Presteady-state Analysis of Avian Sarcoma Virus Integrase

II. REVERSE-POLARITY SUBSTRATES IDENTIFY PREFERENTIAL PROCESSING OF THE U3-U5 PAIR*

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.M111314200

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

The integrase-catalyzed insertion of the retroviral genome into the host chromosome involves two reactions in vivo: 1) the binding and endonucleolytic removal of the terminal dinucleotides of the viral DNA termini and 2) the recombination of the ends with the host DNA. Kukolj and Skalka (Kukolj, G., and Skalka, A. M. (1995) Genes Dev. 9, 2556–2567) have previously shown that tethering of the termini enhances the endonucleolytic activities of integrase. We have used 5′-5′ phosphorothioate oligonucleotides to design reverse-polarity tethers that allowed us to examine the reactivity of two viral long terminal repeat-derived sequences when concurrently bound to integrase and, additionally, developed presteady-state assays to analyze the initial exponential phase of the reaction, which is a measure of the amount of productive nucleoprotein complexes formed during preincubation of integrase and DNA. Furthermore, the reverse-polarity tether circumvents the integrase-catalyzed splicing reaction (Bao, K., Skalka, A. M., and Wong, I. L. (2002) J. Biol. Chem. 277, 12089–12098) that obscures accurate analysis of the reactivities of synapsed DNA substrates. Consequently, we were able to establish a lower limit of 0.2 s⁻¹ for the rate constant of the processing reaction. The analysis showed the physiologically relevant U3/U5 pair of viral ends to be the preferred substrate for integrase with the U3/U3 combination favored over the U5/U5 pair.

The insertion of a DNA copy of the viral RNA genome into the host chromosome is a critical step in the reproduction cycle of retroviruses (1). Integrase catalyzes this reaction via a 2-step process, 1) the processing reaction, which is the recognition and endonucleolytic “trimming” of DNA sequences at the two 3′-ends of linear viral DNA and 2) the joining reaction, which is the concerted cleavage-ligation of the processed ends into the host chromosomal DNA. In avian sarcoma virus (ASV), the processing reaction produces site-specific cuts at the CATT sequence of the viral 3′-ends, removing the terminal TT dinucleotide to create new recessed 3′ OH ends. The two 3′ OH groups then serve as nucleophiles in the joining reaction to attack the phosphate bonds of the cellular target DNA in a single-step transesterification to produce a gapped covalent intermediate with 2-nucleotide overhangs. Overhang removal, gap fill-in, and ligation to complete the integration are likely mediated by host repair mechanisms (2), although participation by viral proteins has been suggested (3–5). The final product is a 4-base pair shortened viral genome inserted in the host DNA flanked by 6-base pair inverted repeats derived from the six base pairs separating the sites of concerted joining. The CA near the ends of the viral long terminal repeats (LTR) is conserved among retroviruses, whereas the length of the flanking host repeats is virus-specific and is thought to be a structural consequence of a particular integrase (6–9). For clarity, the remainder of this report will refer to the dinucleotide TT-trimming activity as the “processing” reaction and the subsequent sequence-dependent insertion of processed ends into a double-stranded DNA target as the “joining” reaction. The novel recombination activity specific to synapsed DNA substrates (10) will be referred to as the “splicing” reaction.

The U5 and U3 regions in ASV LTRs include two separate nearly perfect inverted repeat sequences in these noncoding sections of the retroviral RNA genome. In the course of retroviral replication, the RNA genome is reverse-transcribed into a duplex DNA copy, and as a result of the strand-transfer mechanism of reverse transcriptase, the U5 and U3 sequences become the termini of the DNA genome (11). Purified integrase along with Mn²⁺ or Mg²⁺ as a cofactor is sufficient to catalyze both processing and joining reactions in in vitro assays using synthetic oligodeoxynucleotide substrates with DNA sequences derived from the U3 and/or U5 viral LTR ends (12–14). Typical in vitro integrase processing assays use radiolabeled substrates, with the reaction(s) allowed to proceed for 30–90 min before quenching since integrase shows low reactivity in such assays (13, 15, 16). More recently, Vora and Grandgenett (17) studied the integrase-catalyzed joining reaction using an assay time of 5 min after a period of preincubation, aimed at “more closely reflecting the effects of the initial assembly events.” In either case, the reaction products and unreacted substrates are separated by electrophoresis on agarose or denaturing polyacrylamide gels to resolve the dinucleotide-shortened processing products and the extended joining products from the original substrates. Subsequent quantitation of the two products is used to measure the extent of catalytic activity. Single-end substrate assays show poor processing efficiency, and they also fail to produce measurable amounts of concerted integration products (where, in the case of ASV, two viral ends are inserted six base pairs apart) as is observed with preintegration complexes purified from infected cells (18, 19). Consequently, products larger than the original substrate in these...
assays that use a substrate containing only one LTR-end-derived sequence are considered to be products of only half-reactions, as only one viral DNA end is joined to a target site (19). Assays involving substrates designed to resemble the linear retroviral genome, with an LTR end-derived sequence at either terminus, have resulted in >95% of the joining products coming from the recombination of two separate substrate molecules integrating into a single target molecule (14, 20, 21). Despite the lack of similarity to the nucleoprotein complex assembled in vivo, the assays involving such trimolecular reactions have been used to suggest the preference of integrase for a U3/U3 combination of ends over that of a U5/U5 combination, with the specificity for the biologically significant U3/U5 arrangement intermediate between the two (14). More recently, Brin and Leis (22), using a reconstituted HIV-1 integration system, demonstrated that concerted DNA integration requires the presence of both U3 and U5 ends in the donor DNA. DNase I footprint analysis of the assembled integrase-DNA complex required for integration has revealed that the region of protection at the U3 LTR end (~20 base pairs) was at least twice that at the U5 LTR end (~10 base pairs), suggesting that the nucleoprotein complex is asymmetric when assembled in a fashion capable of full-site integration (17).

