Nucleophile Selection for the Endonuclease Activities of Human, Ovine, and Avian Retroviral Integrase*

Received for publication, August 3, 2000, and in revised form, October 5, 2000
Published, JBC Papers in Press, October 6, 2000, DOI 10.1074/jbc.M007032200

Lynn M. Skinner‡, Malgorzata Sudol‡, Amy L. Harper§, and Michael Katzman‡§¶

From the ‡Department of Medicine and §Department of Microbiology and Immunology, Pennsylvania State University College of Medicine, The Milton S. Hershey Medical Center, Hershey, Pennsylvania 17033

† To whom correspondence should be addressed: Dept. of Medicine, Division of Infectious Diseases, Pennsylvania State University College of Medicine, The Milton S. Hershey Medical Ctr., P. O. Box 830, Mail Code H036, Hershey, PA 17033-0850. Tel.: 717-531-8881; Fax: 717-531-4633; E-mail: mkatzman@psu.edu.

‡ The abbreviations used are: IN, integrase; HIV-1 and -2, human immunodeficiency virus types 1 and 2; RSV, Rous sarcoma virus.

Retroviral integrases catalyze four endonuclease reactions (processing, joining, disintegration, and nonspecific alcoholysis) that differ in specificity for the attacking nucleophile and target DNA sites. To assess how the two substrates of this enzyme affect each other, we performed quantitative analyses, in three retroviral systems, of the two reactions that use a variety of nucleophiles. The integrase proteins of human immunodeficiency virus type 1, visna virus, and Rous sarcoma virus exhibited distinct preferences for water or other nucleophiles during site-specific processing of viral DNA and during nonspecific alcoholysis of nonviral DNA. Although exogenous alcohols competed with water as the nucleophile for processing, the alcohols stimulated nicking of nonviral DNA. Moreover, different nucleophiles were preferred when the various integrases acted on different DNA targets. In contrast, the nicking patterns were independent of whether integrase was catalyzing hydrolysis or alcoholysis and were not influenced by the particular exogenous alcohol. Thus, although the target DNA influenced the choice of nucleophile, the nucleophile did not affect the choice of target sites. These results indicate that interaction with target DNA is the critical step before catalysis and suggest that integrase does not reach an active conformation until target DNA has bound to the enzyme.

Integration of a DNA copy of the retroviral genome into cellular DNA is necessary for retrovirus replication and for the development of retrovirus-related diseases. This recombination event is mediated by the retroviral integrase (IN)† protein. Purified integrase catalyzes one-step transesterifications in which the nucleophilic oxygen of an OH group attacks a DNA phosphodiester bond and covalently joins to the target DNA at the site of nicking (1). To mediate integration in vivo, integrase must catalyze two reactions that use distinctly different nucleophiles and targets. The first reaction, termed processing (see Fig. 1A), prepares the viral DNA ends for subsequent attachment to cellular DNA (2). The target for this site-specific endonuclease reaction is the phosphodiester linkage immediately 3′ to the A of highly conserved CA bases near the 3′-ends of unintegrated retroviral DNA. Nicking at this site removes the terminal nucleotides (usually two) that follow the CA and creates a recessed 3′-hydroxyl group at each DNA end. In vitro studies indicate that the nucleophile for this reaction can be provided by the OH group of a water molecule (3), by certain alcohols (3, 4), or by the OH group at the 3′-end of the unintegrated DNA itself (1) (see Fig. 2). Thus, processing is a site-specific alcoholysis reaction that uses a variety of nucleophiles (3). The second reaction, referred to as joining or strand transfer (see Fig. 1B), inserts the processed viral DNA ends into the two strands of cellular DNA at sites that are separated by a few base pairs (5, 6). Sequencing of multiple insertion sites has not revealed any target consensus sequence, even though certain sites within any sequence may be preferred during in vivo and in vitro integration (7). Thus, joining can be considered nonspecific with respect to target DNA. However, only one nucleophilic donor, i.e. the processed viral DNA end, can be used for successful integration.

In addition to the biologically relevant processing and joining activities, integrase exhibits two other endonuclease activities in vitro. During disintegration (see Fig. 1C), which is a reversal of the joining reaction, the 3′-OH of an oligonucleotide representing nicked cellular DNA attacks the bond linking the CA at the viral DNA end to nonviral DNA. This action seals the nick in the cellular DNA mimic and releases a processed viral DNA end (8). Because the nucleophile and target for this reaction are determined by the oligonucleotide complex, which juxtaposes the reactive groups, disintegration is specific for both substrates. The final and most recently described endonuclease activity of integrase is nonspecific alcoholysis (see Fig. 1D). During this activity, integrase uses a variety of small alcohols as nucleophiles that nick and join to any internal 5′-phosphate group of DNA (9). As with DNA joining, any position in nonviral DNA can be attacked even though certain sites within a given target are preferred. Both disintegration and nonspecific alcoholysis can be catalyzed by the isolated central domain of the integrase protein, unlike processing and joining, which generally require the complete enzyme (9, 10). The biological significance of disintegration and nonspecific alcoholysis is unknown. However, because integrase has a single catalytic site (11), these reactions have proved useful for analyzing the mechanism and substrate interactions of this important enzyme.

