Disruption of the terminal base pairs of retroviral DNA during integration

Brian P. Scottoline,1 Samson Chow,2,4 Viola Ellison,1,3 and Patrick O. Brown1,2,5

1Department of Biochemistry, and 2Howard Hughes Medical Institute, Stanford University Medical School, Stanford, California 94305-5428 USA

Integrase catalyzes two essential steps in the integration of the retroviral genome—end processing and strand transfer—both of which require the interaction of integrase with viral att sites located at the ends of viral genomic DNA. These two different polynucleotidyl transfer reactions are apparently carried out by a single active site. The end product of these reactions, the integrated provirus, does not undergo transposition and remains a stable part of the host cell genome. A central question in understanding the mechanism of integration is how a single active site accomplishes two distinct polynucleotidyl transfer reactions. We propose that integrase distorts DNA substrates to accommodate both reactions within the active site. Evidence is provided for disruption of base-pairing at the terminus of viral DNA during end processing. Furthermore, we show that this end fraying is a required step in end processing and that it appears to occur after initial binding of the viral DNA end. This requirement for base-pair disruption may account for the inability of integrase to use internal sites on DNA molecules as viral att sites. The specificity of integrase for DNA ends solves a problem posed by the long terminal repeat structure of the viral genome, and may help to prevent transposition of integrated proviruses.

[Key Words: Recombination; retroviral integration; end processing; DNA distortion]

Received October 7, 1996; revised version accepted December 17, 1996.

Analysis of integration in vivo and in vitro has revealed that the reaction proceeds in three steps [Brown et al. 1987, 1989; Fujiwara and Mizuuchi 1988; Roth et al. 1989]: [1] 3'-end processing, in which the terminal two nucleotides are cleaved from the 3' ends of the viral DNA such that the processed ends terminate with the phylogenetically conserved CA dinucleotide; [2] strand transfer, in which the recessed 3' ends of the viral DNA are joined to staggered 5' sites on opposite strands of target DNA, yielding a gapped intermediate in which the 5' ends of the viral DNA and the 3' ends of the target DNA remain unjoined; and [3] 5' end-joining, in which the gaps flanking the 5' ends of the viral DNA are repaired to seal the proviral DNA. The mechanism of this last step remains undefined. The final product of these steps is a provirus that is flanked by 4- to 6-bp repeats of the host chromosomal DNA.

In vitro studies with diverse retroviral integrases using model DNA substrates have shown that integrase alone can catalyze end processing and strand transfer [Craigie et al. 1990; Fitzgerald et al. 1990; Katz et al. 1990; Sherman and Fyfe 1990; Bushman and Craigie 1991; van Gent et al. 1991; Pahl and Flugel 1993; Katzman and Sudol 1994; Vink et al. 1994]. The reactions require no exogenous energy source. Integrase can also catalyze the reversal of the strand transfer reaction in vitro using a substrate that mimics the strand-transfer reaction intermediate, a process termed disintegration [Chow et al. 1992].

Mutational analyses of retroviral integrases have iden-
tified at least three functional domains, one of which is the central core domain, containing the D,D[35]E sequence motif. This motif is conserved among retroviral and retrotransposon integrases, as well as the transposases of some bacterial transposons (Fayet et al. 1990; Rowland and Dyke 1990; Khan et al. 1991; Kulkosky et al. 1992; Baker and Luo 1994; Polard and Chandler 1995). Mutation of the phylogenetically invariant aspartate or glutamate residues of this motif can abolish all catalytic activity [Drellich et al. 1992; Engelman and Craigie 1992; Kulkosky et al. 1992; LaFemina et al. 1992; van Gent et al. 1992; Bushman et al. 1993, Leavitt et al. 1993; Vink et al. 1993; Baker and Luo 1994] suggesting a direct role in catalysis. The parallel effects of mutations at these residues on end processing, strand transfer, and disintegration, as well as their physical proximity [Dyda et al. 1994; Bujacz et al. 1995], suggest that there is a single catalytic site per integrase protomer.

