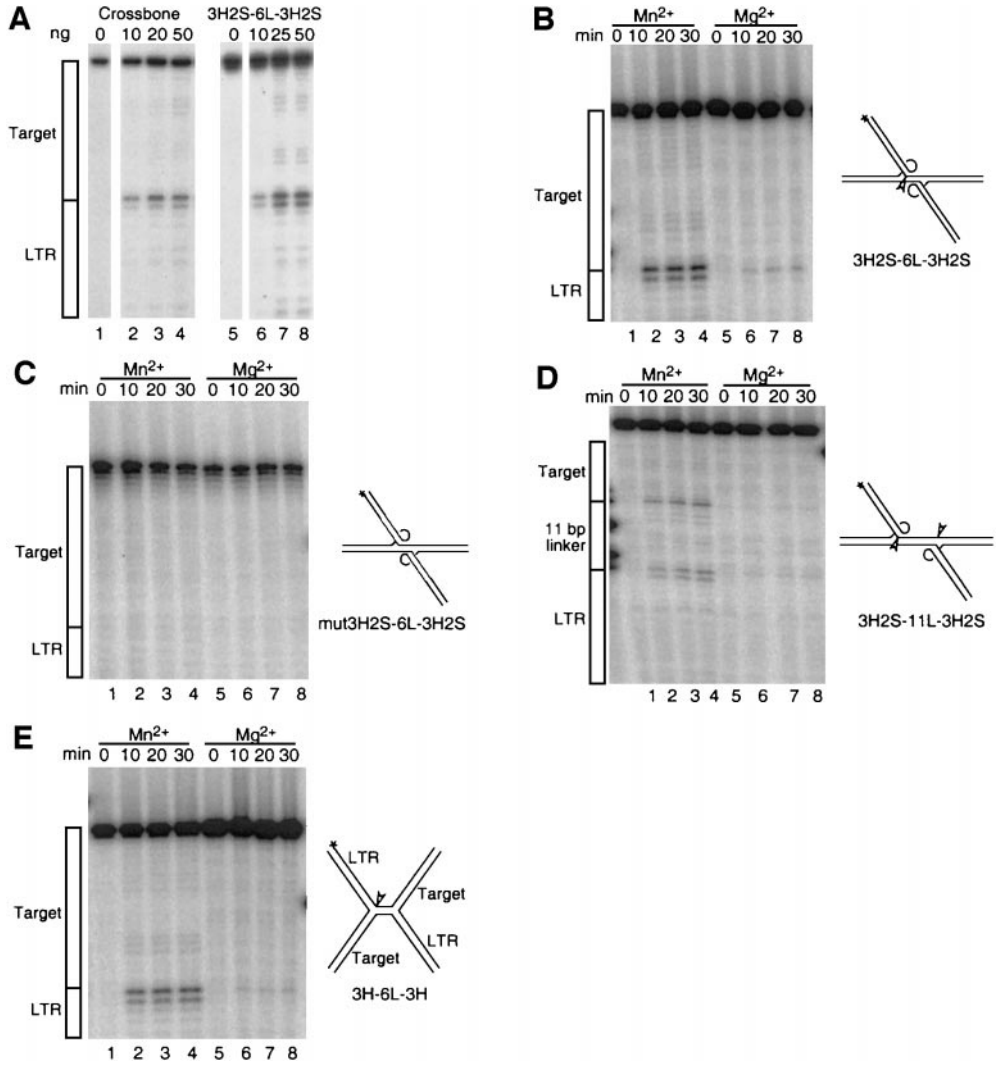

FIG. 4. Junction cleavage of the crossbone substrate, 3H2S-6L-3H2S, mut3H2S-6L-3H2S, 3H2S-11L-3H2S, and 3H-6L-3H with full-length RSV integrase. Substrates were labeled on the U5 viral arm. (A) Comparison of the cleavage of the crossbone substrate to cleavage of 3H2S-6L-3H2S. The concentrations of integrase added are indicated above the lane. All reactions were carried out for 30 min. (B) Time course of cleavage of 3H2S-6L-3H2S. (C) Cleavage of mut3H2S-6L-3H2S. (D) Cleavage of 3H2S-11L-3H2S. (E) Cleavage of the 3H-6L-3H substrate. A diagram of each substrate is shown adjacent to the panel. Reaction buffers contained Mn^{2+} (lanes 1–4) or Mg^{2+} (lanes 5–8), and incubation was performed at 37°C for 0, 10, 20, and 30 min. The locations of LTR and target sequences for 3H2S-6L-3H2S, mut3H2S-6L-3H2S, and 3H2S-6L-3H2S are shown at left. The locations of the LTR, 11-bp linker, and target sequences for the 3H2S-11L-3H2S substrate are also shown.

Optimizing the ectopic connection between the LTR 5′ end and the target DNA 3′ end

Several substrates were assayed to optimize the ectopic connection between the LTR 5′ end and the target DNA 3′ end at the paired three-way junction. The nature of the ectopic connection between the viral cDNA 5′ end and the target 3′ end could have several effects on substrate conformation, such as constraining rotation of the LTR arm around its long axis relative to the target DNA part of the substrate or affecting stacking of the DNA arms at the junction. The bulge region, while necessary for stable assembly of the paired three-way junctions, has the potential to interfere with integrase binding to the LTR end. In the 3H-6L-3H substrate, the 2-base bulge regions were removed (Figs. 4E, 5I,and 5J). The 3H-6L-3H three-way junction was cleaved with the same efficiency and specificity as with 3H2S-6L-3H2S, despite the removal of the 2-base bulge.

Substrates with 2-, 4-, or 6-base bulges were next compared (Fig. 6). All were cleaved by both RSV IN 49–286 and RSV integrase in the presence of both Mn^{2+} and Mg^{2+}. Cleavage efficiency increased with increasing bulge length when treated with RSV integrase in the presence of Mg^{2+} (Fig. 6. lanes 6, 12, and 18). Evidently. the longer bulge facilitated integrase-mediated hydrolysis.