Despite their mechanistic similarities, the four reactions catalyzed by integrase represent all possible combinations of specificity for the attacking nucleophile and the target DNA (see Fig. 1). In particular, processing uses various nucleophiles and a specific target site, joining uses a specific nucleophile and various target sites, disintegration uses a specific nucleophile and a specific target site, and nonspecific alcoholysis uses var-
Nucleotide(s) (DNA end. During nonspecific alcoholysis (DNA acts as the nucleophile to nick and join to any site in target DNA (similar reactions occur on each cellular DNA strand). During disintegration (C), which is a reversal of the joining reaction, a juxtaposed 3’-DNA-end attacks after the CA to seal the nick and regenerate the processed viral DNA end. During nonspecific alcoholysis (D), a variety of nucleophiles can nick and join to any internal site in DNA. The first two reactions occur in vivo, whereas all four reactions occur in vitro. Characteristics of the attacking nucleophile and target DNA are indicated at the right.

EXPERIMENTAL PROCEDURES

Purified Integrase—The integrase coding regions of HIV-1BH10 and visna virus were cloned into plasmid pQE-30 and expressed in Escherichia coli M15[pREP4] (Qiagen, Inc., Chatsworth, CA), as described previously (12, 13). The RSV integrase coding region from a derivative of plasmid pH-RCAN-HiSV (14), which was obtained from Rebecca Craven of Pennsylvania State University, was cloned and expressed similarly. Proteins were purified as described previously (12, 13) except that glycerol was omitted from all buffers when the integrases were to be tested in the absence of alcohols or with other alcohols. Such glycerol-free proteins were not frozen. The concentrations of purified proteins were measured by comparison to a series of Coomassie Blue-stained standards following SDS-polyacrylamide gel electrophoresis and densitometry of dried gels. The concentrations of the purified HIV-1 and visna virus integrases were ~200–700 ng/μl (6–20 pmol/μl), yielding a final concentration in the typical 10-μl reaction of ~1–2 pmol/μl (1–2 μM). Because RSV integrase always purified at higher concentrations, it was diluted 5- to 25-fold to yield comparable concentrations in all reactions.

Oligodeoxynucleotides—The sequences of the substrates for the processing assays were as follows (only the labeled strand is shown, and the CA is underlined): HIV-1 U5 plus strand 18-mer, TTGAAAATCTCTA-GCAG; visna virus U3 minus strand 18-mer, GTCATCTGCTCTGAC-AGT; RSV U3 minus strand 18-mer, ATTTAGCATGCTACATT. The concentration of the attacking nucleophile and target DNA are indicated at the right.

**FIG. 1.** Endonuclease activities of integrase. Curved arrows denote the coupled nucleophilic cleavage and joining that is common to all actions catalyzed by integrase. Heavy lines represent viral DNA and light lines represent any DNA. 5’-Phosphate groups are depicted by closed circles and 3’-hydroxyl groups by open circles. The invariant CA bases near the viral DNA end are in boldface. During processing (A), the terminal nucleotides (NN) can be removed by a variety of nucleophiles (represented as ROH). During joining (B), the recessed 3’-end of the processed viral DNA acts as the nucleophile to nick and join to any site in target DNA (similar reactions occur on each cellular DNA strand). During disintegration (C), which is a reversal of the joining reaction, a juxtaposed 3’-DNA-end attacks after the CA to seal the nick and regenerate the processed viral DNA end. During nonspecific alcoholysis (D), a variety of nucleophiles can nick and join to any internal site in DNA. The first two reactions occur **in vivo**, whereas all four reactions occur **in vitro**. Characteristics of the attacking nucleophile and target DNA are indicated at the right.

| Endonuclease Activities of Integrase | Nucleophile | Target |
|-------------------------------------|-------------|--------|
| Processing                           | Non-specific| Specific|
| DNA Joining or Strand Transfer       | Specific     | Non-specific|
| Disintegration                       | Specific     | Specific|
| Nonspecific Alcoholysis              | Non-specific| Non-specific|

