Presteady-state Analysis of Avian Sarcoma Virus Integrase

I. A SPliciNG ACTIVITY AND STRUCTURE-FUNCTION IMPLICATIONS FOR COGNATE SITE RECOGNITION

Received for publication, November 28, 2001, and in revised form, January 29, 2002
Published, JBC Papers in Press, January 30, 2002, DOI 10.1074/jbc.M111315200

Kogan K. Bao‡, Anna Marie Skalka§¶, and Isaac Wong‡¶

From the ‡Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331 and ¶Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Integrase catalyzes insertion of a retroviral genome into the host chromosome. After reverse transcription, integrase binds specifically to the ends of the duplex retroviral DNA, endonucleolytically cleaves two nucleotides from each 3'-end (the processing activity), and inserts these ends into the host DNA (the joining activity) in a concerted manner. In first-turnover experiments with synapsed DNA substrates, we observed a novel splicing activity that resembles an integrase joining reaction but uses unprocessed ends. This splicing reaction showed an initial exponential phase ($k_{\text{splicing}} = 0.02$ s$^{-1}$) of product formation and generated products macroscopically indistinguishable from those created by the processing and joining activities, thus bringing into question methods previously used to quantitate these reactions in a time regime where multiple turnovers of the enzyme have occurred. With a presteady-state assay, however, we were able to distinguish between different pathways that led to formation of identical products. Furthermore, the splicing reaction allowed characterization of substrate binding and specificity. Although integrase requires only a 3' hydroxyl with respect to nucleophiles derived from DNA, it specifically favors the cognate sequence CATT as the electrophile. These experimental results support a two-site "switching" model for binding and catalysis of all three integrase activities.

After infection, retroviruses create a linear DNA copy of their RNA genome that, through the strand-transfer mechanism of reverse transcriptase, places the U3 region of the LTR sequence at one terminus and the U5 region of the LTR sequence at the other terminus (1). Retroviral replication is dependent on the viral protein integrase catalyzing the recombination of the viral DNA genome into the host genomic DNA. Integrase binds to the two blunt-ended viral LTRs, hydrolyzes the terminal two nucleotides to expose a recessed 3'-OH of the conserved CA dinucleotide at each of the ends (the processing activity), and inserts these "processed" ends into the host DNA (the joining activity) at sites separated by a virus-specific stagger of six base pairs for avian sarcoma virus (ASV). The location of the insertion is nearly random as there is little sequence specificity for the site of recombination within the host genome (2–4). The processing and joining activities are biochemically similar in that both use a hydroxyl group as the nucleophile in an endonucleolytic cleavage. In the case of the processing activity (3'-dinucleotide removal), the enzyme is specific in its choice of electrophile (the cognate CATT), whereas the nucleophile (the processed CA-OH) is specified in the joining reaction (strand transfer). Both the ends-processing and joining activities have been reproduced in vitro with purified recombinant integrase and oligonucleotides whose sequences are derived from the retroviral U3 and U5 LTR sequences (5–7). Detailed examination of the processing activity in vitro has revealed that integrase requires the physiologically relevant configuration of both U3 and U5 ends to be bound concurrently for maximal efficiency of processing catalysis, with the U3 sequence undergoing both the cleavage and recombination reactions earlier than the U5 sequence (8).

Structural investigations of integrase suggest that the enzyme possesses three structural domains, 1) an N-terminal domain characterized by a zinc-stabilized helix-turn-helix, 2) a central core domain with a D,D(35)E motif, and 3) a C-terminal domain with structure resembling a Src homology 3 (SH-3) domain (for review, see Refs. 9–11). Although there is much evidence that active integrase functions as a multimer (12–15), it is still undetermined how the three domains interact with each other within a single integrase monomer or a multimer integrase-DNA complex. Wang et al. (16) have recently reported the structural solution of a two-domain fragment of human immunodeficiency virus-1 integrase that suggests a dimer of dimers resembling Tn5 transposase. However, the actual oligomeric state and geometric arrangement of integrase monomers within the active DNA-protein complex remain unconfirmed, and the detailed catalytic mechanism of the coordinated cleavage of four different DNA segments, resulting in the concerted insertion of the two LTR ends, remains unsolved. Two general models of the organization of the multimeric complex necessary to catalyze the insertion of the two ends of DNA have been proposed (17–19). In one model, separate integrase molecules bind each of the two processed viral DNA ends with two separate integrase molecules binding host DNA to prepare the two host phosphates for attack. In the other model, each integrase monomer contains separate binding sites for host and viral DNA, and a single active site catalyzes both processing and joining reactions, thus requiring the active integrase complex to only be a dimer. An unanswered question with both of the models, however, is the manner in which the LTR ends, which serve as electrophiles in the processing reaction, come to

The abbreviations used are: ASV, avian sarcoma virus; IN, integrase A260 and A280 absorbances at 260 nm and 280 nm, respectively; LTR, long terminal repeat; TBE, Tris-Borate-EDTA; PEI, polyethyleneimine.
reside in the active site as nucleophiles in the joining reaction (where the host DNA is the electrophile). This question becomes increasingly complex when one considers the possible combinations of nucleophiles recognized by integrase (20) and the fact that there must be specific and nonspecific DNA binding sites.