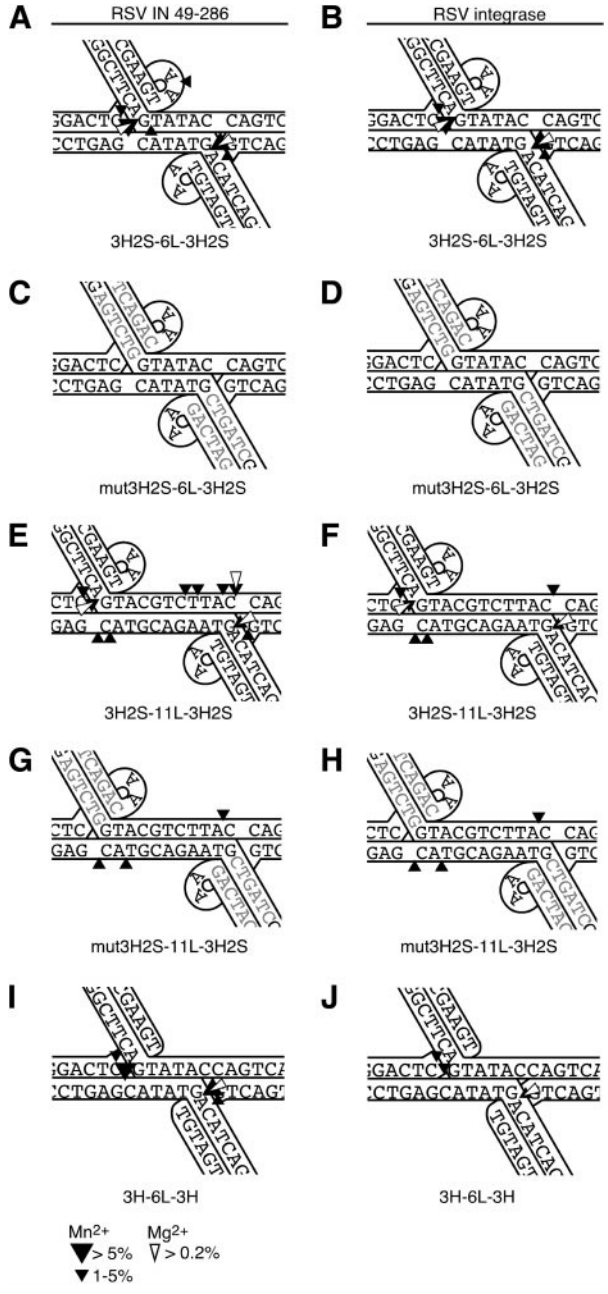

FIG. 5. Summary of three-way junction cleavage data. (A and B) Cleavage of 3H2S-6L-3H2S substrate. (C and D) Cleavage of mut3H2S-6L-3H2S. (E and F) Cleavage of 3H2S-11L-3H2S. (G and H) Cleavage of the mut3H2S-11L-3H2S substrate. (I and J) Cleavage of 3H-6L-3H substrate. Treatment by RSV IN 49–286 is shown in (A), (C), (E), (G) and (I). Treatment by full-length RSV integrase is shown in (B), (D), (F), (H) and (J). Cleavage of substrate is expressed in percent of substrate cleaved. A large filled triangle indicates a site at which >5% of the substrate is cleaved in the presence of Mn^{2+}, a small filled triangle indicates 1–5% cleavage in Mn^{2+}, and an open triangle indicates >0.2% cleavage in Mg^{2+}.

Assays of a substrate lacking base pairing in the linker region

As mentioned above, it is unknown whether the linker region in the 3H2S-6L-3H2S substrate is stably base paired, so, to probe this issue, a substrate incapable of base pairing was constructed and tested. The substrate, named 3H2S-6BL-3H2S for "bubbled linker," was cleaved with efficiency indistinguishable from the wild type (Fig. 7). Thus, integrase is equally capable of converting the substrate to a conformation suitable for hydrolysis regardless of whether or not the linker region is capable of base pairing. Whether or not the linker region in 3H2S-6L-3H2S is base paired normally is unknown.

DNase I footprint analysis of complexes

DNase I footprint experiments were performed to determine whether RSV integrase binds specifically to the double three-way junction substrates and to localize sites of protein–DNA interactions. The LTR DNA arms were lengthened in the substrates used to provide more DNA sequence on which to assess DNase I cleavage (indicated by "D" added to the name). The reaction mixture contained 5 mM $MgCl_2$ and 100 mM NaCl, conditions which allow cleavage by DNase I while permitting formation of tetrameric complexes of RSV integrase (Coleman *et al.,* 1999).

3H2S-6L-3H2S substrate DNAs were incubated on ice with full-length RSV integrase and then treated with DNase I at 37°C (Figs. 8A–8E). Substrates were labeled either on the U5 LTR arm (Figs. 8A and 8B) or the U3 LTR arm (Figs. 8C and 8D, diagrammed in 8E). The pattern of

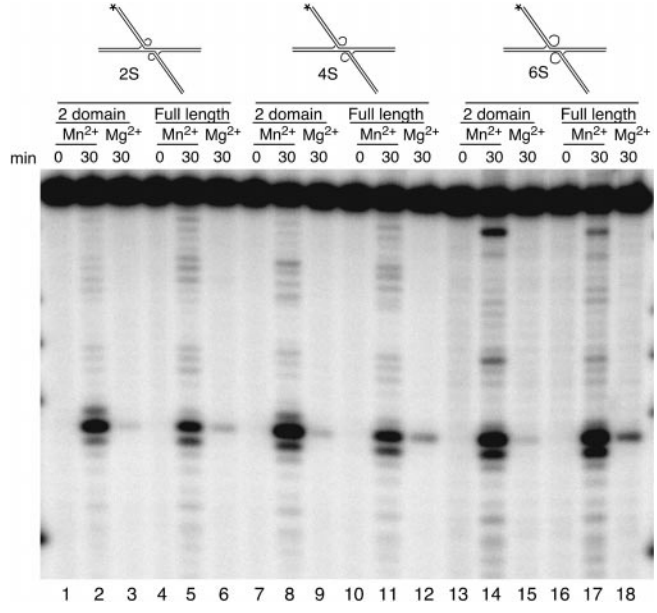

FIG. 6. Cleavage of 2-, 4-, and 6-base bulge substrates. (Lanes 1–6) Cleavage of 3H2S-6L-3H2S substrate. (Lanes 7–12) Cleavage of 3H4S-6L-3H4S substrate. (Lanes 13–18) Cleavage of 3H6S-6L-3H6S substrate. A diagram of each substrate is shown above the appropriate lanes. Substrates were labeled on U5 viral arm and were treated with RSV IN 49–286 (lanes 1–3, 7–9, and 13–15) or with full-length RSV integrase (lanes 4–6, 10–12, and 16–18). Reactions were completed in the presence of Mn^{2+} in all lanes except lanes 3, 6, 9, 12, 15, and 18, in which Mg^{2+} was used.