Quantitation of Results—Results were quantified by measuring the radioactivity of bands in wet gels with a Betascope (Betagen, Waltham, MA).
RESULTS
Activity of Enzymes Purified without Glycerol—The integrase proteins from HIV-1, visna virus, and RSV were purified from a standard bacterial expression system. To permit testing of the enzymes in the presence of various nucleophilic alcohols, most purifications were performed without glycerol in the extraction or dialysis buffers, which necessitated storage of the proteins at 4 °C. Such preparations had comparable protein concentrations and enzymatic activity to our standard preparations purified in the presence of 10% glycerol and stored in 40% glycerol at −70 °C (data not shown). Although glycerol-free RSV integrase remained active for weeks when stored at 4 °C, the HIV-1 and visna virus proteins lost activity during similar storage and thus were usually tested immediately following purification. Each integrase was tested on oligonucleotides derived from one end of its cognate viral DNA in the presence of different concentrations of various nucleophilic alcohols. To facilitate comparisons, the viral DNA end that was previously established as most susceptible to each enzyme was used as the appropriate substrate for processing, i.e., the HIV-1 U5 end (the downstream or right end) and the visna virus and RSV U3 ends (the upstream or left ends) (12, 13, 15). Alcohols that were used for these experiments included three diols previously shown to support integrase-mediated alcoholysis (3, 9): 1,2-ethanediol (ethylene glycol), 1,2-propanediol (propylene glycol), and 1,2,3-propanetriol (glycerol). Reactions conducted with 5′-labeled substrates demonstrated that the presence of these alcohols at concentrations up to 30% generally had minimal or no effect on the extent of processing by any of the three integrases during standard reactions. Similarly, the yield and patterns of the longer strand-transfer products were not influenced by the exogenous alcohols (data not shown).

Nucleophiles Used for Processing Viral DNA Ends—The departing nucleotides at the viral DNA ends are released by integrase as a simple dinucleotide when water serves as the nucleophile, as an alcohol adduct when an exogenous alcohol is the nucleophile, or as a cyclic dinucleotide when the OH at the 3′-DNA-end loops around to create the nick after the CA (Fig. 2). Although assays performed with 5′-labeled substrates cannot reveal which nucleophiles are used for this reaction, the various products can be distinguished if the DNA is radiolabeled near the 3′-end (Fig. 2) (1, 3). Under such conditions, the linear dinucleotide comigrates on denaturing polyacrylamide gels with a specific oligonucleotide marker, the alcohol adducts migrate more slowly to novel positions that are a function of the particular alcohol (i.e. the identity of the R group in Fig. 2), and the cyclic product migrates at an intermediate position just above the alcohol adducts. We first confirmed that each of the integrases created all three types of products. Initial reactions were conducted in the presence of Mn2+, because all retroviral integrases exhibit maximal activity in standard oligonucleotide assays when this divalent cation is present. For the HIV-1 and visna virus integrases, three major products were evident (Fig. 3, A and B). These bands were previously identified as the alcohol adducts (A), the cyclic dinucleotide (C), and the linear dinucleotide (L). None of these bands were observed when reactions were conducted without enzyme, whether or not an exogenous alcohol was present (lanes 2 and 3 in Fig. 3, A and B). Under these conditions, the HIV-1 and visna virus integrases predominantly used water as the nucleophile for processing but also used the exogenous alcohols efficiently, whereas the cyclic products were produced to a lesser extent (lanes 5-7 in Fig. 3, A and B). When reactions were conducted in the absence of the alcohols, each of these enzymes preferentially used water rather than the DNA end as the attacking nucleophile (lane 4 in Fig. 3, A and B).

In contrast to the above findings, RSV integrase had a strong preference for using the 3′-OH at the DNA end to make cyclic products, whether or not exogenous alcohols were present (Fig. 3C, lanes 4-7). Moreover, the alcohol adducts were produced inefficiently and their detection required a darker autoradiographic exposure than that shown. As expected, the bands identified as the cyclic products and alcohol adducts were resistant to alkaline phosphatase and polynucleotide kinase (data not shown). RSV integrase also created a moderate amount of linear trinucleotide products (T), and the cyclic products likely are a mixture of circular di- and trinucleotides. This result was expected, because the avian retroviral integrases are known to nick between the conserved C and A nucleotides in addition to the biologically relevant site after the A when reactions contain Mn2+ (2, 6, 16). In fact, RSV integrase created comparable amounts of 15- and 16-mer products under these reaction conditions when the 18-mer substrate was labeled at the 5′-end (not shown). Combining this observation with the finding that the amount of linear dinucleotides (L) evident in Fig. 3C was less than the amount of linear trinucleotides (T) indicates that most of the cyclic products resulted from circularization of a departing dinucleotide rather than a trinucleotide. This conclusion is consistent with the observation that the cyclic products sometimes resolved into a predominant faster migrating component and a minor slower component.

We also examined the choice of nucleophile for processing by all three enzymes in reactions that contained Mg2+. Although RSV integrase is very active when Mg2+ serves as the cofactor, most integrases are inactive with Mg2+ under standard reaction conditions. However, we recently reported that our preparations of HIV-1 and visna virus integrase exhibit high levels of Mg2+-dependent activity when processing assays are supplemented with 30–40% dimethyl sulfoxide (Me2SO) (17). By com-
prominent in Mn\(^{2+}\) products are indicated: L, linear dinucleotide; C, cyclic product; A, alcohol adducts. For RSV IN (panel C), the trinucleotide product (T) that is prominent in Mn\(^{2+}\)-dependent reactions also is indicated, the cyclic products migrate as a broad band, and visualization of the alcohol adducts required a longer exposure. The nature of the integrase-dependent but alcohol-independent band migrating just behind the 9-position in panel B is unknown (12). Other minor bands are present in the negative control lanes or reflect secondary cleavage sites.