In the course of undertaking a presteady-state investigation of the processing reaction using synapsed-end substrates, we developed an assay that features a preincubation step to form DNA-protein complexes before the initiation of catalytic activity. Along with analysis of product formation during the first enzymatic turnover, the presteady-state assay enabled direct comparison of the reactivity of substrates while minimizing the complications involved in the quantitation of enzymatic activity. Using this assay for the processing reaction with synapsed-end substrates modeled after those of Kukolj and Skalka (5), we report here the discovery of a novel splicing reaction with products nearly indistinguishable from those of the processing and joining reactions. The identification of the splicing reaction resolved complications with the accurate quantitation of enzymatic activity. Additionally, the splicing reaction was used as a tool to gain significant insight into the structure-function relationship in the mechanism of sequence recognition by integrase. Specifically, the splicing reaction allowed us to probe the selection of nucleophiles and electrophiles by the structurally defined, sequence-specific binding sites of integrase. The results from these experiments allow us to propose a model for the configuration of the nucleophile and electrophile within the enzyme active site that satisfies the varying specificity requirements of the three integrase-catalyzed reactions.

For clarity, the remainder of this report will refer to the 3'-end dinucleotide TT-trimming activity as the "processing" reaction, the subsequent sequence-dependent insertion of processed ends into a double-stranded DNA target as the "joining" reaction, and this novel activity with synapsed substrates as the "splicing" reaction.

### EXPERIMENTAL PROCEDURES

**Reagents and Buffers—** Except where noted, all buffers were made with reagent grade chemicals and Milli-Q Plus (Millipore, Bedford, MA) purified distilled-deionized water. Urea, SDS, and dithiothreitol were Ultrapure grade obtained from U. S. Biochemical Corp. Kanamycin sulfate was obtained from Amresco (Solon, OH). EDTA, HEPES, and ammonium sulfate were SigmaUltra grade obtained from Sigma. EDTA, HEPES, and Ultrapure grade obtained from U. S. Biochemical Corp. Kanamycin with reagent grade chemicals and Milli-Q Plus (Millipore, Bedford, MA) was obtained from Sigma (St. Louis, MO). Restriction enzymes and T4 DNA ligase were obtained either from Fermentas (Hanover, MD) or Invitrogen. T4 polynucleotide kinase and acetylated bovine serum albumin were obtained from U. S. Biochemical Corp. Lysozyme was obtained from Sigma. ASV integrase was purified as described under "Purification of Integrase."

### Synthetic Oligodeoxyribonucleotides—
Oligodeoxyribonucleotides were synthesized by the Center for Gene Research and Biotechnology Central Services Laboratory (Oregon State University) and purified by denaturing PAGE (20 or 13% acrylamide, 8 m urea in TBE) as previously described (21). Reversed-polarity oligodeoxyribonucleotides were synthesized using 5'-β-cyanoethyl phosphoramidites (Glen Research, Sterling, VA). Concentrations were determined spectrophotometrically in Tris-buffered using the calculated extinction coefficients at 260 nm (22) listed in Table I. Radiolabeled oligodeoxyribonucleotides were 5'-end-labeled except where otherwise specified and are designated by an asterisk (*). 5'-Radiolabeled oligodeoxyribonucleotides were prepared using 3 units of T4 polynucleotide kinase, 70 μCi of 1γ-32P-ATP (Amer sham Biosciences), and 50 pmol of DNA in 10 μl using the manufacturer's reaction buffer. After incubation at 37 °C for 15 min, reactions were quenched by the addition of 10 μl of 500 mM EDTA. Complementary strands of DNA were then added to equimolar amounts and annealed by heating to 100 °C followed by slow cooling to room temperature over a period of 15–20 min. Kinase was removed by extraction with phenol: chloroform:isoamyl alcohol (25:24:1) followed by a chloroform-only extraction. Unincorporated nucleotides and residual organic material were removed by purification through a Bio-Spin-6 micro-column (Bio-Rad). Final yield and purity of radiolabeled substrates were determined by thin-layer liquid chromatography on PEI cellulose plates developed in a mobile phase of 0.75 M LiCl, 1.25 M formic acid, 40% ethanol. Under these conditions, labeled DNA remains at the origin, whereas unincorporated nucleotides are eluted by the mobile phase. Accurate quantitation of DNA yield is achieved by comparing the amount of radioactivity bound at the origin for samples obtained before and after purification steps. Typically, 87–90% of the DNA is recovered with complete removal of unincorporated nucleotides. In addition, comparison of the radioactivity in unincorporated label and DNA in the pre-purification sample shows that the efficiency of the labeling reaction exceeds 90% under these conditions.

The naming convention used for annealed DNA substrates is as follows. 1) Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and a "3," respectively, 2) strands of duplex DNA containing ASV integrase cognate sequence, CATT, are designated with an asterisk (*) to denote 5'-end radiolabeling.