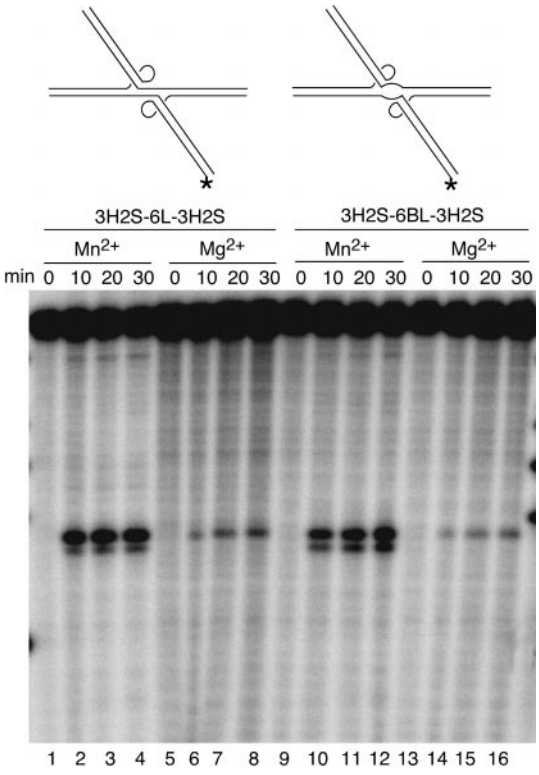

FIG. 7. Cleavage of 3H2S-6BL-3H2S, the "bubbled" substrate. Cleavage of 3H2S-6L-3H2S is shown in lanes 1–8. Cleavage of 3H2S-6BL-3H2S is shown in lanes 9–16. A diagram of each substrate is shown above the appropriate lanes. Substrates were labeled on the U3 arm. Reactions were carried out for increasing amounts of time in either Mn^{2+} (lanes 1–4 and 9–12) or Mg^{2+} (lanes 5–8 and 13–16).

DNase I digestion of the naked DNA substrate is shown in Fig. 8A, lane 1 and Fig. 8C, lane 1. RSV integrase was titrated from 40–140 ng per reaction. At 120 ng, 3H2S-6L-3H2S incubated with RSV integrase showed a pronounced reduction of cleavage at the last 5 bp of the LTR DNA (Figs. 8A and 8C, lane 6). A band appearing just above the protection is an integrase cleavage product, since this band appears in controls containing integrase but lacking DNase I (data not shown). Some reduction in cleavage is also apparent in the target region of the U3-labeled substrate 5 bases downstream of the U5 LTR end. Some reduction of overall cleavage was also observed, as expected for nonspecific binding. At higher concentrations of integrase (140 ng and higher), DNase I cleavage is apparent at all sites in the DNA, indicative of nonspecific coating of DNA by integrase. Assays of substrates labeled on the other DNA strand showed little cleavage by DNase I at the LTR-target junction, precluding analysis (data not shown).

Quantitation of autoradiograms (0 or 100 ng of RSV integrase) was carried out by PhosphorImager and ImageQuant software (Figs. 8B and 8D). The gray rectangles indicate regions of protection from DNase I cleavage at the ends of the LTR DNA. Similar results were observed when substrates were treated with RSV IN 49–286 (data not shown). As a control, DNase protection experiments were performed by using a linear substrate which modeled the viral end joined to target DNA (Fig. 9). No specific protection was observed when RSV integrase was incubated with these substrates. Nonspecific binding was seen at high concentrations as indicated by a nonspecific reduction of cleavage across the entire DNA substrate. Further studies of linear DNAs mimicking the end of the unintegrated cDNA (not joined to target) also did not show detectable specific protection (data not shown). These findings indicate that integrase binds specifically to the DNA three-way junctions at concentrations below those sufficient to coat the substrate.

DISCUSSION

Here, we present oligonucleotide scaffolds that organize RSV integrase complexes on DNA. Physical assays of integrase binding to DNA are typically dominated by nonspecific interactions, probably mimicking target DNA binding. In an effort to constrain the position of integrase on DNA, we have devised branched DNA structures mimicking a reaction intermediate. Normally, the action of RSV integrase yields an intermediate in which the 3′ ends of the viral cDNA are joined to points on the target spaced 6 bases apart on opposite strands (Fig. 2A). In the 3H2S-6L-3H2S substrate, the 5′ viral end is also joined to the 3′ target end to produce a pair of three-way DNA junctions with a 2-base bulge (Figs. 2B and 2C). The DNA strands in the structure are stably base paired under reaction conditions *in vitro,* unlike the authentic integration intermediate (crossbone substrate). The true integration intermediate DNA, if assembled by annealing from oligonucleotides, would be held together only by the 6 bp between the LTR DNA arms (Fig. 2A). Attempts to study integrase binding to such substrates were not successful, probably due at least in part to melting of the linker region (data not shown). We reasoned that stabilizing the DNA strands normally present in the integration intermediate might allow DNA contacts to constrain integrase at defined location on DNA.

Assays of integrase-mediated hydrolysis revealed that integrase could bind productively to the paired DNA three-way junctions, and that the substrate features mimicking the integration intermediate were required for optimal catalysis. The mut3H2S-6L-3H2S substrate lacking the LTR sequences was not cleaved, consistent with previous findings that mutations in the terminal LTR nucleotides greatly decrease activity on other types of substrates (Bushman and Craigie, 1992; Chow and Brown, 1994a; Gerton and Brown, 1997; Katzman and Katz, 1999; van den Ent *et al.,* 1994). The 3H2S-11L-3H2S junction, containing a lengthened linker region between three-way junctions, also showed disrupted cleavage. Weak cleavage was seen at the expected location at the end of the labeled viral DNA arm and also at a site one turn