Quantitative Analysis of Nucleophile Selection during Processing.—The results in Figs. 3 and 4 suggest distinct preferences by the three integrases for using different nucleophiles (data not shown). Retroviral integrases can create the alternative products in reactions performed with Mg\(^{2+}\) (compare the amounts of dinucleotides and trinucleotides in lanes 6 and 7 of Fig. 4, A and B). These faint bands, which are better appreciated on the original autoradiograms, align with the cyclic dinucleotide markers and are not evident from the reactions that did not contain integrase. The yield of alcohol products also was low, perhaps because the large amount of Me\(_2\)SO in these reactions permitted the addition of relatively low alcohol concentrations (data not shown). That retroviral integrases can create the alternative products in reactions performed with Mg\(^{2+}\) is made clear by reactions using RSV integrase. This enzyme was active with Mg\(^{2+}\) even without the addition of Me\(_2\)SO and readily made all three types of nucleophilic donors for the processing reactions (Fig. 4C). As expected, greater specificity for nicking after the CA was exhibited with Mg\(^{2+}\) (compare the amounts of dinucleotides and trinucleotides in lanes 6 and 7 of Fig. 4C), which was confirmed in reactions that used 5'-labeled substrates (not shown). Taken together, Figs. 3 and 4 demonstrate that each of these integrases can use all three types of nucleophilic donors for the processing reaction, albeit with different efficiencies, whether reactions are conducted with Mn\(^{2+}\) or Mg\(^{2+}\).

Nucleophile Selection by Retroviral Integrase

![Graph](image.png)
proximately equal amounts were produced (Fig. 5, A and B). In a similar manner, visna virus integrase substituted the alcohols for water as the attacking nucleophile about one-third of the time (Table I, compare the 90% in the first column for visna virus with the 32% in the second column). In contrast, RSV integrase substituted the alcohols for water less than 20% of the time, such that 5–7% of the products were formed by the use of these nucleophiles in the presence of either cation (Table I).

In addition to varied usage of different types of nucleophiles, varied usage of different exogenous alcohols was observed (Fig. 6). The HIV-1 and visna virus integrases readily used glycerol and 1,2-ethanediol (ethylene glycol), whereas 1,2-propanediol (propylene glycol) was used somewhat less efficiently. Firm conclusions about the relative usage of these three alcohols by RSV integrase could not be drawn from the data. Each of these diols presents OH groups on adjacent carbon atoms. Previously, Vink et al. (3) reported that 1,3-propanediol, which does not have adjacent OH groups, was not used by HIV-2 integrase for processing. We therefore tested this chemical with all three enzymes and found that it was used much less efficiently than the 1,2-diols. However, each of the enzymes was able to utilize it as the nucleophile for processing (Fig. 6). Thus, retroviral integrase does not have an absolute requirement for adjacent OH groups on acceptable nucleophiles, although there is a marked preference for such substrates.

**Nucleophile Selection by Retroviral Integrase**

**TABLE I**

| Product | HIV-1 IN/Mn²⁺ | Visna IN/Mn²⁺ | RSV IN/Mn²⁺ | RSV IN/Mg²⁺ |
|---------|---------------|---------------|-------------|-------------|
| Alcohol adduct | -a | +a | - | + | - | + |
| Cyclic | 11 | 10 | 11 | 9 | 74 | 73 | 39 | 38 |
| Linear dinucleotide | 89 | 50 | 90 | 59 | 4 | 3 | 61 | 54 |
| Linear trinucleotide | - | - | - | - | 23 | 19 | - | - |
| Total processing | (45%) | (37%) | (43%) | (58%) | (68%) | (68%) | (36%) | (24%) |

* *a*, alcohol present; *-*, alcohol absent.

Numbers represent the average amount of each product formed (normalized to 100%) in reactions with the indicated integrase (IN) and divalent cation. The percentage of substrate that was shortened by two nucleotides is shown in parentheses at the bottom as Total processing, except that nicking to release trinucleotides also was included in the calculations for RSV IN with Mn²⁺. Reactions with alcohol typically contained 25–40% of ethylene glycol, propylene glycol, or glycerol. The calculations for reactions with alcohol used the maximum amount of each alcohol adduct measured in 8–10 experiments. Reactions without alcohol were done four to five times.