### Table I

**Nomenclature and μM extinction coefficients at 260 nm DNA substrates**

| Name | Sequence | ε₂₆₀ |
|------|----------|------|
| 5t   | 5'-GCTGAGCAGAAGGTCCATT-3' | 0.20 |
| 5t-2 | 5'-GCTGAGCAGAAGGTCCATT-3' | 0.19 |
| 5b   | 5'-AATGAGCTTCTGGACTCACG-3' | 0.19 |
| 3t   | 5'-GCTATTGCAATTGACTGACTC-3' | 0.21 |
| 3b   | 5'-AATGAGCTTCTGGACTCACG-3' | 0.21 |
| m5t  | 5'-GCTGAGCAGAAGGTCCATT-3' | 0.21 |
| m3b  | 5'-TTTGGCTATATGCAATAGC-3' | 0.20 |
| 5b(23t) | 5'-GCTATTGCAATTGACTGACTC-3' | 0.43 |
| 5b(23b) | 3'-CGATAAGCTTTCTGATGTAAT-5' | 0.43 |
| 5b(23t) | 5'-GCTATTGCAATTGACTGACTC-3' | 0.41 |
| 5b(23t) | 5'-GCTATTGCAATTGACTGACTC-3' | 0.43 |

* The following is the naming convention used for annealed DNA substrates. Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and a "3," respectively. Strands containing the cognate sequence, CATT, are designated with a "t." Strands containing the complementary GTAA sequence are designated with a "b." Synapsed strands are designated with the length of the tether within parentheses. Duplex names consist of a concatenation of the names of all oligodeoxyribonucleotide strands annealed separated by slashes (/). Sequences labeled with 32P will be represented in the text with an asterisk (*) to denote 5'-end radiolabeling.
differences have been attributed to strain variance. Because of the possible importance of one of these differences, a glutamic acid at position 256 was altered to match the lysine of the reported sequence by PCR site-directed mutagenesis. The altered gene was subsequently subcloned into pET24a (Novagen Madison, WI) to produce the recombinant integrase (Novagen) (Sigma, P8849). Overexpression of integrase revealed a multi-gene construct was obtained in E. coli BL21 (DE3) (Novagen) by using isopropyl β-D-thiogalactopyranoside induction at 25 °C. Although recombinant integrase from either variant is biologically competent, in pre-steady state in vitro assays, integrase with Lys-256 produced 2-fold greater processing and splicing burst amplitudes (data not shown).

Protein Purification—E. coli BL21(DE3)/pET24a(IN) was grown and added to a quenching buffer (8 mM urea, 0.25 mM EDTA, 20% sucrose) in a 2:1 quench:reaction mix ratio. Generally, seven time points were taken in the first 30 s of the reaction with seven more time points in the ensuing 29 min. The reaction products were then analyzed by denaturing sequencing PAGE (20% acrylamide, 8 μm urea, TBE) using a Sequi-Gen GT 35 × 39 cm apparatus (Bio-Rad). Bands in the gel were visualized by using a Molecular Dynamics PhosphorImager (Amersham Biosciences), and the intensity of DNA bands was quantitated using ImageQuant software (Amersham Biosciences).

Quantitation—The intensity of each product band, $I(t)$, at each time, $t$, was first normalized with respect to the sum of intensities in the starting substrate band, $I_0$, plus all product bands according to Equation 1.

$$F_{corr}(t) = \frac{I(t)}{I_0}$$

$$F_{norm}(t) = \frac{\sum_{i=1}^{n} I(t)}{n} - \frac{F_{corr}(0)}{1 - \sum_{i=1}^{n} F_{norm}(0)} = \frac{1}{\sum_{i=1}^{n} F_{norm}(0)}$$

The number of exponential phases observable for any particular experiment increased with the length of the time regime under examination (8). Accordingly, the resulting time courses were fitted to Equation 3 consisting of $n$ exponential terms, with amplitudes $A_i$ and apparent rate constants $k_{norm,i}$, plus a linear term with an apparent rate constant $k_{lin}$ to fit the linear portion of the ensuing exponential phase.

$$y = \sum_{i=1}^{n} A_i (1 - e^{-k_{norm,i}t}) + k_{lin}t$$

Non-linear least squares fittings were performed using Kaleidograph software (Synergy, Redding, PA).

Preparation and Purification of Splicing Product for Maxam-Gilbert Sequencing—A 300-μl reaction was performed under standard conditions with 3 μM integrase and 3 μM [55S/52B/23S/3B or *51S-25B/23S/3B for oligodeoxyribonucleotide definitions see Table I] and allowed to proceed for 6.5 min. The reaction was stopped with quench solution, and the products were separated by denaturing sequencing PAGE in a 20% acrylamide, 8 μm urea gel. The product bands were recorded on Kodak X-Omat film, and the largest product band was excised from the gel and isolated with an Elutrap electroelution device (Schleicher & Schuell). Maxam-Gilbert sequencing was performed on the purified product DNA according to the published protocol (25).

Double-filter Binding—Nitrocellulose filter binding experiments were performed with only minor modifications to the double-filter method previously described (26). Standard assay buffer without Mn$^{2+}$ was used for binding studies. Integrase was first incubated on ice for 30 min with radiolabeled oligonucleotide substrates. After incubation, the solution was filtered through a combination of nitrocellulose (BA-583) and DEAE NA-45 or 4% DE-81 membranes (Schleicher & Schuell or Whatman, Maidstone, England) by using a 96-well Bio-Dot microfiltration apparatus (Bio-Rad). Immediately after filtration of the reaction solution, the wells of the apparatus were rinsed twice with an equal volume of ice-cold buffer. The filters were imaged by using a Molecular Dynamics PhosphorImager (Amersham Biosciences) and quantitated as described (26). For experiments using DE-81 membranes, the radioactivity of all three membranes was quantified and summed.