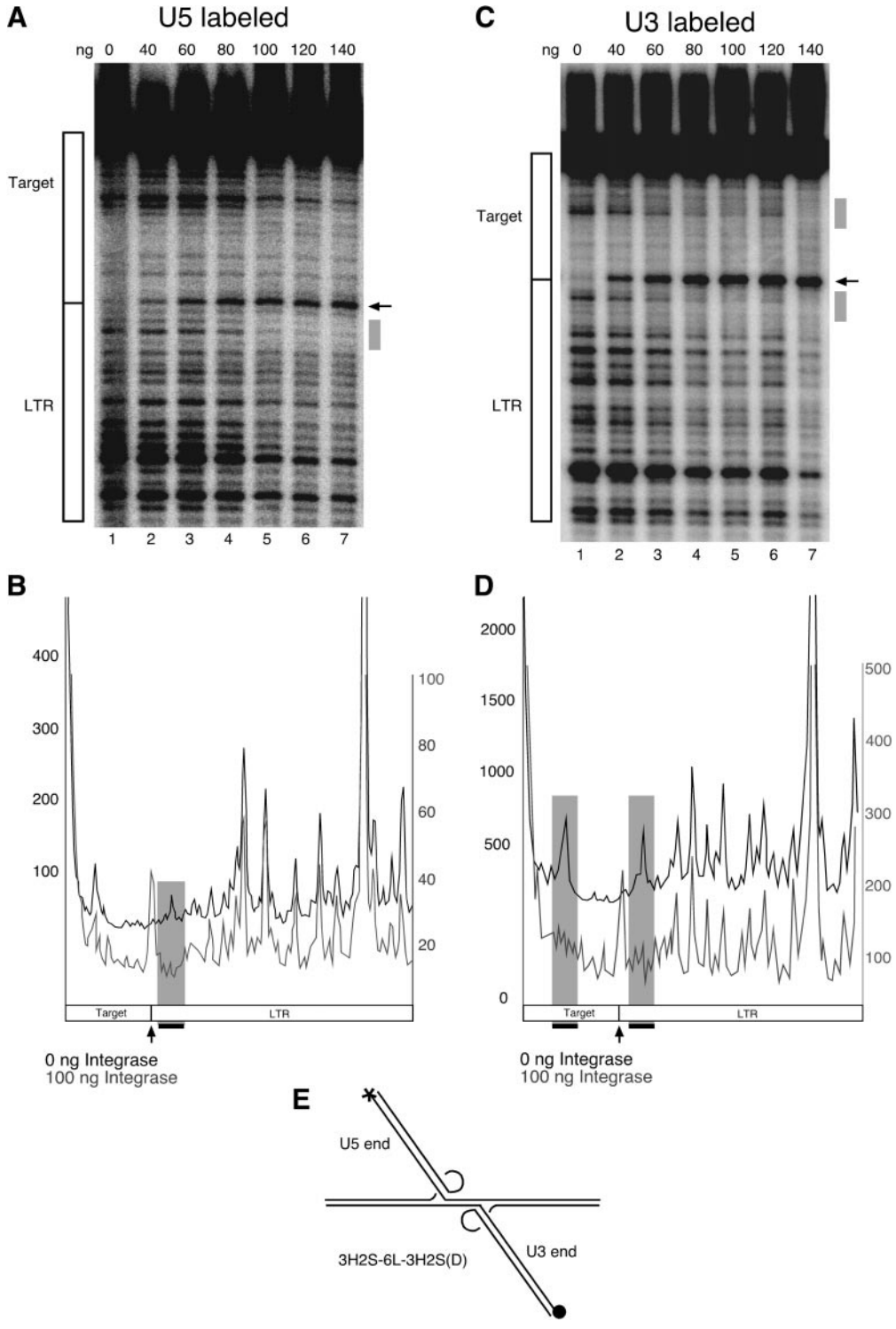

FIG. 8. DNase I footprinting of RSV integrase on three-way junction substrates. (A and C) Footprinting on the 3H2S-6L-3H2S substrate. (A) Footprinting on the U5 labeled substrate (indicated in E with an asterisk). (C) Footprinting on the U3 labeled substrate (indicated in E with a filled circle). The positions of the target and LTR DNA sequence are shown on the left. The location of protection by integrase is shown by the gray box on the right. Amounts of protein in ng are indicated. (B and D) Density plots from (A) and (C), respectively, lanes 1 and 5, showing DNase I products when the 3H2S-6L-3H2S substrate was preincubated with 0 or 100 ng of integrase. The upper plots show DNase I cleavage with 0 ng integrase (counts per minute indicated on the left), while the lower plots show cleavage in the presence of 100 ng integrase (counts per minute indicated on the right). The positions of the target and LTR DNA sequence is shown on the bottom. Vertical gray bars show the position of reduced cleavage by DNase I. The position of substrate cleavage by integrase is indicated by an arrow.