In an attempt to mimic the geometric organization of the viral LTR ends of the in vivo preintegration complex at a more molecular level, Kukolj and Skalka (12) designed a series of substrates that covalently linked two single-end substrates together in a head-to-head configuration using 1–3 nucleotides of single-stranded DNA (see Fig. 1A). It was hypothesized that these single-stranded tethers would provide sufficient flexibility to alleviate torsional or rotational strains arising from the structural alignment of the two viral ends bound within the integrase active site(s). By designing the substrates asymmetrically with respect to the length of the two ends and 5′ radiolabeling both ends, it was possible to quantitate processing products for both ends simultaneously in addition to what appeared to be extended joining products. Using in vitro integrase assays similar to those described above, these authors observed enhanced processing efficiencies with these synapsed-end substrates and concluded that the tether effectively brought together integrase subunits bound separately to the two cognate sites, thereby coordinating the formation of a requisite higher order oligomeric structure with enhanced activity. These observations were consistent with the suggestion by Murphy and Goff (23) that integrase must recognize both DNA end substrates and concluded that the tether effectively spliced reaction and consequently simplified the comparisons of the LTR ends. Analysis of presteady-state assays of these reverse-polarity substrates revealed that, although the U3 LTR sequence appears to be preferred by avian integrase when the termini are studied individually (25–28), the U3/U5 combination of retroviral ends is the preferred substrate for integrase microscopically within a single turnover. Results from the first-turnover exponential analyses also have important bearing on previous interpretations of multiple-turnover product distributions. An examination of the exponential phases and their usefulness to a further understanding of the mechanism of integrase activity and the binding of substrates is also discussed.

**EXPERIMENTAL PROCEDURES**

**Synthetic Oligodeoxynucleotides—Oligodeoxynucleotides** were synthesized by the Center for Gene Research and Biotechnology Central Services Laboratory (Oregon State University). Reversed-polarity oligodeoxynucleotides were synthesized using 5′-β-cyanomethyl phosphoramidites (Glen Research, Sterling, VA). Concentrations were determined spectrophotometrically in Tris-EDTA using the calculated extinction coefficients at 260 nm (29) listed in Table I.

The naming convention used for annealed DNA substrates is as follows. strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a “S” and “A,” respectively; strands of duplex DNA containing ASV integrase 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; duplexes are denoted as the concatenation of the names of the component single-stranded oligodeoxynucleotides separated by slashes (/); sequences 5′-end-radiolabeled with 32P will be specified in the text with an asterisk (*) at the beginning.

**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, \( I_d(t) \), plus all product bands according to Equation 1 to determine the normalized product fraction \( F_{i,norm}(t) \).

\[
F_{i,norm}(t) = \frac{I(t)}{I_d(t) + \sum_{i=1}^{n} I_i(t)}
\]

Equation 1

\( F_{i,norm} \) was then corrected for background intensity present at \( t = 0 \) for the \( i \)th band and renormalized for background intensities of all product bands to obtain the final corrected product fraction, \( F_{i,corr} \), according to Equation 2.

\[
F_{i,corr} = \frac{F_{i,norm}(t) - F_{i,norm}(0)}{1 - \sum_{i=1}^{n} F_{i,norm}(0)}
\]

Equation 2

Experimental time courses were fitted to Equation 3 consisting of \( n \) exponential terms, with amplitudes \( A_i \) and apparent rate constants \( \lambda_i \), plus a linear term with an apparent rate constant \( \lambda_{lin} \) to fit to the linear portion of the ensuing exponential phase.

\[
y = \sum_{i=1}^{n} A_i (1 - e^{-\lambda_i t}) + \lambda_{lin} t
\]

Equation 3

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

**Reagents, Buffers, Purification of Oligodeoxynucleotides, 5′-32P Labeling, Presteady-state Assays, Product Analysis by Denaturing Acrylamide Gel Electrophoresis, and ASV Integrase Overexpression and Purification**—These materials and protocols were identical to those described in detail in the first paper of this series (10).

**RESULTS**

**Effect of Reverse-polarity Tethers on Splicing Activity**—The normal-polarity substrate, used originally to develop the presteady-state assay for integrase (10), positioned the cognate CATT of the U3 sequence internally (see Fig. 1A). As a result, this sequence became the site of a spurious site-specific splicing reaction (10). Because the site preference of the splicing reaction is identical to that of the processing reaction, i.e. the internal CATT, its presence interfered with accurate quantita-
tion of enzymatic processing activity. A reverse-polarity substrate (Fig. 1B), incorporating a 5′-5′ reverse-polarity tether and containing the same sequence as that of the normal-polarity substrate (Table I), was therefore designed specifically to circumvent the splicing reaction while maintaining the advantages of tethering the viral LTR sequences (Fig. 1B).