The single catalytic site in integrase must accomplish two separate reactions—end processing and strand transfer. Although both are one-step polynucleotidyl transfer reactions (Engelman et al. 1991), the two reactions are quite different in the identity and organization of the reactants relative to the attacking hydroxyl nucleophile and the phosphate groups undergoing substitution. In end processing, phosphodiester bond cleavage (most likely by water in vivo) yields the 3’ hydroxyl of the conserved subterminal A nucleotide as the leaving group, whereas in strand transfer, the same 3’ hydroxyl is the attacking nucleophile. Moreover, in end processing, the reactive phosphate is part of the viral DNA end and is subsequently removed from the end, whereas in strand transfer, the reactive phosphate is part of the target DNA and becomes joined to the viral DNA end.

The differing requirements of end processing and strand transfer make it difficult to envision a single catalytic site accommodating these two reactions if the substrates retain B-form structure. The variations in the reactions suggest a requirement for flexibility in the structure of the viral and target DNAs to accommodate these variations and to minimize the differences between the substrates and reactions. Such flexibility would accommodate similar transition states for reactions catalyzed by the integrase active site. An indication of substrate flexibility in end processing is provided by the observation that the terminal 3’ hydroxyl of the viral DNA end can, under some conditions, serve as a nucleophile in the end-processing reaction in vitro (Engelman et al. 1991; Vink et al. 1991b; Dotan et al. 1995), generating a 3’→5’ cyclic dinucleotide product. Attack by the terminal 3’ hydroxyl on the penultimate 3’ phosphodiester bond suggests disruption of base-pairing at the ends of viral DNA to allow the hydroxyl group access to the phosphate undergoing substitution.

We have tested the hypothesis that integrase disrupts base-pairing at the viral DNA end during end processing. Our results demonstrate that noncomplementarity of terminal base pairs stimulates end processing and indicate that the end-processing reaction requires disruption of the terminal base pairs [end fraying]. The requirement for disruption of base pairing at the site of cleavage may account for the ability of integrase to act specifically at the ends of viral DNA and not at identical sequences at internal sites in the viral DNA, or at the viral-target DNA border in the integrated provirus.

**Results**

**Base-pair mismatches can increase end processing**

The formation of a 3’→5’ cyclic dinucleotide product of end processing (Engelman et al. 1991) suggested that the terminal base pairs of the viral DNA end might be disrupted [frayed] before or during the chemical step of end processing, rendering the terminal 3’ hydroxyl available to attack the penultimate phosphodiester bond phosphate. This raised the possibility that integrase might facilitate base-pair disruption at the viral DNA ends as an early step in end processing. We initially tested this hypothesis by testing DNAs with mismatched ends as substrates in the integration assay (Fig. 1A). Our initial experiments involved testing substrates into which randomized single-base-pair mismatches were introduced near the terminus of the viral DNA end substrate. The mismatches were introduced by using an equimolar mixture of noncomplementary bases paired with wild-type bases to minimize the idiosyncratic effects of base identity on the reaction. These altered substrates, containing terminal base-pair mismatches, were processed by integrase at rates up to severalfold higher than their wild-type counterparts with fully complementary base pairs (Fig. 1B). The increase in processing activity was observed only if one or more of the terminal 3 bp were mismatched. We also tested model HIV-1 viral DNA end substrates that incorporated single terminal mismatches composed of an altered base complemented with a wild-type counterpart. These substrates were also processed at rates up to severalfold higher than the fully wild-type substrate (data not shown). The increase in end processing was once again observed only if the mismatches were incorporated in the terminal three base pairs.

**Effect of specific base-pair mismatches on end processing**

Although the presence of mismatches composed of a wild-type base opposite an altered base in model HIV-1 DNA end substrates resulted in an increase in end processing, the increases were not dramatic. We speculated that this was attributable to two effects operating in opposition to each other: a positive effect of the mismatch countered by the negative effect of altering a natural base. A more appropriate comparison of the effect of a mismatch on end processing would be to relate the end processing of substrates with noncomplementary base-pair substitutions to those with complementary base-pair substitutions. To distinguish the effects of mismatches between base pairs from the effects of the identity of the altered bases, each of the terminal 6 bp was
Distortion of viral DNA ends by HIV-1 integrase