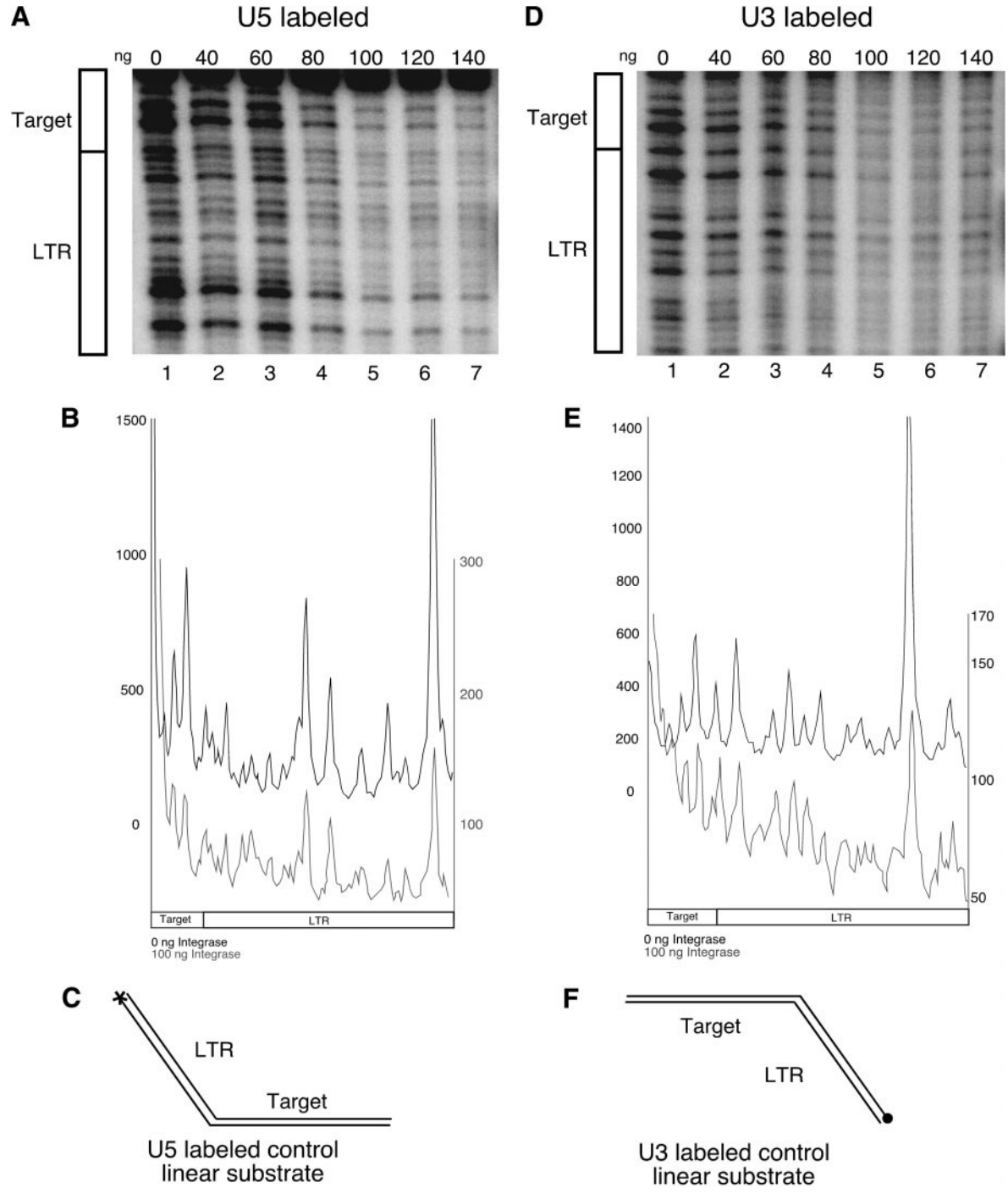

FIG. 9. DNase I footprinting of RSV integrase on linear substrates. (A and D) Footprinting on linear substrates matched to the 3H2S-6L-3H2S substrate. (A) Footprinting on the U5-labeled substrate (indicated in C with an asterisk). (D) Footprinting on the U3-labeled substrate (indicated in F with a filled circle). The positions of the target and LTR DNA sequence are shown on the left. Amounts of protein in ng are indicated. (B and E) Density plots from (A) and (D), respectively, lanes 1 and 5, showing DNase I products when the linear substrates were preincubated with 0 or 100 ng of integrase. The upper plots show DNase I cleavage with 0 ng integrase (counts per minute indicated on the left), while the lower plots show cleavage in the presence of 100 ng integrase (counts per minute indicated on the right). The positions of the target and LTR DNA sequence are shown on the bottom.

away at the distal three-way junction (Figs. 4D, 5E, and F). In assays with mut3H2S-11L-3H2S, only cleavage at the aberrant site occurs, indicating that, in this substrate, LTR sequence is important for normal LTR cleavage, but not for the ectopic downstream cleavage (Figs. 5G and 5H). These findings emphasize the importance of the correct DNA scaffold for organizing integration complexes on the paired three-way junctions.

The connection between the viral DNA 5′ end and the target DNA 3′ end, while necessary for stable assembly of the substrate, nevertheless introduces several complications. Increasing the length of the ectopic DNA bulge increased cleavage in Mg^{2+} (Fig. 6), probably because longer bulges minimize the constraints on DNA conformation. Several groups have reported that integration takes place preferentially on distorted target DNA (Bor *et al.,* 1995; Pruss *et al.,* 1994; Pryciak *et al.,* 1992a,b; Pryciak and Varmus, 1992) and that integrase preferentially cleaves frayed DNA sites (Bushman and Craigie, 1992; Katz *et al.,* 1998; Scottoline *et al.,* 1997). Perhaps distortion due to integrase binding brings the substrates into active conformations for cleavage, and the longer the bulge, the less energy required. Base pairs at the three-way junctions may also be unpaired in the absence of integrase.

Using DNase I footprinting, we observe specific binding by integrase to the last 5 bp of the LTR DNA arm (Fig. 8). At higher integrase concentrations, we also observed nonspecific binding. Models of RSV IN 49–286 binding to DNA have recently been proposed based on the structure of the RSV IN 49–286 dimer solved by x-ray crystallography (Yang *et al.,* 2000). In these models, a tetramer of integrase binds to the integration intermediate. The catalytic domain binds approximately 5 bp at the tip of the LTR, while the C-terminal domain binds to more distal LTR sequences. Crosslinking studies in the HIV system (Esposito and Craigie, 1998; Heuer and Brown, 1998; Gao *et al.,* in press) also implicate the catalytic domain in binding this region. Our data indicate protection of approximately 5–7 bp at the tip of the LTR, suggesting particularly strong binding by the catalytic domain to this region.

The ability to assemble integrase with paired three-way junctions may facilitate ongoing attempts to crystallize integrase bound to DNA, since the paired three-way junctions promote formation of integrase complexes. Stable complexes of integrase bound to three-way junctions may also allow contacts between integrase amino acids and substrate bases to be studied more readily.

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 polynucleotide kinase, and DNase I endonuclease were obtained from commercial suppliers and used according to the manufacturer's instructions. RSV integrase and RSV IN 49–286 were obtained from the laboratory of Dr. Craig Hyde, and were used as described below. The RSV IN 49–286 enzyme used in this study also contains a F199K mutation which functions to further solubilize the protein.

Construction and characterization of substrates

Table 1 lists oligonucleotides used in this study. All oligonucleotides were synthesized by IDT (Coralville, IA) and purified by standard means (Sambrook *et al.,* 1989). When necessary, oligonucleotides were radiolabeled by treatment with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP. For annealing of substrates, 1 μM of the labeled nucleotide and 1.2 μM of unlabeled nucleotides were mixed in a solution containing 100 mM NaCl and 5 mM $MgCl_2$ in TE buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA). Samples were then incubated at 95°C for 10 min, and cooled to 4°C over the course of 60 min.