At high concentration of NaCl (400 mM), the activity of integrase with normal-polarity substrates was predominated by the splicing activity (Fig. 2A). To demonstrate that the reverse-polarity substrate is not susceptible to this splicing reaction, presteady-state assays comparing different DNA substrates with and without a tether were performed at high salt conditions favorable to the splicing reaction. Reactions were performed as described under “Experimental Procedures” with all substrates radiolabeled on *5t, a 21-mer that is endonucleolytically cleaved by integrase at the normal processing, minus 2 position and at a second, minus 3 position to yield radiolabeled 19- and 18-mer products. The splicing reaction would yield a 46-mer product with the normal-polarity substrate *5/t/3b/*3t and an anticipated 23/24-mer with the reverse-polarity substrate *5t/5b(2)/3b/*3t.

Fig. 2A shows that the major product, at 400 mM NaCl, with the normal-polarity substrate was the 46-mer splicing product, whereas the 19-mer processing and shorter products were present in minor but detectable amounts. In contrast, reactions performed under identical conditions with the reverse-polarity substrate, *5t/5b(2)/5b/3t, yielded predominantly processing products (19- and 18-mers) (Fig. 2B). Integrase-mediated splicing activity of this reverse-polarity substrate would have resulted in +2 and +3 products of 23- and 24-nucleotide lengths, respectively; products of such sizes accumulated in minute amounts and only at the two longest reaction times examined (1200 and 1800 s). By comparison, the processing products appeared within 5–10 s and in much greater quantities. These data demonstrated that even under reaction conditions chosen specifically to favor splicing over processing, the 5′-5′ tethering preferentially mediates integrase-catalyzed processing.

To rule out the possibility that the increased NaCl concentration itself was capable of having promoted the initial exponential phase of processing activity with the reverse-polarity substrate, a control experiment was performed with an unsynapsed, single-ended substrate *5t/5b. Fig. 2C shows clearly the lack of any detectable integrase activity with this substrate under these conditions. Even at the longest time point assayed, 1800 s, *5t/5b was minimally processed. These results showed that without the scaffolding provided by the tethering of the two ends, integrase was unable to utilize the cognate sequence for enzymatic activity at this higher NaCl concentration.

**Effect of Tethering on Reactivity**—Fig. 3A shows the results from a typical first-turnover experiment performed as described under “Experimental Procedures.” Assay results with synapsed 5t/5b(2)/3b/*3t and unsynapsed *3t/3b substrates (see Table I) are shown to illustrate the effect of tethering two viral end sequences in a head-to-head manner via a 5′-5′ linkage. The radiolabeled 21-mer, *3t, with a sequence corresponding to the viral U3 LTR, used in both substrates, contained the 3′ terminal cognate sequence CATT-OH. In the processing reaction, this sequence was endonucleolytically cleaved by integrase at the minus 2 position to yield a 19-mer and a TT dinucleotide as the major products. Additionally, an 18-mer minor product was also observed that is consistent with the less specific processing activity at the minus 3 position observed when Mn²⁺ is used as the metal cofactor (19, 30, 31). Both 19- and 18-mer products were summed and plotted to quantitate total integrase-catalyzed processing activity.

Kinetic experiments revealed multi-phasic reactions at all integrase concentrations tested. The time-dependent appearance of the processing product was best described by a series of presteady-state exponential phases followed by a linear phase according to Equation 3. The number of exponential phases observed was dependent on both the substrate used and the duration of time course examined. For all time courses, we have chosen to report the model-independent, actual molar concentration of product observed without further interpretation for accuracy.

Compared with an unsynapsed substrate (*3t/3b), the initial exponential product formation phase of the synapsed substrate (5t/5b/5b/3b/*3t) reaction occurred with larger amplitude and a significantly faster rate. The difference in burst amplitudes and rate constants for the synapsed versus unsynapsed substrates directly reflected differences in the extent of productive complex formation between enzyme and DNA during the preincubation period (before the initiation of the reaction with Mn²⁺). Furthermore, the synapsed substrate contained two CATT-OH cognate sites at which processing could have occurred, one at the 3′-end of the 3t strand and one at the 3′-end of the 5t strand. Because only the 3t strand was radiolabeled, however, processing at the cognate site of the U5 sequence was silent in these assays. As a consequence, the amplitude measured for the synapsed substrate did not include the amount of productive complexes formed at the U5 cognate site, and the large difference in amplitudes observed in this experiment actually underestimated the true enhancement of productive nucleoprotein complex formation attributable to the tethering of the two viral LTR ends. At time points extended beyond the early exponential phases, both time courses paralleled each other for the synapsed and unsynapsed substrates, indicating similar rates of reaction in subsequent steps of the enzyme.

Experiments performed using radiolabeled *5t showed similar but lower reactivity under the same conditions.