---

2 R. Katz, personal communication.
RESULTS

Presteady-state Assay Reveals Splicing of Synapsed Substrate by ASV Integrase—Fig. 1A shows results from a typical first-turnover processing activity assay at 0.5 μM integrase, 0.5 μM *5t/5b(2)3t/3b, 100 mM NaCl with samples subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel (A). Lanes represent reaction times of 0, 5.9, 11.1, 17.1, 22.6, 28.0, 34.1, 46.0, 60.0, 75.0, 90, 120, and 300 s. The amount of processing products was quantified using a PhosphorImager, and a plot of the results shown in B. The solid line represents the best fit of the data as described under “Experimental Procedures” with A = 5.9 ± 0.1 nM, λ = 0.20 ± 0.03 s⁻¹, and λ = 0.010 ± 0.001 nM/s.

Fig. 1. Processing reactivity at the terminal CATT cognate site on the 21-mer *5t of the dual-site synapsed substrate. First-turnover processing activity assay at 0.5 μM integrase, 0.5 μM *5t/5b(2)3t/3b, 100 mM NaCl with samples subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel (A). Lanes represent reaction times of 0, 5.9, 11.1, 17.1, 22.6, 28.0, 34.1, 46.0, 60.0, 75.0, 90, 120, and 300 s. The amount of processing products was quantified using a PhosphorImager, and a plot of the results shown in B. The solid line represents the best fit of the data as described under “Experimental Procedures” with A = 5.9 ± 0.1 nM, λ = 0.20 ± 0.03 s⁻¹, and λ = 0.010 ± 0.001 nM/s.

RESULTS

Presteady-state Assay Reveals Splicing of Synapsed Substrate by ASV Integrase—Fig. 1A shows results from a typical first-turnover experiment performed with a synapsed DNA substrate, *5t/5b(2)3t/3b, as detected by electrophoresis on an 8 M urea, 20% acrylamide sequencing gel. This substrate contains DNA sequences from the U3 and U5 ends of the viral LTR organized with the cognate CATT sites held together by a two-nucleotide single-stranded tether in a head-to-head configuration (5). The substrate was radiolabeled at 5t, the 21-mer DNA strand containing the U5 sequence. This 21-mer terminates at the 3'-end with the processing cognate sequence, CATT, which is endonucleolytically cleaved by integrase at the −2 position to exponentially yield an expected 19-mer as the major product at 0.2 s⁻¹ (Fig. 1B) plus a dinucleotide TT product, which is silent in this assay (27–29). The 5.9 nM burst amplitude reflects only 1.2% of total integrase monomers. Although the concentrations used (500 nM integrase, 500 nM DNA) were known to be not saturating, the propensity of integrase to aggregate at higher concentrations precluded experiments under saturating conditions (8). In addition, a minor 18-mer side product was also observed. Because the ratio between the two products was approximately constant, the 18-mer was likely produced in parallel to the 19-mer major product rather than from the degradation of the 19-mer. This 18-mer is consistent with “near miss” processing at the −3 position observed with Mn²⁺ as the metal cofactor as reported (28, 30).

Unexpectedly, a third product appeared that was larger than the starting 21-mer. Isolation from a semi-preparative scale assay reaction and sequencing (Fig. 2A) showed that this product was a 46-mer derived from the direct splicing of an unprocessed *5t 21-mer into a 5b(2)3t 44-mer strand at its internal cognate CATT site (Fig. 2C). Interestingly, the electrophilic site of splicing occurred at the expected processing position for the U3 cognate sequence, CATT, of 5b(2)3t.

To determine whether this splicing reaction maintains the same sequence specificity as the processing reaction or if the position of attack resulted from the opportunistic placement of the 3’-OH of *5t as a convenient nucleophile near the process-
more importantly, the rates of product formation of the initial exponential phase, representing the first turnover of catalytically competent pre-formed complexes, were comparable for both substrates despite the possible difference in the location of the nucleophiles relative to the site of splicing. The cognate sequence specificity of the electrophile binding site alone appears to be sufficient to determine the specificity of the splicing reaction.

Sequence Requirements of the Splicing Reaction—Two mutant single-site synapsed substrates were synthesized by changing either the U5 or U3 cognate CATT sequence to GCAA to examine nucleophile versus electrophile selectivity. The substrate *5t/m5b(2)/3t/3b contained a radiolabeled on the 21-nucleotide-long m5t strand, in which the terminal CATT sequence of the 5t strand was substituted with GCAA. A complementary mutation was made on the DNA strand of the substrate containing the tether, m5b(2)/3t, to maintain base pairing. The internal U3-derived cognate sequence, where splicing occurs, was unmodified. Alternatively, the substrate *5t/m5b(2)/3m/3m3b was synthesized by making the cognate to GCAA substitution at the internal cognate sequence, derived from the U3 LTR, while leaving the sequence at the U5-derived terminus unaltered.