**Fig. 4. Choice of nucleophile during Mg²⁺-dependent processing.** Double-stranded 18-mers derived from the indicated viral DNA ends were labeled near the 3'-end of the strand with the invariant CA and used as the substrate for the corresponding IN in the presence of one nucleophilic alcohol (E, ethylene glycol, lanes 4; P, propylene glycol, lanes 5; or G, glycerol, lanes 6). Control reactions were without alcohol or IN (lanes 1), with one alcohol (ethylene glycol) but without IN (lanes 2), or with IN but without alcohol (lanes 3). Reactions in lanes 1–6 of panel A were conducted with 7.5 mM MgCl₂ and 30% Me₂SO, those for lanes 1–6 in panel B contained 10 mM MgCl₂ and 35% Me₂SO, and those for lanes 1–6 in panel C had 5 mM MgCl₂ and no Me₂SO. As markers for the various products, reactions conducted with 10 mM MnCl₂ and one alcohol are included as lanes 7 in each panel. The 18-mer substrate and specific cleavage products (as in Fig. 3) are indicated. Additional bands are due to the extended autoradiographic exposures necessary to demonstrate the various products.
whereas the relative choice of nucleophile is revealed by the use of substrates labeled near the 3’-end. Thus, the two types of experiments yield complementary information. Nonspecific alcoholysis assays were performed with Mn²⁺ as the divalent cation, because RSV integrase catalyzes this reaction inefficiently in the presence of Mg²⁺ (9) and we have not detected Mg²⁺-dependent nonspecific alcoholysis by HIV-1 or visna virus integrase.

Several important observations resulted from experiments that used a 5’-labeled nonviral DNA substrate. Perhaps most striking, the total amount of DNA nicking was dramatically augmented by increasing concentrations of exogenous alcohols. In the presence of no or minimal amounts of the alcohols, HIV-1 integrase (Fig. 8, A and D) and visna virus integrase (Fig. 8, B and E) nicked about 15% of the DNA substrate by hydrolysis, whereas more than 50% of the substrate was nicked with high concentrations of the alcohols. Even RSV integrase, which nicked about 50% of the DNA in the absence of alcohols, was stimulated to nick an additional 20% of the substrate with optimal amounts of the alcohols (Fig. 8, C and F). When reactions were performed with lower concentrations of RSV integrase to reduce the baseline amount of hydrolysis, stimulation similar to that seen for the HIV-1 and visna virus enzymes was observed (data not shown). Importantly, no nicking occurred when the DNA was incubated with high concentrations of alcohol in the absence of integrase (lane 1 in Fig. 8, A–C). Similar results were obtained with glycerol, 1,2-ethanediol, 1,2-propanediol, and 1,3-propanediol (although total nicking was not as high with the last chemical). Thus, these agents do not merely compete with water but also stimulate the nonspecific nuclease activity of integrase.

The distinct target site preferences of the various integrases are evident in Fig. 8. HIV-1 integrase preferentially nicked this substrate at positions that are 8, 11, 15, and 17 nucleotides from the 5’-end, whereas visna virus integrase preferred positions 5, 15, and 19 and RSV integrase preferred positions 8, 12, 15, and 19. The patterns of prominent products were generally independent of the presence or amount of the alcohols (for example, compare lanes 2–4 with lanes 7–9 in panels A, B, and C of Fig. 8). The only apparent exceptions are the diminished amounts of 19-mers and 15-mers produced by RSV integrase in reactions with the highest alcohol concentrations (Fig. 8C, lanes 8 and 9). However, time-course studies proved that these findings actually were due to further nicking of the 19- and 15-mers to yield the very prominent 8-mer product under conditions in which the 23-mer substrate was being depleted (not shown). Thus, the nicking pattern was independent of the alcohol concentration. It should also be noted that, although 1,2-propanediol was used for Fig. 8, the same patterns of preferential sites were obtained when the other three alcohols were used (not shown). These data suggest that the selection of target sites is not influenced by which nucleophile integrase uses to nick nonviral DNA. However, the experiments do not reveal which nucleophile actually was used, because it is possible that all of the nicked products resulted from hydrolysis.

To ascertain which nucleophile was used, experiments were
performed with substrates labeled near the 3'-end (as depicted in Fig. 7). Under these conditions, products of hydrolysis comigrate with an oligonucleotide ladder (Fig. 9A, lane M), whereas alcohol adducts migrate to novel positions that are a function of the particular alcohol (Fig. 9). Results are shown only for 1,2-propanediol. As expected, each integrase exhibited the same site preferences with the 3'-labeled substrate as with the 5'-labeled DNA. The positions of the prominent alcoholysis products in Fig. 9 were interpolated from the oligonucleotide positions. Each of these bands correlates with a prominent band of labeled DNA. For example, the band migrating as a "9.7-mer," created from the 3'-labeled 24-mer substrate, reflects the same nicking event that created a 15-mer product from the 5'-labeled 23-mer substrate (as demonstrated diagrammatically in Fig. 9D). However, additional information was provided by the experiments with 3'-labeled substrates. For example, the nick at the position 15 nucleotides from the 5'-end, which created the labeled 15-mer in Fig. 8A, produced bands at the 9 and 9.7 positions in Fig. 9A, reflecting hydrolysis and alcoholysis, respectively. Similarly, the nick at the position 11 nucleotides from the 5'-end, which created the labeled 11-mer in Fig. 8A, produced bands at the 13 and 13.5 positions in Fig. 9A. Thus, the amounts of hydrolysis and alcoholysis at specific sites can be compared.