To assess the effect of enzyme concentration on the differences in reactivities of the two substrates, titrations of enzyme concentration were performed for both substrates. Typical results are shown in Figs. 3, B and C, for assays conducted at integrase concentrations varying from 0.5 to 5.0 µM and a substrate concentration of 0.5 µM. The data were fit to Equation 3, and the best fit amplitudes and rate constants of the initial phase were plotted as a function of integrase concentrations. In the case of *3t/3b (closed squares, solid lines), both the apparent rate constant, λ₁, and amplitude, A₁, increased linearly with increasing integrase concentration with slopes of 0.0287 (r = 0.99) and 0.0049 (r = 0.99), respectively. In contrast, although the amplitude of the initial exponential phase in the synapsed substrate reaction increased with increasing
protein concentration (with the exception of the highest protein concentration where protein aggregation becomes a problem), the rate constant for the initial exponential phase remained 0.2 s⁻¹, independent of integrase concentration for the range of concentrations tested. In contrast, the single-end substrate only approached this rate at the highest integrase concentrations tested. In contrast, the single-end substrate s the rate constant for the initial exponential phase remained 0.2 s⁻¹, independent of integrase concentration for the range of concentrations tested. In contrast, the single-end substrate only approached this rate at the highest integrase concentration range. These data show that the lower limit of the rate constant for the processing reaction is 0.2 s⁻¹.

Effect of U3 and U5 Sequences on Processing and Joining—The reverse-polarity substrates allowed exclusive examination of the processing activity and, therefore, made possible the direct comparison of the reactivities of the U5 and U3 sequences. The reactivities of all four possible combinations of synapsed end sequences, U3/U3, U3/U5, U5/U5, and U5/U3, were separately measured using *3t/3b(2)5b/*3t, 5t/5b(2)3b/3t). At 3 μM integrase, 0.5 μM reverse-polarity substrate DNA, and 130 mM NaCl, the best fits of the time-dependent appearance of processing products of the U3 end in the context of a synapsed U3/U3 pair (open circles, *3t/3b(2)3b/*3t) versus that of the same U3 sequence in a synapsed U3/U5 combination (closed circles, 5t/5b(2)3b/*3t). At 3 μM integrase, 0.5 μM reverse-polarity substrate DNA, and 130 mM NaCl, the best fits of the time-dependent appearance of processing products were characterized by two initial exponential phases followed by a slower linear phase representing the beginning of a third exponential phase.

The results at 400 mM NaCl were optimal for demonstrating the preference in favor of U3 processing is an intrinsic property of the U3 sequence and to verify the results obtained at lower NaCl/integrase concentrations, additional experiments were performed to compare the reverse-polarity substrates containing either two synapsed U3 ends, *3t/*3t, or two synapsed U5 ends, *5t/*5t. As expected, Fig. 5B shows that the substrate which contained two synapsed U3 sequences (closed circles, solid line) was processed to a greater extent than was the substrate that contained two synapsed U5 sequences (closed squares, dashed line). However, to compare these results to those in Fig. 5A, it was necessary to sum the best fits from the data of the U3/U5 (5t/5b(2)3b/*3t) and U3/U3 (5t/5b(2)3b/*3t) substrates to determine the total reactivity of the U3 and U5 termini when they are concurrently bound. This adjustment was necessary because the radiolabeling of these substrates at the cognate sites in decreasing order is: U3 when synapsed with U3, U3 when synapsed with U5, U5 when synapsed with another U3, and U5 when synapsed with another U5. These results are similar to those obtained from data at lower NaCl/integrase concentrations (Fig. 4); however, the reactions at higher NaCl/integrase concentrations displayed increased reactivities for all substrates and enhanced the differences between the LTR ends.

The results at 400 mM NaCl were optimal for demonstrating

| Name | Sequence | ε<sub>260</sub> |
|------|----------|--------------|
| 5t   | 5′-GCTGAAAGCAAGGCTCTATT-3′ | 0.20 |
| 5b   | 5′-AATGAAAGCTCTGCTTCAACG-3′ | 0.19 |
| 3t   | 5′-GCTATTGCTAAAGCTATT-3′ | 0.21 |
| 3b   | 5′-AATGGTATTCTTATGCAACTG-3′ | 0.21 |
| 5b/3b | 5′-GCTTATGCTAAAGCTATT-3′ | 0.43 |
| 3b/5b | 3′-GCTTATGCTAAAGCTATT-3′ | 0.43 |
| 3b/3b | 3′-GCTTATGCTAAAGCTATT-3′ | 0.41 |

* 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 “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.
these preferences; however, similar results, albeit less dramatic, were observed at all NaCl conditions examined. Fig. 6 shows polyacrylamide gels of reaction aliquots from assays performed at 130 mM NaCl with the U3/U5 (5t/5b(2)3b/*3t, A) and U5/U3 (*5t/5b(2)3b3t, B) substrates. As expected, the U3 end was processed to a greater extent relative to the U5 end. Interestingly, a ladder of bands (indicated by a “J” in Fig. 6) larger than the starting material appeared after the initial appearance of the processing products. These larger products were present in a broad distribution of sizes, suggesting that they were joining products based on the observation that the integrase-catalyzed joining reaction lacks much sequence specificity (20, 33, 34). Unfortunately, the distribution of the radioactive label over a broad range of product sizes prevented accurate quantitation of the individual bands. However, a qualitative visual inspection of the gels revealed the rapid appearance of U3 joining products within 15 s and a delayed onset of U5 joining products for 120 s even though the formation of both U3 and U5 processing products occurred within the first 5–10 s of the reaction. These results extend previous reports that integrase favors U3 over U5 (14, 25) and illustrate the importance of having both U3 and U5 sequences present for the correct assembly of an integrase-LTR productive complex. This suggests that after processing, the joining activity of integrase is ordered such that the U3 end of the viral LTR is inserted prior to the U5 LTR.