Fig. 4 shows the results from assays, at low (100 mM, A) and high (400 mM, B) NaCl concentrations, with the unmodified dual-site substrate along with those of the two mutant single site substrates. The single-site substrate *5t/m5b(2)/3t/3b (squares), comprising a non-specific splicing nucleophile and a cognate splicing electrophile, was observed to have significantly higher reactivity than the conventional dual-site substrate *5t/m5b(2)/3t/3b (circles), resembling the non-specific splicing nucleophile and a non-specific splicing electrophile sequence, *5t/m5b(2)/3m/3t (diamonds), had barely measurable splicing activity. Fig. 4B also shows that the amounts of splicing product observed in the presteady-state assays are greatly increased at 400 mM NaCl. Although the rate constant for the initial appearance of these products was smaller, the actual rate of product formation when expressed as a mole of product formed/s was similar. This suggests that the observed reduction in rate constant may merely be a reflection of the increase in amplitude. Consistent with this result, preliminary experiments beyond the scope of this report indicate that the inhibitory effects due to the aggregation of integrase into higher assembly states can be partially reversed by the addition of salt (8).

Because the choice of nucleophile used in the splicing reaction can be CATT-OH, a processed CA-OH (Fig. 3), or the non-cognate GCAA-OH, these results showed that the nucleophile binding site on integrase is not sequence-specific. The cognate CATT sequence was, however, required at the electrophile site in order for efficient splicing to occur. Given the lack of discrimination of the enzyme regarding selection of a nucleophile in these reactions, the 5'-OH in spatial proximity to the splicing site was also examined for reactivity. This hydroxyl is at the 5' terminus of the 21-mer strand, 3b, annealed to the 44-mer 5b(2)/3t that contains the internal CATT splicing site. In an experiment where 3b* was 3'-endo-radiolabeled using α32P-TTP and wild type T7 DNA polymerase, no new bands representing any reaction products were observed under the conditions tested (data not shown). This result indicated that although any 3'-OH can act as a nucleophile, the 5'-OH is non-reactive.

Effect of Synapsed Ends on Equilibrium Binding—The equilibrium binding of inteigrase to these substrates was investigated by the nitrocellulose double-filter binding method (26). In addition to the usual dual-site synapsed substrate *5t/m5b(2)/3t/3b, the single-site 20-mer substrate *3t/3b and the single-site
synapsed substrate \(*m5t/m5b(2)3t/3b\) were examined, and the data were phenomenologically fitted to \(y = [IN]n/(K + [IN]n)\), which describes binding with infinite cooperativity to \(n\) sites.

The parameter \(n\), akin to the Hill coefficient, is a direct measure of the apparent cooperativity of binding and provides a lower limit for the minimal number of binding sites required for a fit. The two single-site substrates titrated identically despite their difference in length (Fig. 5). By comparison, the dual-site synapsed substrate showed both an increase in binding affinity and an increase in apparent cooperativity of binding as indicated by a shift to the left and an increase in the steepness of transition from \(n = 1.6 \pm 0.06\) to \(2.0 \pm 0.08\). In the absence of an independent measure of the assembly state of integrase, these titration data alone were insufficient to define a specific model for binding. However, these results were indicative of a change in binding modes by integrase due to the presence of two cognate sites on the DNA substrate. In contrast, when only a single cognate site was present on a synapsed substrate, binding was identical to that of two different substrate molecules despite a significant difference in their sizes.

**Intramolecular Versus Intermolecular**—Theoretically, the splicing reaction could have occurred either intermolecularly, between two different synapsed substrates, or intramolecularly, with the \(*5t\) strand splicing into the CATT of the \(5b(2)3t\) annealed to it (Fig. 6A). To address this issue, an integrase activity assay was performed using an equimolar solution of radiolabeled “preprocessed” \(*5t-2/5b(2)3t/3b\) and unlabeled full-length \(5t/5b(2)3t/3b\). The rationale for the assay stems from the assumption that both the \(5t\) and the \(5b(2)3t\) strands can serve as electrophiles because they both contain a cognate sequence CATT. Therefore, in an intermolecular splicing reaction, the 3′-OH from a given 19-mer \(*5t-2\) strand would have the choice of splicing into either the internal CATT of \(5b(2)3t\) (Fig. 6A, Intermolecular option I) or the terminal CATT-OH of \(5t\) (Fig. 6A, Intermolecular option II) to yield products with expected sizes of 44 and 21 nucleotides, respectively. Although a substantial amount of 44-mers was observed in the assay, no 21-mer products were detected (Fig. 6B).

To rule out the trivial possibility that the splicing reaction
could not occur at a terminal cognate site, a reversed-polarity substrate was synthesized utilizing a 5′-5′ phosphate linkage in the tether region. The substrate 5t-2 replaces the internal CATT of the synapsed substrate with a terminal CATT-OH (Fig. 7A). The preprocessed 19-mer 5t-2 strand was radiolabeled; splicing into the only available cognate site would extend this 19-mer to a 21-mer. Fig. 7B shows an exponential phase of product formation with product sizes ranging from 21 to 25 appearing within 20 s, a range of products smaller than the 19-mer starting material appearing after a lag of 1 min, and negligible amounts of products larger than a 25-mer even at the longest reaction time. The insignificant amounts of larger products are consistent with a lack of nonspecific joining by the radiolabeled preprocessed end. The delayed appearance of the smaller products was likely the consequence of sequential processing-joining reactions originating from the cognate site of the unlabeled 3t 21-mer. More importantly, the exponential appearance of 21–25-mer initially revealed the presence of splicing activity with this substrate. This demonstrated that the terminal CATT-OH functions as a viable splicing electrophile, although integrase apparently is unable to splice exclusively into the 5′-5′ position. In contrast, the reaction with equimolar labeled 5t-2/5b(2)3t/3b and unlabeled 5t/5b(2)3t/3t failed to produce any products in the 21- to 25-mer range. The lack of