Inspection of the autoradiograms in Fig. 9 and similar experiments permitted several conclusions. First, the presence of the alcohols generally had little effect on the amount of hydrolysis by integrase. Almost all of the additional DNA nicking was from integrase-mediated alcoholysis, which was evident with as little as 4% of the alcohols. We have even detected alcohol adducts with 0.1% of exogenous alcohol (data not shown). Second, at high alcohol concentrations, integrase sometimes used the exogenous nucleophile in place of water, as indicated by diminished amounts of hydrolysis and increased amounts of alcoholysis at the same position (e.g. compare the bands at the 16 and 16.4 positions in Fig. 9C). Third, under optimal conditions, the integrases used the alcohols comparably to or more often than water. This fact is revealed by comparing the sum of the intensities of the alcoholysis products with the corresponding hydrolysis products in Fig. 9 (more evident in panels A and B but also seen in other experiments with RSV integrase). Finally, the target site preferences exhibited by each integrase were the same for hydrolysis and alcoholysis (compare lanes 2 and 9 in each of Figs. 9A, B, and C). Moreover, these preferences were independent of the type of alcohol used as the nucleophile, because similar results were obtained with all four alcohols tested (data not shown). These results prove that selection of target sites was not influenced by which nucleophile integrase used to nick nonviral DNA.

**DISCUSSION**

Retroviral integrase is a multifunctional enzyme that is responsible for the permanent incorporation of a DNA copy of the retroviral genome into cellular DNA. Knowledge of the biochemistry of how integrase functions is critical for obtaining a more complete view of the virus life cycle and for new treatments for the acquired immunodeficiency syndrome and other retrovirus-related diseases. In particular, how integrase interacts with its various substrates and how these substrates interact with each other are poorly understood. To examine how target DNA and nucleophile affect each other during reactions catalyzed by integrase, the experiments in this report compared nucleophile selection by different integrases during the activities that accommodate a variety of nucleophiles.

Processing of viral DNA ends in preparation for insertion into cellular DNA is the first enzymatic action required of integrase. In *vitro* studies have identified three types of nucleophiles that can provide the OH group for nicking near the viral DNA ends, i.e. water (3), the 3'-end of the unintegrated DNA (1), and certain alcohols. Exogenous alcohols shown previously to participate in this reaction either have OH groups on adjacent carbon atoms, including glycerol, 1,2-ethanediol, and 1,2-propanediol (3), or have OH and amino groups on adjacent carbons, such as the amino acids serine and threonine (3, 4). In contrast to a report that 1,3-diols do not participate, we found that the retroviral integrases from two lentiviruses (HIV-1 and visna virus) and an oncovirus (RSV) can use 1,3-propanediol for processing, although less efficiently than 1,2-diols (Fig. 6). It is likely that the structure of acceptable nucleophiles reflects the configuration of the site on the protein that interacts with the nucleophilic donor before or during catalysis.

We found that the three integrases studied in this report exhibited different nucleophile selectivities during processing. Consistent with previously published data for Mn<sup>2+</sup>-dependent reactions (1, 3, 12, 18), HIV-1 integrase and visna virus integrase primarily used water rather than the viral DNA ends as the nucleophile, whereas RSV integrase had a strong preference for using the 3'-OH at the viral DNA end (Table I). These results have now been documented in reactions performed in the absence of exogenous nucleophiles by using glycerol-free integrase preparations, which had previously been done only for the HIV-1 enzyme (3). Moreover, when reactions were sup-

![Fig. 7. Hydrolysis versus alcoholysis of nonviral DNA. A single nick on one DNA strand is shown. The nucleophilic oxygen (with its donor electrons indicated by two dots) is provided by an OH group of water (HOH) or an alcohol (ROH). In both schemes, the attacking oxygen nicks on one side of a phosphodiester bond and joins to the newly exposed 5'-phosphate group at the site of DNA cleavage. If the radiolabel is on the 5'-end of the substrate (as indicated by asterisks), all labeled products from either pathway comigrate on gels with similarly labeled oligonucleotide markers. If the label is near the 3'-end of the DNA (as indicated by circles), only the labeled hydrolysis products comigrate with these markers and the labeled alcoholysis products migrate as a function of the attached alcohol.](image-url)
implemented with optimal concentrations of alcohols, HIV-1 integrase used the alcohols almost as often as water, whereas visna virus integrase used the alcohols somewhat less often than water. In marked contrast, RSV integrase used the alcohols inefficiently. When processing reactions were conducted with Mg$^{2+}$ as the divalent cofactor, the preference of the HIV-1 and visna virus integrases for water became even stronger and minimal amounts of the other products were detected, consistent with previous reports for HIV-1 or HIV-2 integrase (3, 19, 20). Although the use of water by RSV integrase also was facilitated in reactions with Mg$^{2+}$, as described by others (18), we found that the avian enzyme made cyclic products and alcohol adducts even with this cation. The distinct preferences of the three enzymes are unlikely to be due to membership in different retrovirus subfamilies, because preferential usage of the 3' viral DNA end as the nucleophile for processing has also been described for murine leukemia virus (21) and feline immunodeficiency virus integrases (22). Rather, these preferences must reflect subtle differences in the structure of the various proteins, as suggested by crystallographic studies of the HIV-1 and RSV enzymes (23–25). This conclusion is consistent with the finding that certain amino acid substitutions