**Figure 1.** Mismatches can increase end processing. (A) Diagram of the integration assay with an HIV-1 DNA end oligonucleotide substrate. Shown is the sequence of the last 4 bp of the att DNA, with the conserved CA dinucleotide underlined. The solid lines represent the rest of the oligonucleotide sequence. The substrate is labeled at the 5' end of the CA-containing strand with \(^{32}\)P. Integrase (IN) catalyzes end processing (indicated by the vertical arrow), resulting in the loss of the 3' terminal 2 nucleotides. The newly generated 3'-recessed hydroxyl serves as the nucleophile in the strand-transfer reaction, which joins the viral DNA end substrate to target DNA in a non-site-specific manner, resulting in strand-transfer products of heterogeneous lengths. The substrate serves as both donor and target. (B) Mismatches can increase end processing. Integration reaction with a wild-type substrate (lanes 1, 2) and with viral DNA end substrates incorporating terminal randomized bases to generate mismatches (lanes 3, 4). Shown are the final 4 bp of the substrate. The 5' \(^{32}\)P label. Randomized base pairs are indicated in outline. The outlined H indicates a mixture of A, C, and T bases; the outlined B indicates a mixture of C, G, and T bases. The positions of the substrate (sub), the end-processing product (-2), and strand-transfer products (st) are as indicated. Lanes 1 and 3 are reactions in the absence of integrase; lanes 2 and 4 are reactions in the presence of integrase. Below lanes 2 and 4 are indicated the end processing activities for the wild-type and the randomized mismatch mutation substrates relative to the wild-type substrate.

investigated using a set of four oligonucleotide substrates, incorporating all four combinations of two alternative bases on each strand. Thus, the strategy involved a four-element matrix of substrates for each position in the viral DNA end from the terminal base pair to the sixth base pair inward from the terminus (Fig. 2). For each base pair tested, all four substrates differed from the wild type: two were fully complementary, and two were mismatched. This approach allowed the effects of structure (matched or mismatched) to be distinguished from the effects of specific base substitutions.

Complementary base-pair substitutions at the first or second positions (relative to the terminal base pair) in the viral DNA end substrate resulted in moderate (three- to fivefold) reductions in end-processing activity when compared to the wild-type substrate (Fig. 2). Substrates with noncomplementary base-pair substitutions at the same positions had higher rates of end processing than the complementary base-pair substituted substrates from which they were derived (Fig. 2B). The magnitude of the increase varied from less than fourfold for mismatches at the terminal base pair to approximately eightfold for mismatches in the second base pair. In addition, the substrates with noncomplementary base-pair substitutions were better substrates for end processing than the wild-type substrate. The most dramatic effect of a base-pair mismatch was observed with mismatches at the third position, the site of the phylogenetically conserved A:T base pair (Fig. 2B). Complementary base-pair substitutions at this position resulted in a profound reduction in end processing, which could be rescued to near wild-type levels by a base-pair mismatch. The magnitude of the difference in end-processing activity between base-paired and mismatched substrates was from 10- to 17-fold. A favorable effect was not observed for mismatches in the fourth base pair, the position of the conserved C:G base pair [Fig. 2B]. Complementary base-pair substitutions at this site resulted in a drastic loss of end-processing activity, and base-pair mismatches at this position conferred no consistent advantage. Similarly, mismatches in base pairs 5 and 6 did not enhance end processing [Fig. 2B]. Thus, end processing was stimulated by base-pair mismatches at any of the terminal 3 bp, but mismatches at more internal sites, from the fourth base pair inward, had no apparent favorable effect.

A 3' → 5' cyclic dinucleotide is produced as a product of end processing only in the presence of Mn\(^{2+}\) and not Mg\(^{2+}\) [Engelman et al. 1991]. This raises the possibility that end-fraying does not occur when the metal ion cofactor is Mg\(^{2+}\). Because the metal ion cofactor in vivo is most likely Mg\(^{2+}\), we tested whether terminal mismatches enhanced end processing when Mg\(^{2+}\) was the divalent cation. The results were fundamentally similar to those observed in reactions performed in the presence of Mn\(^{2+}\): Mismatches in any of the terminal 3 bp increased end processing [data not shown; B. Scottoline and P. Brown, in prep.].