To assay whether substrates were completely annealed, the 3H2S-6L-3H2S substrate was analyzed by electrophoresis on 5% polyacrylamide gels in Tris/Borate/EDTA buffer or on 8% DNA sequencing-type gels following cleavage with restriction enzymes (Sambrook *et al.,* 1989). Reaction products were visualized by autoradiography.

Integrase cleavage assay

Twenty nanograms of RSV integrase or RSV IN 49–286 were incubated with 0.01 μM of the appropriate substrate in a 10-μl final volume of cleavage buffer [10 mM Tris, pH 8.0, 2 mM β-mercaptoethanol (βME), and 100 mM NaCl] at 4°C for 30 min. Either $MgCl_2$ or $MnCl_2$ at 5 mM was then added to the samples which were then incubated at 37°C for 0 or 30 min. Reactions were then stopped by the addition of 1 μl 100 mM EDTA and 10 μl stop buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol in formamide). Reaction products were then analyzed by PAGE and autoradiography. Reactions were quantitated with ImageQuant version 1.2 software.

DNase I footprinting assay

DNase I footprinting assays were performed essentially as described (Galas and Schmitz, 1978). Substrates used in this study were 3H2S-6L-3H2S(D) and the linear substrate (Table 1). Full-length RSV integrase or the two domain RSV IN 49–286 in amounts ranging from 0 to 140 ng were mixed with 0.01 μM of the appropriate substrate in a final volume of 10 μl of footprinting buffer [10 mM Tris, pH 8.0, 2 mM βME, 100 mM NaCl, 5 mM $MgCl_2$, and 1 mg/ml bovine serum albumin (BSA)]. After incubation at 4°C for 30 min, 0 or 0.01 unit of DNase I was added in 1 μl of DNase solution (10 mM Tris, pH 8.0, 2 mM βME, 5 mM $MgCl_2$) and the sample incubated at 37°C for 10 min. Reactions were stopped by addition of 10 μl stop buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol in formamide) and analyzed by electrophoresis on a 6% denaturing gel and visualized using autoradiography (Sambrook *et al.,* 1989).

TABLE 1

Oligonucleotides Used in this Study

Oligo	Sequence	Comment
EPJ41	5′-GAGCCTCTTCGCAGAAGGCTTCAGTATACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target of 3H2S-6L-3H2S
EPJ42	5′-GCATGCGGCTGACTGAATGTAGTCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of 3H2S-6L-3H2S
EPJ43	5′-GTGTCCGCTCGCATAAGACTACAGTATACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target of 3H2S-6L-3H2S
EPJ44	5′-GCTGCTCGTGGACTCAATGAAGCCTTCTGCGAAGAGGCTC-3′	Left target strand and U5 strand of 3H2S-6L-3H2S
EPJ45	5′-GAGCCTCTTCGCAGAAGGCTTCACGGCGCCAGTCAGCCGCATGC-3′	U5 strand duplication site and right target of crossbone substrate
EPJ46	5′-GCATGCGGCTGACTG-3′	Right target strand of crossbone
EPJ47	5′-AATGTAGTCTTATGCGAGCGGACAC-3′	U3 strand of crossbone
EPJ48	5′-GTGTCCGCTCGCATAAGACTACAGCGCCGGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target arm of crossbone substrate
EPJ49	5′-GCTGCTCGTGGACTC-3′	Left target strand of crossbone
EPJ50	5′-CTCGGAGAAGCGTCTTCCGAAGTAA-3′	U5 arm of crossbone
EPJ64	5′-GAGCCTCTTCGCAGAAGAGTCTGGTATACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target of mutant 3H2S-6L-3H2S
EPJ65	5′-GCATGCGGCTGACTGAAGACTAGCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of mutant 3H2S-6L-3H2S
EPJ66	5′-GTGTCCGCTCGCATAAGCTAGTCGTATACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target of mutant 3H2S-6L-3H2S
EPJ67	5′-GCTGCTCGTGGACTCAACAGACTCTTCTGCGAAGAGGCTC-3′	Left target strand and U5 strand of mutant 3H2S-6L-3H2S
EPJ68	5′-GAGCCTCTTCGCAGAAGGCTTCAGTACGTCTTACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target strand of 3H2S-11L-3H2S
EPJ69	5′-GTGTCCGCTCGCATAAGACTACAGTAAGACGTACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target strand of 3H2S-11L-3H2S
EPJ115	5′-GAGCCTCTTCGCAGAAGAGTCTGGTACGTCTTACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target strand of mut3H2S-11L-3H2S
EPJ116	5′-GTGTCCGCTCGCATAAGCTAGTCGTAAGACGTACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target strand of mut3H2S-11L-3H2S
EPJ42a	5′-GCATGCGGCTGACTGGAATGTAGTCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of 3H2S-6L-3H2S for Fig. 3C
EPJ43a	5′-GTGTCCGCTCGCATAAGACTACATATACGAGTCCACGACAGC-3′	U3 strand, duplication site, and left target of 3H2S-6L-3H2S for Fig. 3C
EPJ420S	5′-GCATGCGGCTGACTGTGTAGTCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of 3H-6L-3H
EPJ440S	5′-GCTGCTCGTGGACTCTGAAGCCTTCTGCGAAGAGGCTC-3′	Left target strand and U5 strand of 3H-6L-3H
EPJ424S	5′-GCATGCGGCTGACTGAATTTGTAGTCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of 3H4S-6L-3H4S
EPJ426S	5′-GCATGCGGCTGACTGAATTTTTGTAGTCTTATGCGAGCGGACAC-3′	Right target strand and U3 strand of 3H6S-6L-3H6S
EPJ444S	5′-GCTGCTCGTGGACTCAATTTGAAGCCTTCTGCGAAGAGGCTC-3′	Left target strand and U5 strand of 3H4S-6L-3H4S
EPJ446S	5′-GCTGCTCGTGGACTCAATTTTTGAAGCCTTCTGCGAAGAGGCTC-3′	Left target strand and U5 strand of 3H6S-6L-3H6S
EPJ75	5′-CTCAGTCGAAGGACACGAAGTCCTGGACATGATGCTAAGTACTGCACTAC-GAGCCTCTTCGCAGAAGGCTTCAGTATACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target of 3H2S-6L-3H2S(D)
EPJ76	5′-GCATGCGGCTGACTGAATGTAGTCTTATGCGAGCGGACACCAGTCGACTCTG-CAGAATGCAGTGGTCTACTTGTGTACCGTACCTGATCG-3′	Right target strand and U3 strand of 3H2S-6L-3H2S(D)
EPJ77	5′-CGATCAGGTACGGTACACAAGTAGACCACTGCATTCTGCAGAGTCGACTGG-TGTCCGCTCGCATAAGACTACAGTATACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target of 3H2S-6L-3H2S(D)
EPJ78	5′-GCTGCTCGTGGACTCAATGAAGCCTTCTGCGAAGAGGCTCGTAGTGCAGTA-CTTAGCATCATGTCCAGGACTTCGTGTCCTTCGACTGAG-3′	Left target strand and U5 strand of 3H2S-6L-3H2S(D)
EPJ79	5′-CTCAGTCGAAGGACACGAAGTCCTGGACATGATGCTAAGTACTGCACT-ACGAGCCTCTTCGCAGAAGAGTCTGGTATACCAGTCAGCCGCATGC-3′	U5 strand, duplication site and right target of mutant 3H2S-6L-3H2S(D)
EPJ80	5′-GCATGCGGCTGACTGAAGACTAGCTTATGCGAGCGGACACCAGTCGA-CTCTGCAGAATGCAGTGGTCTACTTGTGTACCGTACCTGATCG-3′	Right target strand and U3 strand of mutant 3H2S-6L-3H2S(D)
EPJ81	5′-CGATCAGGTACGGTACACAAGTAGACCACTGCATTCTGCAGAGTCG-ACTGGTGTCCGCTCGCATAAGCTAGTCGTATACGAGTCCACGAGCAGC-3′	U3 strand, duplication site, and left target of mutant 3H2S-6L-3H2S(D)
EPJ82	5′-GCTGCTCGTGGACTCAACAGACTCTTCTGCGAAGAGGCTCGTAGTGC-AGTACTTAGCATCATGTCCAGGACTTCGTGTCCTTCGACTGAG-3′	Left target strand and U5 strand of mutant 3H2S-6L-3H2S(D)
EPJ87	5′-GCATGCGGCTGACTGGTATACTGAAGCCTTCTGCGAAGAGGCTCGTA-GTGCAGTACTTAGCATCATGTCCAGGACTTCGTGTCCTTCGACTGAG-3′	When annealed to EPJ75, forms linear substrate with U5 viral arm
EPJ93	5′-GCTGCTCGTGGACTCGTATACTGTAGTCTTATGCGAGCGGACACCAG-TCGACTCTGCAGAATGCAGTGGTCTACTTGTGTACCGTACCTGATCG-3′	When annealed to EPJ77, forms linear substrate with U3 viral arm
EPJ125	5′-GTGTCCGCTCGCATAAGACTACAACGCGTGAGTCCACGAGCAGC-3′	U3 strand, duplication site and left target arm of 3H2S-6BL-3H2S