---

**Fig. 8.** Alcohol stimulates nonviral DNA nicking by integrase. A–C, double-stranded nonviral 23-mers (the sequence is shown in Fig. 9D) were 5'-labeled on one strand and incubated with glycerol-free preparations of the indicated integrase (IN) and increasing amounts of 1,2-propanediol, as indicated by % ROH. Sequence-specific markers (M) in panel A are relevant for all three sets of reactions, which are from adjacent parts of a single gel. Numbers beneath the lanes are aligned with the lower parts of the lanes. The sizes in nucleotides of the preferential cleavage sites for each IN are indicated to the right of the panels. D–F, the total amount of DNA nicking by the indicated enzymes in the presence of various nucleophilic alcohols was quantified. Each curve represents the average of all experiments in which >5% of the substrate was nicked. The average number of reaction sets for each curve was 3 (range, 2–7).

---
in the central region of HIV-1 or HIV-2 integrase affect the relative amounts of the various dinucleotide products created by processing (19, 26). Similarly, the influence of the divalent cation on nucleophile selection for processing may reflect metal-induced conformational changes of the enzyme (27). Steric effects and electrostatic interactions between integrase and substrate DNA also have been suggested to contribute to the choice of nucleophile during processing (20).

Nonspecific alcoholysis, the recently discovered fourth activity of integrase, is the most potent action of several integrase proteins (13, 28). This activity uses a variety of nucleophiles (as does processing), nicks almost any target DNA site (as does joining), and only requires the central domain of the protein (as does disintegration). Thus, nonspecific alcoholysis has a unique combination of characteristics that makes it a useful tool for understanding how integrase selects nucleophiles during catalysis. We have previously emphasized the relationship of this activity to joining, because both activities attack multiple DNA sites, lack a target consensus sequence, avoid 5'-ends of substrates, and display site preferences that are a function of the target DNA sequence and the viral source of integrase (29). We have now found that HIV-1 integrase and visna virus integrase exhibit similar patterns of nucleophile selection during nonspecific alcoholysis reactions, whereas RSV integrase performs greater hydrolysis. The high baseline level of hydrolysis by RSV integrase was not due to contaminating nucleases, because it demonstrated the signature nicking pattern (9) characteristic of the avian enzyme (Figs. 8C and 9C). In fact, the patterns of preferential nicks created by each enzyme were the same whether the enzymes were catalyzing hydrolysis or alcoholysis, consistent with the idea that these reactions reflect one catalytic mechanism. Moreover, the nicking patterns were independent of which exogenous alcohol was used. Thus, the nucleophile did not affect the choice of target DNA sites.

In contrast to the above conclusion, the target DNA did affect selection of the nucleophile. In fact, several differences in nucleophile usage were noted between processing and nonspecific alcoholysis reactions that differ merely by the DNA that is presented to integrase. Most striking was the finding that exogenous nucleophiles stimulated the nonspecific nuclease activity of integrase. Much of the additional nicking of nonviral DNA was due to the use of exogenous alcohols as the attacking nucleophile (Fig. 9), whereas the same alcohols only competed with water during processing (e.g., Fig. 5). In addition, the various 1,2-diols were used more equivalently during nonspe-

**Fig. 9. Choice of nucleophile with nonviral DNA.** A–C, double-stranded nonviral 24-mers labeled near the 3'-end of one strand were incubated with glycerol-free preparations of the indicated integrases and increasing amounts of 1,2-propanediol, as indicated by % ROH. Hydrolysis markers (M) in panel A are relevant for all three sets of reactions, which are from adjacent parts of a single gel. The apparent sizes in nucleotides of prominent products that migrate differently from the hydrolysis markers are interpolated from the autoradiogram and are indicated to the right of the panels. D, the sequence of the labeled strand of the nonviral 24-mer substrate used for the reactions in panels A–C is shown; the radiolabel is near the 3'-end (as indicated by the circle). The 23-mer substrate for the reactions shown in Fig. 8 was labeled at the 5'-end (as depicted by the asterisk) and did not contain the final 3'-base pair. Also indicated are the relationships of the prominent alcoholysis products seen in panels A–C (numbered from the 3'-end and shown above the sequence) to the preferred cleavage products seen in Fig. 8 (numbered from the 5'-end and shown below the sequence). Note that the effect of the attached alcohol on migration through the gel diminished as the length of the 3'-end-labeled products increased.
specific alcoholysis than during processing (Figs. 6 and 8). Moreover, the difference in usage of a 1,3-diol compared with the 1,2-diols appeared to be less during nonspecific alcoholysis than during processing. Finally, RSV integrase used water relatively inefficiently as the nucleophile for processing viral DNA ends (Table I) but performed high levels of hydrolysis on nonviral DNA (Figs. 8 and 9).