DISCUSSION

In this work, we report the design, synthesis, and use of reverse-polarity synapsed substrates in presteady-state analysis of the ASV integrase reaction mechanism. Both unsynapsed and reverse-polarity synapsed substrates displayed multiple exponential phases leading to the initial formation of the dinucleotide-shortened processing product. Analysis of the earliest/fastest exponential phase revealed that the reverse-polarity tethering of viral LTR ends led to a faster rate and greater amplitude, consistent with an increase of productive complex formation. In addition to the benefits of a faster exponential rate and greater amplitude, the reverse-polarity substrate enabled comparative examination of combinations of LTR sequences when concurrently bound to integrase. The results showed integrase to have a preference for an asymmetric pair of ends (U3/U5). The mechanistic implications of these findings will be discussed.

First-turnover Kinetics of Integrase-mediated Processing—In presteady-state analyses, the enzyme is treated as a reactant and is used at concentrations comparable with or in excess over that of the substrate. Under first-turnover conditions, experiments are performed with enzyme in excess over substrate to allow the direct observation of the conversion of substrates to intermediates and products through a single catalytic cycle along the reaction pathway to the release of product. Preincubation circumvents the effect of the substrate binding and product release rates from the initial appearance of product and allows measurement of the slowest chemistry or conformational change step. In contrast, steady-state kinetics can define only the rate of conversion of bulk substrate to product as a function of a catalytic quantity of enzyme. Because the products measured are time-averaged over many catalytic turnovers, in steady-state experiments it is generally difficult to resolve intermediates and, thus, the chronology of events at the catalytic site.

Because of solubility and aggregation difficulties with the enzyme in the presence of DNA, it was not possible to determine the reaction stoichiometry in active site titration experiments (for an example of such experiments, see Ref. 35). Accordingly, we were unable to unambiguously identify the number of exponential phases corresponding to a complete single turnover event, and it was not possible to predict the expected total concentration of bound complexes in a single turnover.
The lack of active site titration data also introduced ambiguities that made direct mechanistic information unobtainable. It was not possible to resolve whether the multiple exponentials represented different serial steps occurring at different rates, different conformations of the enzyme-substrate complex operating in parallel with different reactivities, or the interconversion between active and non-active complexes. The processing reaction can be fitted to a single exponential phase followed by a linear phase, which represents the initial portion of the subsequent exponential phase only within the first 120 s of the reaction coordinate. For reaction times greater than this, a given product begins to become the time-averaged products of the reaction coordinate. For reaction times greater than this, the exponential was used to compare the relative reactivities of the DNA substrates as it has the smallest number of additional reactions complicating the quantitation of product.

The reaction quench mixture contained both EDTA to chelate the metal cofactor and urea to denature the enzyme, stopping further flux of DNA through the reaction pathway and releasing all DNA bound to the enzyme. As such, the signal measured was not ambiguous and reflected directly the amount of DNA hydrolyzed by the enzyme. The product providing the signal to measure this fastest phase is the physiologically relevant processing product, and thus, its rate of $0.2 \text{ s}^{-1}$ represents the lower limit estimate of the rate of the chemistry step for integrase-catalyzed processing. Both the rate of reaction and amplitude of the fastest phase for single-site substrates approach those of the synapsed substrate at higher concentrations of integrase, suggesting that this first exponential step lies on the same mechanistic pathway for both substrates and is not unique to the synapsed substrates.

Amplitude—The need for a large excess of integrase to obtain a significant initial exponential phase is unfortunate but consistent with the following known DNA binding properties of integrase. 1) In nitrocellulose filter binding experiments, a minimum of a 10-fold excess of integrase over DNA is required.
to achieve saturating conditions (10), and 2) both processing and joining activities of integrase require a multimeric form of the enzyme (17, 24, 36) with up to 11 monomers (16) required per LTR end. The typical range of initial processing exponential amplitudes was on the order of 0.5–12% of the total DNA in the range of integrase concentrations examined. Unfortunately, this amplitude is too small to be useful in a detailed mechanistic study. To examine the possibility that the small amplitude size was due to inactive protein, we have compared the amplitude size from integrase purified from different preparations, different clones, and different enzyme purification protocols as well as protein preparations from different laboratories (using different vectors, different clones, and different purification protocols). In all cases, the amplitude size was similar and reproducible (data not shown). The only differences observed in integrase activity, as measured by a change in the size of the amplitude of the first exponential, was upon the inclusion of detergent during the preparation of the enzyme.

One explanation for the low amplitude is the formation of nonproductive aggregates. In addition to the observed low solubility of the enzyme (12, 32, 37), we and others have observed that although purified integrase exists as monomers, dimers, and tetramers, large aggregates/multimers do form upon the introduction of DNA at integrase concentrations above 20 nM (Ref. 17 and data not shown). Furthermore, we have examined DNA substrates of varying lengths and found that the processing exponential amplitude size decreases with increasing substrate length, suggesting that noncognate DNA sequences facilitate the formation of nonproductive integrase-DNA aggregates. In the integrase titration experiments (Fig. 3B), we suspect the suppressed amplitude at the highest concentration of integrase to have been due to nonproductive protein aggregation in the presence of the longer DNA strands of the synapsed substrate.