ACKNOWLEDGMENTS

We thank the laboratory of Dr. Craig Hyde for the kind gift of purified integrase proteins, and Dr. Nadrian C. Seeman for helpful discussions. This work was supported by National Institutes of Health Grants GM56553 and AI34786 (to F.D.B.), the James B. Pendleton Charitable Trust and the Berger Foundation. F.D.B. is a Scholar of the Leukemia and Lymphoma Society of America. E.P.J. was supported by National Institutes of Health Research Service Award #1 F32 AI10165-01.

REFERENCES

Bor, Y.-C., Bushman, F., and Orgel, L. (1995). In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl. Acad. Sci. USA* **92,** 10334–10338.

Bor, Y.-C., Miller, M., Bushman, F., and Orgel, L. (1996). Target sequence preferences of HIV-1 integration complexes in vitro. *Virology* **222,** 238–242.

Brown, P. O. (1997). Integration. *In* "Retroviruses" (J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds.), pp. 161–203. Cold Spring Harbor Laboratory Press, Plainview, NY.

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1989). Retroviral integration: Structure of the initial covalent complex and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* **86,** 2525–2529.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995). High-resolution structure of the catalytic domain of avain sarcoma virus integrase. *J. Mol. Biol.* **253,** 333–346.

Bushman, F. D., and Craigie, R. (1992). Integration of human immunodeficiency virus DNA: Adduct interference analysis of required DNA sites. *Proc. Natl. Acad. Sci. USA* **89,** 3458–3462.

Bushman, F. D., Landau, N. R., and Emini, E. A. (1998). New developments in the biology and treatment of HIV. *Proc. Natl. Acad. Sci. USA* **95,** 11041–11042.

Bushman, F. D., and Wang, B. (1994). Rous sarcoma virus integrase protein: Mapping functions for catalysis and substrate binding. *J. Virol.* **68,** 2215–2223.

Bushman, F. D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249,** 1555–1558.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **4,** 567–577.

Carteau, S., Hoffmann, C., and Bushman, F. D. (1998). Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. *J. Virol.* **72,** 4005–4014.

Chen, J. C.-H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000a). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proc. Natl. Acad. Sci. USA* **97,** 8233–8238.

Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emini, E. A., Hazuda, D., and Kuo, L. C. (2000b). X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293): An initial glance of the viral DNA binding platform. *J. Mol. Biol.* **296,** 521–533.

Chow, S. A., and Brown, P. O. (1994a). Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J. Virol.* **68,** 3896–3907.

Chow, S. A., and Brown, P. O. (1994b). Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. *J. Virol.* **68,** 7869–7878.

Coleman, J., Eaton, S., Merkel, G., Skalka, A. M., and Laue, T. (1999). Characterization of the self association of *avian sarcoma* virus integrase by analytical ultracentrifuge. *J. Biol. Chem.* **274,** 32842–32846.

Craigie, R., Fujiwara, T., and Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. *Cell* **62,** 829–837.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **266,** 1981–1986.

Eijkelenboom, A. P. A. M., Puras Lutzke, R. A., Boelens, R., Plasterk, R. H. A., Kaptein, R., and Hard, K. (1995). The DNA binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* **2,** 807–810.

Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T., Plasterk, R. H. A., Kaptein, R., and Boelens, R. (1997). The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: A three-helix bundle stabilized by zinc. *Curr. Biol.* **7,** 739–746.