**Internalized viral DNA ends**

Because of the LTR structure of the retroviral genome, the terminal att sequences that serve as the substrates for end processing and joining to target DNA are duplicated at internal sites in the retroviral genome. Despite being exact duplicates of the actual att sites at the viral DNA ends, these internal copies are not used as attach-
Figure 2. End fraying enhances end processing. Mismatch matrix experiments. (A) Integration reaction with the wild-type viral DNA end substrate. [-IN] A reaction without HIV-1 integrase; [+IN] a complete reaction. The positions of strand-transfer products (st), substrate (sub), and end-processing product (-2) are as indicated. The wild-type substrate, with the terminal 6 bp shown and numbered 1–6 from the terminus inward, is diagramed below. (●) The 5'-32P label. The site at which integrase cleaves the wild-type viral end is indicated with a vertical arrow. (B) Groups of four reactions are shown for each position from the first through the sixth base pair as indicated at the upper left of each group. Each group of four is composed of reactions involving four mutant 21-mer substrates based on the wild-type substrate, of which two are fully base-paired and two are mismatched. The identity of the altered base pair is indicated above the gel image for each reaction. Mismatched bases are indicated as outlined letters; matched (but still altered) bases are normal letters. Below the gel images for each reaction is a four-element table, with one element representing each of the four reactions. The wild-type bases at each base-pair position are indicated at the upper left hand corner of the corresponding table, with the upper letter representing the base in the processed strand and the lower letter representing the base in the opposite strand. The identity of the altered bases in the processed strand is indicated above the corresponding column; the altered bases for the opposite strand are identified to the left of the corresponding row. The number in each cell represents the end-processing activity for the corresponding substrate (sub), and end-processing product (-2) are as indicated. The wild-type substrate, with the terminal 6 bp shown and numbered from the terminus inward, is diagramed below. (•) The 5'-32P label. The site at which integrase cleaves the wild-type viral end is indicated with a vertical arrow. (●) The 5'-32P label. The site at which integrase cleaves the wild-type viral end is indicated with a vertical arrow. (●) The 5'-32P label. The site at which integrase cleaves the wild-type viral end is indicated with a vertical arrow.
Figure 3. Subterminal processing sites can be cleaved if mismatched or nicked. (A) Internalized end substrates. Shown are a wild-type substrate (wt) and three groups of internalized viral DNA end substrates (a–j)—those with 5 bp beyond the wild-type terminus (a–d, +5 internalized ends), substrates with 15 bp beyond the wild-type end (e–h, +15 internalized ends), and substrates with 15 bp but with a nick in either strand at the junction of viral DNA end sequences and the additional DNA (i and j, nicked ends). The final 4 bp of the wild-type substrate are indicated, with the rest of the substrate symbolized by solid lines. For the +5 and +15 substrates with wild-type viral DNA end sequence, the final four viral DNA end base pairs are given, as is the sequence of the additional DNA, which is indicated by the dark lines above the added DNA. The position of the end-processing cleavage site for each class of substrates is indicated by a vertical arrow. For internalized-end substrates with mutations in the viral DNA end sequence, or with nicks, the additional DNA is denoted by shaded rectangles. Mutations in the viral DNA end sequence are given as outlined letters, whether mismatched or complementary. Nicks in nicked substrates are symbolized by arrows with the letter n above or below. Letters assigned to each substrate are given to the left of the substrate. [B] An integration reaction with the wild-type substrate. (-IN): Integration reaction without integrase; (+IN) integration reaction with integrase. The positions of strand-transfer products (st), substrate (sub), and end-processing products (-2) are as indicated. (C) Integration reactions with +5 internalized-end substrates (lanes a–d). Substrates in each reaction are indicated above the corresponding lane by a letter as assigned to the substrates in A. Abbreviations to denote the positions of substrates and products are the same as in A, with the addition of ep, indicating the position of the -7 product of end processing. (D) Integration reactions with +15 internalized-end substrates (lanes e–h), nicked substrates (lanes i,j), and a reaction with +15 internalized-end substrate f with Mg^2+ as cofactor [lane f/Mg]. The abbreviations to denote substrates and products are as in B and C with the exception of ep, indicating the position of the -17 product of end processing for +15 substrates [lanes e–i, and f/Mg] or the -2 product of end processing for lane j.