**Fig. 6.** Intermolecular versus intramolecular splicing mechanisms. A, intermolecular splicing can occur either at an internal CATT (option I) or at a terminal CATT sequence (option II). In contrast, intramolecular splicing can only occur at an internal CATT sequence. B, single-turnover activity assay at 5.0 μM integrase, 0.5 μM 5t-2/5b(2)3t/3t, 400 mM NaCl with samples subjected to electrophoresis on a 20% polycrylamide, 8 M urea, TBE sequencing gel. Lanes represent reaction times of 0, 15, 30, 45, 60, 75, 90, 120, 240, and 360 s. The substrate was made by simultaneously annealing 5b(2)3t 1:1 with 3b, 1:0.5 with 5t, and 1:0.5 with 5t-2 to create an equimolar population of 5t-2/5b(2)3t/3t and 5b/5b(2)3t/3t. The radiolabel on the 19-mer 5t-2 strand allows detection of splicing products at either the terminal CATT-OH of 5t (intermolecular option I; expected product size of 21-mer) or the internal CATT of 5b(2)3t (intermolecular option II or intramolecular; expected product size of 44-mer). No 21-mer products above the background amount present at t = 0 were detected over time.

**Fig. 7.** Reactivity of a processed reversed-polarity substrate. A, a reversed-polarity synapsed substrate containing sequences derived from ASV U5 and U3 LTR sequences attached via a 2-nucleotide single-stranded tether with a 5′-5′ phosphate linkage (facing arrows). B, single-turnover activity assay at 5.0 μM integrase, 0.5 μM 5t-2/5b(2)3t/3t, 400 mM NaCl with samples subjected to electrophoresis on a 20% polycrylamide, 8 M urea, TBE sequencing gel. Lanes represent reaction times of 0, 5.8, 10.6, 15.3, 20.1, 25.7, 30.9, 45.3, 60, 120, 240, 600, 1200, and 1802 s.
splicing products characteristic of intermolecular option II (Fig. 6A) suggests an intramolecular mechanism of splicing.

**Processing Versus Splicing of the Internal Cognate Site**—Assays in which the synapsed 5b/2/3 t*-44-mer was radiolabeled yielded radiolabeled 19-mer product from either the splicing or the processing reactions since both are targeted at the identical phosphate bond between the A and T of the internal CATT cognate sequence. To determine the relative contributions of the two reactions to total 19-mer products observed, a first-turnover experiment was performed with 5t/5b/2/3t*/3b. The radiolabel was placed exclusively on the synapsed 5b(2)3t*/3b with respect to both rate of product formation and processing at 0.020 nM/s (A) and 0.20 nM/s (B). The amount of processing product; and 4) splicing occurs as an exponential phase with a characteristic rate constant of 0.020 s⁻¹, which is 10 times slower than processing (k_{processing} = 0.20 s⁻¹).

Although the splicing reaction appears to resemble the joining activity of integrase, the sequence specificity for a cognate CATT as the electrophile represents a significant mechanistic difference between the two. The integrase-catalyzed joining reaction shows very little sequence specificity with regard to target site selection (4, 31), although it does seem to show a structural preference, e.g., for bent DNA (32) or cruciform stem-loops (31). In contrast, the splicing activity of integrase is specific for the cognate sequence CATT such that splicing preferentially occurs between the A and T of this sequence. If the splicing reaction were representative of the true joining reaction, then selective integration at CATT sites in the host genome would be expected. This has not been observed. We have examined the sequence of putative "hot" sites (33) and were not able to correlate sites of integration with the existence of CATT sequences (data not shown).

On the other hand, the preference for a cognate CATT as the electrophile is shared in common with the processing reaction, the difference being that in the splicing reaction, a 3'-OH from DNA serves as the nucleophile, a role played by H₂O in the processing reaction. However, even in this aspect, the two reactions are similar. The promiscuity regarding the choice of nucleophiles in the processing reaction is well documented; non-water nucleophiles ranging from glycerol to serine and threonine (29) and even the 3'-OH of the leaving dinucleotide (18) have been observed. In the case of the splicing reaction, we have additionally shown that the DNA sequence attached 5' to the attacking 3'-OH is irrelevant. Thus, the splicing reaction may in fact represent a "rogue reaction" closely akin to processing, at least with regard to the specificity requirements of the electrophile and nucleophile binding sites.