Engelman, A., Bushman, F. D., and Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12,** 3269–3275.

Engelman, A., Hickman, A. B., and Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* **68,** 5911–5917.

Esposito, D., and Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17,** 5832–5843.

Esposito, D., and Craigie, R. (1999). HIV integrase structure and function. *Adv. Virus Res.* **52,** 319–333.

Fitzgerald, M. L., and Grandgenett, D. P. (1994). Retroviral integration: In vitro host site selection by avian integrase. *J. Virol.* **68,** 4314–4321.

Fujiwara, T., and Mizuuchi, K. (1988). Retroviral DNA integration: Structure of an integration intermediate. *Cell* **54,** 497–504.

Galas, D. J., and Schmitz, A. (1978). DNAase footprinting: A simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5,** 3157–3170.

Gerton, J. L., and Brown, P. O. (1997). The core domain of HIV-1 integrase recognizes key features of its DNA substrates. *J. Biol. Chem.* **272,** 25809–25815.

Hansen, M. S. T., Carteau, S., Hoffmann, C., Li, L., and Bushman, F. (1998). Retroviral cDNA integration: Mechanism, applications and inhibition. *In* "Genetic Engineering: Principles and Methods" (J. K. Setlow, Ed.), Vol. 20, pp. 41–62. Plenum Press, New York and London.

Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287,** 646–650.

Hazuda, D. J., Wolfe, A. L., Hastings, J. C., Robbins, H. L., Graham, P. L., LaFemina, R. L., and Emini, E. A. (1994). Viral long terminal repeat substrate binding characteristics of the human immunodeficiency virus type 1 integrase. *J. Biol. Chem.* **269,** 3999–4004.

Heuer, T. S., and Brown, P. O. (1998). Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type-1 integrase-DNA complex. *Biochemistry* **37,** 6667–6678.

Katz, R. A., Gravuer, K., and Skalka, A. M. (1998). A preferred target DNA structure for retroviral integrase in vitro. *J. Biol. Chem.* **273,** 24190–24195.

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell* **63,** 87–95.

Katz, R. A., Omer, C. A., Weis, J. H., Mitsialis, A., Faras, A. J., and Guntaka, R. V. (1982). Restriction endonuclease and nucleotide sequence analysis of molecularly cloned unintegrated avian tumor virus DNA: Structure of large terminal repeats in circle junctions. *J. Virol.* **42,** 346–351.

Katzman, M., and Katz, R. A. (1999). Substrate recognition by retroviral integrases. *Adv. Virus Res.* **52,** 371–395.

Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (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.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34,** 9826–9833.

Mathe, G. (2000). The non-enumerable described retrovirus integrase inhibitors are not a lure, as evidenced by ten years of clinical experience. *Biomed. Pharmacother.* **54,** 3–6.

Mazumder, A., Engelman, A., Craigie, R., Fesen, M., and Pommier, Y. (1994). Intermolecular disintegration and intramolecular strand transfer activities of wild-type and mutant HIV-1 integrase. *Nucleic Acids Res.* **22,** 1037–1034.

Pemburton, I. K., Buckle, M., and Buc, H. (1996). The metal ion-induced cooperative binding of HIV-1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II). *J. Biol. Chem.* **271,** 1498–1506.

Pommier, Y., and Neamati, N. (1999). Inhibitors of human immunodeficiency virus integrase. *Adv. Virus Res.* **52,** 427–458.

Pruss, D., Bushman, F. D., and Wolffe, A. P. (1994). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* **91,** 5913–5917.

Pryciak, P., Muller, H.-P., and Varmus, H. E. (1992a). Simian virus 40 minichromosomes as targets for retroviral integration in vivo. *Proc. Natl. Acad. Sci. USA* **89,** 9237–9241.

Pryciak, P. M., Sil, A., and Varmus, H. E. (1992b). Retroviral integration into minichromosomes in vitro. *EMBO J.* **11,** 291–303.

Pryciak, P. M., and Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69,** 769–780.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning, A Laboratory Manual." Second Ed. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Schauer, M., and Billich, A. (1992). The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. *Biochem. Biophys. Res. Commun.* **185,** 874–880.

Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997). Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.* **11,** 371–382.

Sherman, P. A., and Fyfe, J. A. (1990). Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. USA* **87,** 5119–5123.

Thomas, M., and Brady, L. (1997). HIV integrase: A target for AIDS therapeutics. *Trends Biotechnol.* **15,** 167–172.

van den Ent, F. M., Vink, C., and Plasterk, R. H. A. (1994). DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. *J. Virol.* **68,** 7825–7832.

van Gent, D. C., Elgersma, Y., Bolk, M. W. J., Vink, C., and Plasterk, R. H. A. (1991). DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Res.* **19,** 3821–3827.

van Gent, D. C., Vink, C., Oude Groeneger, A. A. M., and Plasterk, R. H. A. (1993). Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12,** 3261–3267.

Vink, C., and Plasterk, R. H. A. (1993). The human immunodeficiency virus integrase protein. *Trends Genet.* **9,** 433–437.

Welch, J. B., Walter, F., and Lilley, D. M. J. (1995). Two inequivalent folding isomers of the three-way DNA junction with unpaired bases: Sequence-dependence of the folded conformation. *J. Mol. Biol.* **251,** 507–519.

Woerner, A. M., Klutch, M., Levin, J. G., and Marcus-Sekura, C. J. (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.

Yang, W., and Steitz, T. A. (1995). Recombining the structures of HIV integrase, RuvC, and RNAse H. *Curr. Biol.* **3,** 131–134.

Yang, Z. N., Mueser, T. C., Bushman, F. D., and Hyde, C. C. (2000). Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296,** 535–548.

Yi, J., Asante-Appiah, E., and Skalka, A. M. (1999). Divalent cations stimulate preferential recognition of a viral DNA end by HIV-1 integrase. *Biochemistry* **38,** 8458–8468.

Yoder, K. E., and Bushman, F. D. (2000). Repair of gaps in retroviral DNA integration intermediates. *J. Virol.* **74,** 11191–11200.

Yoshinaga, T., Kimura-Ohtani, Y., and Fujiwara, T. (1994). Detection and characterization of a functional complex of human immunodeficiency virus type 1 integrase and its DNA substrate by UV cross-linking. *J. Virol.* **68,** 5690–5697.