Substrates with or without mismatched or complementary were labeled at the 3' end of the processed strand and tested in the end-processing assay [Fig. 4, lanes 1,2] to determine the effect of mismatches on the distribution of products [Fig. 4, lanes 3–8]. Complementary base-pair substitutions at either of these positions, whether mismatched or complementary, resulted in a marked increase in the relative abundance of the cyclic dinucleotide product when compared to the ratio of products generated with the wild-type substrate [Fig. 4, cf. lanes 2 and 4 or 2 and 7]. Mismatched base-pair substitutions at the second base pair did not detectably alter the ratio of the cyclic to linear product observed with complementary base-pair substitutions [Fig. 4, lanes 7,8]. Mismatched substitutions for the third base-pair increased the ratio of the cyclic product to the linear product only slightly when compared to complementary base-pair substitutions at this position [Fig. 4, lanes 4,5].

These results indicate that end fraying per se has little
Scottoline et al.

**Figure 4.** Mismatches at the viral DNA end have little effect on the ratio of end-processing products. (Left) A diagram of the end-processing assay with an HIV-1 DNA end oligonucleotide substrate. Viral DNA is indicated by solid lines, with the terminal 4 bp of the att DNA shown. The conserved CA dinucleotide is underlined. The substrate is labeled with $^{32}$P at the terminal phosphodiester bond of the 3’ end as indicated by the outlined letter p. The short vertical arrow indicates the position of the cleavage site. Integrase (IN) catalyzes the endonucleolytic cleavage of the phosphodiester bond immediately 3’ of the CA dinucleotide (indicated by the vertical arrow), resulting in the loss of the 3’-terminal 2 nucleotides. The reaction generates multiple dinucleotide end-processing products, including a linear dinucleotide and a 3’ → 5’ cyclic dinucleotide product, as shown in the adjacent panels. (Right) Display the products of end-processing reactions using 3’ $^{32}$P-labeled wild-type and base pair-substituted complementary and mismatched viral DNA end substrates. Substrates for reactions shown in lanes 1 and 2, 3 and 4, 5 and 6, and 7 and 8, respectively, are indicated in schematic above the lanes, with only the terminal 4 bp shown. Percentages shown in adjacent panels. Lanes 1-8 display the products of end-processing reactions using 3’ $^{32}$P-labeled wild-type or mutant substrates shown in Figure 2, the apparent $K_m$ was determined for sets of substrates with complementary or mismatched base-pair substitutions at the first through fourth base pair. Figure 5 shows the results for a pair of substrates with substitutions at the third base pair. The apparent $K_m$ values for these two substrates were quite similar (4 vs. 9 nM) and also very similar to that for the wild-type substrate. In contrast, the $V_{max}$ values for reactions with mismatched substrates were greater than those for reactions using the corresponding, fully base-paired substrates. The kinetic parameters measured for the reactions using sets of paired (complementary vs. mismatched) substrates with alterations at base pairs 1, 2, and 3 were consistent with the results in Figure 5. The data from these experiments suggest that the stimulation of the end-processing rate by end fraying is...
base pairs underwent end processing more readily than substrates with complementary base pairs. Substrates with mismatches composed of a wild-type base complemented with an altered base were processed at rates which were several-fold higher than wild type, despite the presence of a mutation that would be expected to reduce end-processing activity. Mismatches of a similar type have been observed previously to have little effect on or to increase end processing (van den Ent et al. 1994). More dramatic effects of mismatching were observed if the processing of substrates incorporating noncomplementary base-pair substitutions was compared with the processing of substrates incorporating complementary base-pair substitutions. These substrates allowed the effects of base-pair mismatching to be distinguished from the confounding effects of the necessary alterations in conserved bases. Enhancements of end processing by base mismatches ranged from 4-fold for mismatches at the terminal base pair to nearly 20-fold for mismatches at the third base pair. These results are consistent with a decrease in the effect of mismatching on base-pair opening equilibria from the third base pair to the terminal base pair (Leroy et al. 1988). The increased rate observed with mismatched substrates indicates a lower energy barrier for a rate-limiting step in the reaction. We attribute this lower barrier to the facilitation of end fraying by base-pair mismatches, suggesting that disruption of base-pairing at the terminus of the viral DNA is a step in end processing. Base-pair mismatches internal to base pair 3 produced no consistent positive effect on end processing, suggesting that base pairs internal to this position are not disrupted during end processing.