Another possible explanation for the small amplitude size is the reported dependence of integrase processing activity on the disruption of the terminal base pairs of the substrate (15, 38). If the processing activity were dependent on DNA distortion, the amplitude size would be a reflection of the subpopulation of productive integrase-DNA complexes in which the substrate DNA ends are frayed. Although the absolute sizes of the amplitudes were not useful, the relative amplitudes permitted examination of the relative reactivity of substrates.

**Effect of Reverse-polarity Tethering on Integrase-mediated Processing**—Although the oligomeric state of functionally active integrase is not known, previous reports suggest that integrase functions minimally as a dimer (17, 24, 36, 39, 40) or a dimer of dimers (40). The presteady-state data are consistent with integrase functioning as a multimer in its active form. The initial apparent exponential rate measured for the ends-processing of the reverse-polarity synapsed substrate was 0.2 s⁻¹ and remained constant throughout the integrase concentration range tested (0.5–5.0 μM). In contrast, the exponential rate measured for the unsynapsed substrate increased linearly from 0.14 to 0.14 s⁻¹ with increasing integrase over the same concentration range. Additionally, although the exponential amplitudes of both synapsed and unsynapsed substrates increased monotonically with increasing enzyme concentration, the size of the amplitudes was larger for the synapsed substrate for all but the highest integrase concentration. These results indicate that tethering of the LTR ends facilitates increased formation of productive integrase-substrate complexes during the preincubation period relative to the separate individual LTR ends. Furthermore, the correlation between

---

**Fig. 4. Processing reactivity of integrase with different combinations of LTR ends.** Single-turnover splicing activity assays at 5 μM integrase, 0.5 μM DNA substrate, 130 mM NaCl with samples subjected to electrophoresis in a 20% polyacrylamide, 8 M urea, TBE sequencing gel. A, a plot of the results from substrates radiolabeled at the *3t* strand shows increased processing for the asymmetric U3/U5 (5t/5b/3b/3t, closed circles) combination over that of the U3/U3 (*3t/*3t, open circles) after the best fit of the U3/U3 data has been multiplied by 0.5 (dotted curve) to allow for accurate comparison. The lines represents the best fit of the data to Equation 3 with $A_1 = 0.0106 \pm 0.0005 \mu M$, $A_2 = 0.19 \pm 0.02 s^{-1}$, $A_0 = 0.037 \pm 0.003 \mu M$, $A_2 = 0.0085 \pm 0.0099 s^{-1}$, $A_3 = (7.0 \pm 0.5) \times 10^{-5} s^{-1}$ (dashed line) and $A_1 = 0.0067 \pm 0.0005 \mu M$, $A_2 = 0.19 \pm 0.02 s^{-1}$, $A_0 = 0.016 \pm 0.002 \mu M$, $A_2 = 0.011 \pm 0.002 s^{-1}$, $A_3 = (8.9 \pm 0.4) \times 10^{-5} s^{-1}$ (solid line). B, a plot of the results from substrates radiolabeled at the *5t* strand shows increased processing for the asymmetric U5/U3 (*5t/5b/23b/3t, closed squares) combination over that of the U5/U5 (*5t/*5t, open circles) after the best fit of the U5/U5 data has been multiplied by 0.5 (dotted curve) to allow for accurate comparison. The lines represents the best fit of the data to Equation 3 with $A_1 = 0.0020 \pm 0.0003 \mu M$, $A_2 = 0.19 \pm 0.02 s^{-1}$, $A_3 = 0.013 \pm 0.004 \mu M$, $A_0 = 0.006 \pm 0.002 s^{-1}$, $A_3 = (4.8 \pm 0.7) \times 10^{-5} s^{-1}$ (dashed line) and $A_1 = 0.0017 \pm 0.0002 \mu M$, $A_2 = 0.19 \pm 0.02 s^{-1}$, $A_3 = 0.0063 \pm 0.0006 \mu M$, $A_0 = 0.012 \pm 0.002 s^{-1}$, $A_3 = (7.8 \pm 0.1) \times 10^{-5}$ (solid line).
creasing protein concentration and increasing activity suggests the assembly of a higher ordered active integrase multimer.

In addition to the processing and joining activities catalyzed by integrase, substrates that tether together two viral LTRs are also susceptible to a splicing reaction that produces a specifically sized product (10). The "rogue" splicing and physiologically relevant joining activities differ in that the latter generates products in a range of sizes due to the nonspecific nature of joining site selection. Although the synapsing of the two ends does enhance integrase catalytic activity, the susceptibility of synapsed substrates to the splicing activity complicates accurate quantitation of the extent of the other two reactions (10, 11).