An unexplained feature of the splicing reaction is the observation that no processing, as defined by H₂O cleavage, occurred at the internal CATT site. It would be expected that the internal CATT is a viable site for processing as previous studies (29) have shown that processing can occur at cognate sites with
Analysis of the exponential rates and amplitudes characteristic of the initial exponential phase, arising from both true processing and the rogue splicing reaction, provided the basis for the discovery of exclusive splicing activity at the internal CATT site. In contrast, previous experiments relying on measurements made in a time regime where the enzyme had undergone multiple exponential phases of product formation (5) erroneously attributed all the cleavage products at this site to a processing reaction and, as a result, incorrectly estimated the extent of processing seen with these substrates. Furthermore, under these assay conditions, the splicing reaction has almost twice the linear turnover rate when compared with the processing reaction despite having a similar initial exponential amplitude with a 10-fold slower exponential rate of product formation. Therefore, the net accumulation of splicing products after multiple exponential phases of product formation, mistakenly quantified as processing products, does not represent an accurate measure of the relative extent of the two reactions but may merely reflect the fact that the splicing product is bound less tightly by the enzyme than the processed product.

Although the design of the original synapsed-end substrate introduced complications with respect to quantitation of integrase activity, the binding data reported here do support the hypothesis that the tethering of two cognate sequences facilitates assembly of an integrase-DNA complex. These data show a higher level of cooperativity and tighter binding for the synapsed substrate versus either a single-ended untethered substrate (3t/3b) or a synapse substrate with only one cognate CATT present. In addition, the faster 0.2 s\(^{-1}\) burst rate constant for the processing reaction was observed only using substrates that contained two cognate sites. A more ideal design of synapsed substrates, therefore, requires both cognate CATTs to be situated at the 3′-termini of their respective strands. A 5′-5′ linkage in the tether segment, as found in the reverse polarity substrate, 5t/5b(2)/3b/3t. Mechanistic studies using this new class of synapsed substrates are presented in the accompanying paper of this series (8).

**Similarities to V(D)J Recombinases**—Interestingly, the splicing reaction of integrase closely resembles a reaction catalyzed by the V(D)J recombinases RAG1 and RAG2 (40–43). This site-specific recombination requires a 3′-OH at the end of the attacking strand for joining, uses this 3′-OH as the nucleophile in an intramolecular attack at a sequence specific site to form a hairpin, requires divalent metal ions, and is independent of ATP (42). The parallel observed between these two enzyme-catalyzed transesterification reactions provides the first functional evidence in support of the hypothesis that V(D)J recombinases belong to the same transposase structural superfamily that includes human immunodeficiency virus integrase, ASV-integrase, MuA (42), and would therefore predict a strong structural similarity between these enzymes and retroviral integrases.

**A Model for Functional Switching of Binding Sites**—The characterization of the splicing reaction also provides significant insights into the structure-function relationship of integrase. In particular, it highlights the important distinction between structurally defined binding sites such as the cognate sequence recognition site, where CATT is bound, versus functionally defined sites, such as the electrophile or the nucleophile binding sites. Our results show that in both the splicing and the processing reaction, only the electrophile is bound in the structurally fixed sequence-specific CATT binding site, whereas the nucleophilic 3′-OH occupies a nonspecific binding site. On the other hand, in the true joining reaction, this rela-
tionship between functional and structural sites is reversed. Namely, nonspecific DNA serves as the electrophile, whereas sequence-specific processed-end CA-OHs act as nucleophiles.

In an extreme case, this structure-function switch would require having different sets of active sites for the processing reaction and the joining reaction. In such a model, the processing would take place in a structurally coupled cognate site-specific electrophile binding site. After processing, the viral ends would then be transferred to a nearby cognate site-specific nucleophile binding site. In addition, joining would also require the binding of target host DNA in a third site, one where nonspecific DNA is bound as the electrophile. The host DNA cannot bind in the electrophile site used for processing, since it is cognate-sequence specific.

The current results suggest the possibility of a model that requires only two structurally fixed sites per retroviral DNA end, a sequence-specific site and a sequence-nonspecific site. In contrast to the more complex model, the functionalities of these two sites are not fixed but would switch during the course of catalysis. This is similar to classical acid-base mechanisms for phosphodiester hydrolysis proposed for ribonuclease (44) and alkaline phosphatase (45). In the initial processing state, the constellation of catalytic groups in the sequence-specific site is poised to activate the bound DNA as an electrophile, whereas the sequence-nonspecific site functions as a nucleophile-activating site. A switch in the functionality of the two sites accompanies the catalysis of the processing reaction. This switch obviates the need to translocate the processed DNA to a separate nucleophile-activating site. At the same time, the coupled switch would create an electrophile-activating site in the nontarget DNA in preparation for binding nonspecific target DNA.

The model makes chemical sense, as illustrated in Fig. 9 for a minimal constellation consisting of an acid-base pair. In this hypothetical scenario, a base in the nonspecific site activates the phosphate backbone of the bound CATT sequence as the electrophile. In the course of catalysis, this is generally poised to bind a cognate sequence as the target of nucleophilic attack and use any convenient functional hydroxyl, the spatially proximal 3'-OH of an unprocessed DNA strand in the case of the synapsed-end DNA molecules, as the nucleophile. The result is catalysis of the observed splicing reaction with these synapsed-end substrates.

Acknowledgments—We thank Dr. George Kukolj for insightful discussions, especially of his previous data and results, and Dr. Richard Katz for clarification of the sequence discrepancies between viral variants. We are grateful to Dr. Daniel Gamble, Victor Hsu, P. Shing Ho, C. Andrew Karplus, Christopher Mathews, and Michael Schimerlik for critical reading of the manuscript. K. K. B. also thanks A. Lucius and V. Tam for helpful discussions.