Table 1. Comparison of substrate kinetic parameters for matched and mismatched substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Mutation position</th>
<th>Matched/ mismatched</th>
<th>$K_m$ app (nm)</th>
<th>$\Delta V_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGAT</td>
<td>2</td>
<td>matched</td>
<td>4</td>
<td>+5x</td>
</tr>
<tr>
<td>GTCA</td>
<td>2</td>
<td>matched</td>
<td>8</td>
<td>+5x</td>
</tr>
<tr>
<td>CAT</td>
<td>2</td>
<td>matched</td>
<td>11</td>
<td>+9x</td>
</tr>
<tr>
<td>CATT</td>
<td>3</td>
<td>matched</td>
<td>4</td>
<td>+9x</td>
</tr>
<tr>
<td>CGGT</td>
<td>3</td>
<td>mismatched</td>
<td>9</td>
<td>+9x</td>
</tr>
<tr>
<td>GACA</td>
<td>3</td>
<td>mismatched</td>
<td>36</td>
<td>-3x</td>
</tr>
</tbody>
</table>

Experiments to determine apparent $K_m$ values. ($K_m$ app) and $V_{max}$ values were performed as described in the legend to Fig. 5. The terminal 4 bp of the 21-mer viral DNA end substrates are shown; the remaining nucleotides are represented by solid lines. Mutant bases are shown in outline, and their positions are numbered relative to the viral end terminus. Data points for determination of the $K_m$ and $V_{max}$ values were the averages of three or more determinations. $\Delta V_{max}$ is the fold difference between the $V_{max}$ values measured for matched and mismatched substrates.
Kinetic experiments indicated that mismatches between any of the terminal 3 bp increased the $k_{cat}$ of the end-processing reaction (when compared to $k_{cat}$ values measured for substrates with complementary base-pair substitutions) while leaving the apparent $K_m$ virtually unaltered. Mismatches also did not affect the stability of the initial stable complex [Ellison and Brown 1994] between the substrate DNA and integrase [data not shown]. Thus, disruption of terminal base pairs appears to occur after initial binding. Facilitating terminal base-pair disruption did not significantly alter the distribution of end-processing products, indicating that disruption of base pairs adjacent to the cleavage site is essential not only for cyclic dinucleotide formation, but for end processing in general.

The role of the end in end processing
A terminal location of att sequences is critical to integration. In vivo, Moloney murine leukemia viruses (Mo-MLVs) with 10 supernumerary base pairs at one end of the viral DNA are replication defective, and the internalized att sites are not detectably processed [Colicelli and Goff 1988]. In vitro, the addition of base pairs at the termini of model HIV-1 DNA ends results in a progressive reduction in end-processing activity [B.P. Scottoline et al., this paper; Vink et al. 1991a]. Thus, some property of a viral DNA end other than the proper att site sequence is required for integration. This property appears to be fraying of the viral DNA ends. We found that substrates with internalized att sites were processed only if base-pair disruption was facilitated near the end-processing cleavage site. Mismatches in the terminal 3 bp of the internalized att site sequence, or nicks near the cleavage site, apparently reduced the increased energetic barrier to base-pair disruption associated with internal base pairs and restored accurate end-processing activity. The presence of such modifications, however, is not sufficient to support processing activity in the absence of an att sequence. Thus, end processing by integrase requires both att site sequences and base unpairing near the site of cleavage; the requirement for an end-in-end processing can thus be understood as a consequence of the lower energetic barrier to base disruption at the end of a DNA molecule [Leroy et al. 1988].