**Fig. 5.** Processing reactivity of integrase with different combinations of LTR ends at high NaCl concentrations. Single-turnover splicing activity assays at 20 μM integrase, 0.5 μM DNA substrate, 400 mM NaCl with samples subjected to electrophoresis in a 20% polyacrylamide, 8 M urea, TBE sequencing gel. A, a plot of the results from substrates with both U3 and U5 ends present shows increased processing at the U3 end, 5′/5b/2′3b/3t (closed circles) over the U5 end 5′/5b/2′3b/3t (closed squares). The lines represents the best fit of the data to Equation 3 with $A = 0.057 \pm 0.001$, $\lambda_{mw} = (7.7 \pm 0.4) \times 10^{-5}$ s$^{-1}$ (solid line) and $A = 0.027 \pm 0.001$, $\lambda_{ml} = 0.030 \pm 0.001$ s$^{-1}$ (dashed line). B, a plot of the results with substrates with either U3 or U5 ends shows increased processing for a substrate with only the U3 sequence, 3′/3b/2′3b/3t (closed circles), over a substrate with only the U5 sequence, 5′/5b/2′3b/3t (closed squares). The curves represent best fit of the data to Equation 3 with $A = 0.046 \pm 0.001$, $\lambda = 0.034 \pm 0.001$ s$^{-1}$, $\lambda_{ml} = (6.4 \pm 0.3) \times 10^{-5}$ s$^{-1}$ (solid line) and $A = 0.026 \pm 0.001$, $\lambda = 0.015 \pm 0.001$ s$^{-1}$, $\lambda_{ml} = (5.7 \pm 0.2) \times 10^{-5}$ s$^{-1}$ (dashed line). Additionally, the fits of the data from Fig. 4A are summed (dotted lines) to allow direct comparison of the processing reactivity of an asymmetric substrate versus that of the symmetric substrates.
Fig. 6. Reactivity of synapsed substrates at 130 mM NaCl. Single-turnover activity assay at 5.0 μM integrase, 0.5 μM DNA, 130 mM NaCl samples were subjected to electrophoresis on a 20% polyacrylamide, 8 mM urea, TBE sequencing gel. An S indicates the starting 21-mer-radiolabeled strand of the substrate; a P indicates the shortened processing products; extended joining products are indicated with a J. A, With 5'U5b/2/3b'S'3t as the substrate, lanes represent reaction times of 0, 5.2, 9.6, 14.0, 18.5, 23.4, 28.4, 45.1, 61, 120, 241, 600, 1200, and 1800 s. B, with *5t/5b(2)3b/3t as the substrate, lanes represent reaction times of 0, 5.4, 10.7, 16.3, 21.4, 27.3, 32.1, 45.3, 66, 120, 240, 600, 1200, and 1800 s.

The lack of detectable integrase-catalyzed splicing products from reverse-polarity substrates, even at reaction conditions shown to favor the splicing activity, suggests that these substrates are viable tools for studying the molecular details of the processing activity. Although a small amount of products larger than the starting material was visible in these reactions, these products did not appear until much later, after the initial appearance of processing products. Additionally, the distribution of sizes of these larger products indicates that they are most likely the result of joining rather than splicing activity. We conclude that the reverse-polarity substrates are an improvement over previous normal-polarity substrates because the lack of splicing activity allows for uncomplicated quantitation of presteady-state experiments to make mechanistic rather than phenomenological observations with regard to integrase reactivity.

Comparison of Concurrently Bound Viral LTR Ends—The utilization of reverse-polarity synapsed substrates along with selective radioactive labeling allowed for experiments in which integrase reactivity with a particular DNA sequence could be measured as a function of the other LTR sequence. Burst analysis allows this comparison to be carried out under single-turnover conditions, thus ensuring that the data reflected a single binding and catalysis event. Experiments were conducted to examine the extent of integrase processing reactivity with substrates containing either two U3 ends, a U3 and a U5 end, or two U5 ends. The data from presteady-state experiments identified the order integrase preference for LTR end combinations as U3/U5 > U3/U3 > U5/U5 in addition to confirming previous reports of the preference of integrase for the U3 sequence over the U5 sequence (12, 14, 25).

Our results represent the first direct evidence of the preference for a physiologically relevant U3/U5 combination of LTR ends in in vitro studies on the processing reaction. However, in light of the fact that the asymmetric U3/U5 pair is the “proper” in vivo substrate for the molecule, this preference is not altogether surprising. Additionally, this result is consistent with results from DNase protection analysis, which indicated that integrase likely binds asymmetrically to the U3 and U5 LTR ends when the two ends are coupled for correct full-site integration (17). Examination of the joining reaction products (Fig. 6) further suggests that these preferences extend beyond the processing activity and has implications on the temporal order of the joining reaction. Specifically, the U3 sequence had not only a significantly larger burst amplitude when examined for processing activity, but putative joining products from processed U3 ends also appeared at reaction times much earlier when compared with those from the U5 end. Although the two ends appear to be integrated concurrently on a longer and more macroscopic time scale (17, 22), our data suggest that the U3 cognate sequence is joined to the host DNA before the U5 end, even within a single turnover.

Comparative data of the U3 and U5 ends revealed that productive complex formation is maximal for substrates with both end sequences present. Although this result is consistent with the data reported by Vora et al. (14), our conclusions differ from theirs. We believe that the “macroscopic” nature of the analysis used in their study obscured the preference of integrase for the asymmetric U3/U5 combination of ends. In that work, integrase processing and joining activities were assayed by detection of the integration of radiolabeled 480-base pair mini-viral substrates containing the LTR ends into a 2,867-base pair circular target. Products accumulated after 10 min resulting from both full-site integration, in which two LTR ends are inserted into host DNA, and half-site integration, in which only one LTR end is inserted, were summed to arrive at an order of substrate preference. Full-site integration with these mini-viral substrates could have taken place through either a trimolecular donor reaction, where ends from two different donor molecules are joined to a single target DNA, or a bimolecular reaction, where the two ends of a single molecule are joined to a single target DNA to form the final integration product. It has been observed that the bimolecular reaction is less than 5% as efficient (14, 20, 21, 27). In contrast, the end-synapsed substrates used in our presteady-state assay are bound predominantly in a unimolecular fashion (10), thus allowing direct measurement of the reactivity of one LTR end sequence with another specific LTR end sequence bound at the active site. In addition, using single-turnover conditions allowed the resolution of the specific origins of the processing and joining products and their intermediates. Although from a macroscopic point of view, the integration of the two LTR ends into host DNA is temporally concerted in a single binding event, our presteady-state data show that within this single binding event the U3 end is processed before the U5 end and undergoes the joining activity earlier as well. A more recent study indicating
that concerted DNA integration requires the presence of both U3 and U5 ends in the donor DNA (22) provides additional direct evidence consistent with our results. However, by using the presteady-state assay, we were able to make a distinction between the microscopic order of the events within the integration reaction versus the seemingly macroscopic concerted nature of this reaction.