REFERENCES

1. Peliska, J. A., and Benkovics, S. J. (1992) Science 258, 1112–1118
2. Aiyar, A., Hindmarsh, P., Skalka, A. M., and Leis, J. (1996) J. Virol. 70, 3571–3580
3. Ellison, V., and Brown, P. O. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 7316–7320
4. Bor, Y. C., Miller, M. D., Bushman, F. D., and Orgel, E. L. (1996) Virology 222, 283–288
5. Kukolj, G., and Skalka, A. M. (1995) Genes Dev. 9, 2556–2567
6. Chow, S. A. (1997) Methods 12, 306–317
7. Vora, A. C., Chiu, R., McCord, M., Goodarzi, G., Stahl, S. J., Mueser, T. C., Hyde, C. C., and Grandgenett, D. P. (1997) J. Biol. Chem. 272, 23945–23954
8. Bao, K., Skalka, A. M., and Wong, I. (2002) J. Biol. Chem. 277, 12099–12108
9. Hindmarsh, P., and Leis, J. (1999) Microbiol. Mol. Biol. Rev. 63, 836–843
10. Andrade, M. K., and Skalka, A. M. (1996) J. Biol. Chem. 271, 19633–19636
11. Craigie, R. (2001) J. Biol. Chem. 276, 22310–22316
12. Deprez, E., Taeu, P., Leh, H., Mousseaud, J. F., Aufra, C., and Brochon, J. C. (2000) Biochemistry 39, 9275–9284
13. Jones, K. S., Coleman, J., Mulk, G. W., Leis, J., and Skalka, A. M. (1992) J. Biol. Chem. 267, 16937–16940
14. Vora, A., and Grandgenett, D. P. (2001) J. Virol. 75, 3556–3567
15. Pemberton, I. K., Buckle, M., and Buc, H. (1996) J. Biol. Chem. 271, 1498–1506
16. Wang, J., Ling, H., Yang, W., and Craigie, R. (2001) EMBO J. 20, 7333–7343
17. Katz, R. A., and Skalka, A. M. (1994) Annu. Rev. Biochem. 63, 153–173
18. Engelman, A., Mizuochi, K., and Craigie, R. (1991) Cell 67, 1211–1221
19. Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1996) J. Virol. 70, 425–437
20. Skinner, L. M., Sodol, M., Harper, A. L., and Katzman, M. (2001) J. Biol. Chem. 276, 114–124
21. Wong, I., Patel, S. S., and Johnson, K. A. (1991) Biochemistry 30, 526–537
22. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) Biopolymers 9, 1059–1077
23. Terry, R., Solis, D. A., Katzman, M., Cobrinik, D., Leis, J., and Skalka, A. M. (1988) J. Virol. 62, 2358–2365
24. Gill, S. C., and von Hippel, P. H. (1989) Anal. Biochem. 182, 319–326
25. Ausubel, F. M. (1988) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, New York
26. Wong, I., and Lohman, T. M. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 5428–5432
27. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 2525–2529
28. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) J. Virol. 63, 5319–5327
29. Vink, C., Yehezkely, E., van der Marel, G. A., van Boom, J. H., and Plasterk, R. H. (1991) Nucleic Acids Res. 19, 6691–6698
30. Katz, R. A., Mulk, G., Kalkosky, J., Leis, J., and Skalka, A. M. (1990) Cell 65, 87–95
31. Katz, R. A., Gravuer, K., and Skalka, A. M. (1998) J. Biol. Chem. 273, 24190–24195
32. Muller, H. P., and Varmus, H. E. (1984) EMBO J. 13, 4704–4714
33. Shih, C. C., Stuyt, J. P., and Coffin, J. M. (1988) Cell 53, 531–537
34. Katz, R. A., DiCandeler, P., Kukolj, G., and Skalka, A. M. (2001) J. Biol. Chem. 276, 34213–34220
35. Mueller, B., Jones, K. S., Mulk, G. W., and Skalka, A. M. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 11633–11637
36. Patel, S. S., Wong, I., and Johnson, K. A. (1991) Biochemistry 30, 511–525
37. Johnson, K. A. (1995) Methods Enzymol. 249, 38–61
38. Johnson, K. A. (1995) Methods Enzymol. 249, 3–37
39. Jegg, P. A., Taccioni, G. E., and Jackson, S. P. (1995) Bioessays 17, 949–957
40. Mele, M., Gellert, M., and van Gent, D. C. (1988) Science 290, 301–303
41. Roth, D. B., and Craig, N. L. (1988) Cell 54, 411–414
42. Ramsden, D. A., McBlane, J. F., van Gent, D. C., and Gelbert, M. (1996) EMBO J. 15, 3197–3202
43. Richards, F. M., and Wyrhoff, H. W. (1971) in The Enzymes (Boyer, P. D., ed) Vol. 4, 3rd Ed., pp. 847–860, Academic Press, Inc., New York
44. Reid, T., and Wilson, I. (1971) in The Enzymes (Boyer, P. D., ed) Vol. 4, pp. 373–416, Academic Press, Inc., New York