The dependence of the endonuclease activity of integrase on a DNA end provides a convenient explanation for the apparent permanence of the proviral DNA. Proviral DNA lacks only the last 2 bp of unintegrated viral DNA, which are replaced by target DNA in the provirus. The sequence of the terminal 2 bp can be altered without a drastic effect on end-processing activity [B.P. Scottoline et al., this paper; LaFemina et al. 1991; Vink et al. 1991a; Leavitt et al. 1992; Sherman et al. 1992]; thus, if the endonuclease activity of integrase did not require a DNA end, the proviral DNA might be a suitable substrate for cleavage. The resulting cleavage product would then exactly resemble the donor cleavage product in replicative transposition [Shapiro 1979], and would potentially be competent for subsequent strand transfer. Thus, the requirement for base-pair disruption around the site of end processing, and the apparent inability of integrase to stimulate this disruption away from an end, makes proviral DNA a poor substrate for integrase, and prevents proviral transposition. For the same reason, the duplicated att sites at internal LTR edges are not substrates for cleavage by integrase.

A model for a single active site catalyzing two distinct steps of the integration process
Distortion of target DNA has been suggested to play a role in the strand transfer reaction [Pryciak and Varmus 1992; Muller and Varmus 1994; Pruss et al. 1994a,b]. The evidence for DNA distortion in end processing and strand transfer suggests a means by which the active site could catalyze these two reactions, which differ in the arrangement of the phosphate to be substituted and the identity of the attacking nucleophile [Fig. 6]. Substrates in the two reactions could be distorted such that similarities between the two reactions are maximized. Thus, in the end-processing reaction, end fraying would allow the phosphate group under attack and the adjacent viral DNA to be positioned in the active site in an arrangement homologous to that of the target phosphate and surrounding target DNA during strand transfer. With similar phosphate transition states, the subsequent nucleophilic attacks could then follow similar pathways for both reactions.

Similarity with other proteins involved in polynucleotidyl transfer reactions
The effects of end fraying on the endonuclease activity of HIV-1 integrase are remarkably similar to those that have been observed with Mo-MLV integrase [B.P. Scottoline and P.O. Brown, in prep.]. This suggests that end fraying is likely to be a general feature of the retroviral integration reaction. Evidence for an altered DNA structure near the donor cleavage site has also been observed in several related polynucleotidyl transfer reactions. The bacteriophage Mu transposase protein [MuA], which shares important catalytic and structural features with integrase [Mizuuchi 1992; Baker and Luo 1994; Rice and Mizuuchi 1995], appears to cause DNA deformation at the termini of Mu DNA [Lavoie et al. 1991], and mis-matching or nicking of flanking DNA in model Mu end oligonucleotide substrates facilitates cleavage by MuA protein [Savilahti et al. 1995]. Likewise, unpairing of coding flank or signal sequence DNA near the cleavage site of the V(D)J recombinase, which catalyzes reactions that share mechanistic features with those catalyzed by integrase [van Gent et al. 1996], stimulates cleavage [Ramsden et al. 1996]. The parallels between the roles played by deformation of the double helix near the site of donor cleavage in these diverse systems suggests a common strategy, distortion of substrate DNA, for reconciling the constraints of DNA structure with the flexibility required for interaction between the active site and substrate DNA. Furthermore, integrase, MuA, and the V(D)J/
Distortion of viral DNA ends by HIV-1 integrase

Figure 6. Models for the configuration of the viral DNA end during end processing [A] and for viral DNA end and target DNA during strand transfer [B]. The diagrams show only an integrase protomer, rather than a dimer or larger multimer, to emphasize DNA configuration. The monomer is depicted with a single catalytic site, symbolized by the broken line. The phosphate groups undergoing attack are symbolized as solid circles on the DNA, which is represented in ribbon structure. [A] Viral DNA configuration during end processing. The last 3 bp of the viral DNA end are disrupted, and the scissile phosphodiester bond is positioned in the catalytic site for nucleophilic attack by solvent water (HOH) or the terminal 3' hydroxyl. Endonucleolytic cleavage releases a dinucleotide product. [B] After cleavage, the processed viral DNA end remains stably bound, and the enzyme binds target DNA. The target DNA may be distorted. In this scheme, the strand of the target DNA that becomes joined to viral DNA during the strand-transfer reaction occupies the position that was occupied by the 3' end of the viral DNA prior to end processing. The reactive phosphodiester bond in the target DNA is positioned in the active site in a manner similar to the position of the scissile phosphodiester bond in end processing. Nucleophilic attack by the recessed viral DNA end 3' hydroxyl (OH) on a target DNA phosphodiester bond phosphate accomplishes strand transfer. Thus, strand transfer resembles a reversal of the end-processing reaction.

recombinase catalyze strand transfer as well as donor cleavage; the close analogy between the reactions carried out by these polynucleotidyl transferase proteins suggests that these enzymes may employ common modes of local deformation of DNA structure to allow their active sites to catalyze both reactions.