Acknowledgments—We are grateful to Drs. P. Shing Ho, Christopher Mathews, and Michael Schimerlik for critical reading of the manuscript.

REFERENCES

1. Katz, R. A., and Skalka, A. M. (1994) Annu. Rev. Biochem. 63, 133–173
2. Daniel, R., Katz, R. A., and Skalka, A. M. (1999) J. Virol. 73, 723–729
3. Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992) J. Virol. 66, 3593–3601
4. Kulkosky, J., Katz, R. A., Merkel, G., and Skalka, A. M. (1995) Virology 210, 448–456
5. Hishinuma, F., DeBona, P. J., Astrin, S., and Skalka, A. M. (1981) Cell 23, 155–164
6. Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A., and Capen, D. J. (1985) Nature 313, 450–458
7. Starhech, B., Ratner, L., Josephs, S. F., Okamoto, T., Gallo, R. C., and Wong-Staal, F. (1985) Science 227, 538–540
8. Vink, C., Groenink, M., Elgersma, Y., Fouchier, R. A., Tersmette, M., and Plasterk, R. H. (1990) J. Virol. 64, 5626–5627
9. Tao, K., Skalka, A. M., and Wang, I. (2002) J. Biol. Chem. 277, 12089–12098
10. Peliska, J. A., and Benkovic, S. J. (1992) Science 258, 1112–1118
11. Kukolj, G., and Skalka, A. M. (1995) Genes Dev. 9, 2556–2567
12. Chow, S. A. (1997) Methods 12, 306–317
13. Vora, A. C., Chiu, R., McLeod, M., Goodarzi, G., Stahl, S. J., Maeser, T. C., Hyde, C. C., and Grandgenett, D. P. (1997) J. Biol. Chem. 272, 23938–23945
14. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) Genes Dev. 11, 371–382
15. Pemberton, I. K., Buckle, M., and Buc, H. (1996) J. Biol. Chem. 271, 1498–1506
16. Vora, A., and Grandgenett, D. P. (2001) J. Virol. 75, 3556–3567
17. Craigie, R., Fujiwara, T., and Bushman, F. (1990) Cell 62, 829–837
18. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) Cell 63, 87–95
19. Aiyar, A., Hindmarsh, P., Skalka, A. M., and Leis, J. (1996) J. Virol. 70, 3571–3580
20. Fitzgerald, M. L., and Grandgenett, D. P. (1994) J. Virol. 68, 4314–4321
21. Brin, E., and Leis, J. (2002) J. Biol. Chem. 277, in press
22. Murphy, J. E., and Goff, S. P. (1992) J. Virol. 66, 5092–5095
23. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M. (1992) J. Biol. Chem. 267, 16037–16040
24. Fitzgerald, M. L., Vora, A. C., and Grandgenett, D. P. (1991) Anal. Biochem. 196, 19–23
25. Vora, A. C., and Grandgenett, D. P. (1995) J. Virol. 69, 7483–7488
26. Fitzgerald, M. L., Vora, A. C., Zeh, W. G., and Grandgenett, D. P. (1992) J. Virol. 66, 6257–6263
27. Grandgenett, D. P., Inman, R. B., Vora, A. C., and Fitzgerald, M. L. (1993) J. Virol. 67, 2628–2636
28. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) Biopolymers 9, 1059–1077
29. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) J. Virol. 63, 5319–5327
30. Terry, R., Solis, D. A., Katzman, M., Colbrinik, D., Leis, J., and Skalka, A. M. (1988) J. Virol. 62, 2358–2365
31. Coleman, J., Eaton, S., Merkel, G., Skalka, A. M., and Laue, T. (1999) J. Biol. Chem. 274, 32942–32946
32. Ellison, V., and Brown, P. O. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 7316–7320
33. Patel, S. S., Miller, M. D., Bushman, F. D., and Orgel, L. E. (1996) Virology 222, 283–288
34. Patel, S. S., Wong, I., and Johnsen, K. A. (1991) Biochemistry 30, 511–525
35. Deprez, E., Tuve, P., Leh, H., Mouseadet, J. F., Auclair, C., and Brochon, J. C. (2000) Biochemistry 39, 9275–9284
36. Chen, J. C., Krucinski, J., Miercze, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7320–7325
37. Katz, R. A., DiCandolo, P., Kukolj, G., and Skalka, A. M. (2001) J. Biol. Chem. 276, 34213–34220
38. Pemberton, I. K., Buckle, M., and Buc, H. (1996) J. Biol. Chem. 271, 1498–1506
39. Wang, J., Ling, H., Yang, W., Craigie, R. (2001) EMBO J. 7333–7343