The ability to accommodate a diversity of nucleophiles suggests considerable flexibility for the active site of integrase (1). Indeed, we observed that the use of exogenous alcohols could exceed the use of water during processing or nonspecific alcoholysis (Figs. 5 and 9). Although the alcohol concentrations reached in these experiments are relatively high (40% glycerol or propylene glycol corresponds to 5.4 M and 40% ethylene glycol corresponds to 7.2 M), they are much less than the 33 M concentration of water in a solution that is 60% water. How integrase chooses various nucleophiles for these reactions is unclear, but current evidence indicates that the nucleophiles interact with integrase as specific substrates rather than merely as solvent molecules attacking DNA that is appropriately positioned by the enzyme. Most importantly, use of the nucleophiles involves specificity because not all compounds are used equivalently. For example, ethanol and 2-propanol were not used by integrase (3, 9). Moreover, many compounds that inhibit integrase 

in vitro have adjacent OH groups (30, 31), as do the most active alcohols in these assays, suggesting that these agents act at a similar protein site.

It is important to note that the cyclic dinucleotide and alcohol adducts described in these experiments are stable products that do not convert to hydrolysis products. This fact was confirmed by purifying the different products from gels, incubating them under various conditions, including exposure to alkali, and re-examining them by gel electrophoresis (data not shown). Thus, our measurements reflect the true distribution of the various products. Moreover, the relative distribution of the various products was not affected by the extent of the reaction. In particular, time-course studies of the processing reaction showed similar kinetics of appearance of the three types of dinucleotide products. In addition, the relative amounts of the various products were not affected by the concentration of substrate DNA when initial rates of product formation were measured during the early, linear phase of the time course. Analogous results were obtained for the nonspecific alcoholysis reaction when the kinetics of appearance of the hydrolysis and alcoholysis products and the initial rate of formation of these products as a function of DNA concentration were compared.

Furthermore, the concentrations of integrase used in this report were within the linear part of the reaction curve for both types of assays when initial rates of product formation were compared with enzyme concentration (data not shown). Interestingly, the amount of product formed in these experiments was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected when integrase was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected. Interestingly, the amount of product formed in these experiments was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected when integrase was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected.

Thus, our measurements reflect the true distribution of the various products. Moreover, the relative distribution of the various products was not affected by the extent of the reaction. In particular, time-course studies of the processing reaction showed similar kinetics of appearance of the three types of dinucleotide products. In addition, the relative amounts of the various products were not affected by the concentration of substrate DNA when initial rates of product formation were measured during the early, linear phase of the time course. Analogous results were obtained for the nonspecific alcoholysis reaction when the kinetics of appearance of the hydrolysis and alcoholysis products and the initial rate of formation of these products as a function of DNA concentration were compared.

Furthermore, the concentrations of integrase used in this report were within the linear part of the reaction curve for both types of assays when initial rates of product formation were compared with enzyme concentration (data not shown). Interestingly, the amount of product formed in these experiments was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected when integrase was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected. Interestingly, the amount of product formed in these experiments was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected when integrase was always less than the amount of HIV-1 and visna virus integrase present and only a low level of turnover was detected.
Nucleophile Selection by Retroviral Integrase

31. Pommier, Y., and Neamati, N. (1999) *Adv. Virus Res.* **52**, 427–458
32. Hawkins, M. E., Pfleiderer, W., Mazumder, A., Pommier, Y. G., and Balis, F. M. (1995) *Nucleic Acids Res.* **23**, 2872–2880
33. Jones, R. S., Coleman, J., Merkel, G. W., Luse, T. M., and Skalka, A. M. (1992) *J. Biol. Chem.* **267**, 16037–16040
34. Lee, S. P., Kin, H. G., Censullo, M. L., and Han, M. K. (1995) *Biochemistry* **34**, 10205–10214
35. Tramontano, E., La Colla, P., and Cheng, Y.-C. (1998) *Biochemistry* **37**, 7237–7243
36. Bushman, F. D., and Wang, B. (1994) *J. Virol.* **68**, 2215–2223
37. Ellison, V., and Brown, P. O. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7316–7320
38. Vink, C., Puras-Lutzke, R. A., and Plasterk, R. H. A. (1994) *Nucleic Acids Res.* **22**, 4103–4110
39. Yoshinaga, T., Kimura-Ohtani, Y., and Fujiwara, T. (1994) *J. Virol.* **68**, 5699–5697
40. Miller, M. D., Bor, Y.-C., and Bushman, F. (1995) *Curr. Biol.* **5**, 1047–1055
41. Yi, J., Asante-Appiah, E., and Skalka, A. M. (1999) *Biochemistry* **38**, 8458–8468
42. Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emin, E. A., Hazuda, D., and Kuo, L. C. (2000) *J. Mol. Biol.* **296**, 521–533
43. Yang, Z.-N., Mueser, T. C., Bushman, F. D., and Hyde, C. C. (2000) *J. Mol. Biol.* **296**, 535–548