Materials and methods

Enzymes

Wild-type HIV-1 integrase (IN) was expressed in Escherichia coli and purified as described previously [Vincent et al. 1993]. T4 polynucleotide kinase (T4 PNK) was purchased from New England Biolabs, and exonuclease-free Klenow fragment of E. coli DNA polymerase I (exo-free Klenow) was purchased from U.S. Biochemical.

Preparation of integration and end-processing assay substrates

All oligonucleotides were obtained from Operon Technologies and were purified by electrophoresis through 20% denaturing polyacrylamide gels prior to use. The standard end processing and strand transfer assay substrate (wild-type) mimicked the last 21 bp from the U5 end of HIV-1 DNA. The strand corresponding to the 3' terminus of the viral DNA [sequence 5'-ATGTTGGAAAATCTCTAGCAG-3'] was labeled at its 5' end with [γ-32P]ATP (6000 Ci/m mole, 10 μCi/μl, Amersham) using T4 PNK. The labeled strand was purified away from unincorporated label by gel filtration through two successive 1-ml bed volume Sephadex G-15 (Pharmacia) spin columns [Sambrook et al. 1989], and the recovery was quantitated by scintillation counting. The labeled strand was annealed to a 1.5-fold excess of the 21-mer complement. Other substrates for integration assays were prepared in the same manner.

To prepare substrates for assaying the dinucleotide products of 3'-end processing, 20-mer oligonucleotides corresponding to the 3' terminus of the U5 end of HIV-1 DNA, but lacking the final T of the 3' end, were annealed to 18-mer template strands that corresponded to the 5' terminus of the U5 end. The resulting double-stranded oligonucleotides were filled in with [α-32P]TTP (3000 Ci/m mole, ICN) using exo-free Klenow and then purified over two successive G-15 spin columns followed by electrophoresis through a 20% denaturing polyacrylamide gel to separate the labeled strand away from the template strand. The correctly labeled 21-mer 3' end strand was visualized by autoradiography, excised from the wet gel, eluted, quantitated by scintillation counting, and annealed to a 1.5-fold excess of the desired 21-mer complement.

Mutant oligonucleotides

Mutant viral DNA end oligonucleotides for studying end fraying were 21-mers identical to the wild-type sequence except for alterations at the positions indicated in Figure 2. These substrates were either labeled at the 5' or 3' end as described above. Internalized-end oligonucleotides were also based on wild-type and mutant 21-mer substrates but incorporated additional non-viral nucleotides at the end of viral DNA sequences. The sequence of the extensions and the identity of the mutations are as noted in Figure 3A.

Substrates for end-processing kinetics

Substrates to be used for kinetic analysis of end processing were labeled at the 5' or 3' end, annealed as described above, and then
subjected to electrophoresis through 15% native polyacrylamide gels to purify the double-stranded substrate away from excess complement. The double-stranded substrate was visualized by autoradiography, excised from the wet gel, eluted, and quantitated by scintillation counting prior to use in integration or end-processing assays.

Integration and end processing assays

Integration assays were conducted as follows, unless otherwise noted: One picomole of viral DNA end oligonucleotide substrate was incubated with 0.5 pmole HIV-1 IN protomer in a reaction buffer containing 20 mM HEPES–OH (pH 7.4), 20 mM KCl, 5 mM MmCl₂, 5 mM DTT, and 0.05% NP-40 in a reaction volume of 10 μl at 37°C for 30 min. Reactions were confirmed previously to be linear with time over the course of these reactions. For reactions with internalized-end substrates, 0.5 pmole of substrate DNA and 2 pmoles of HIV IN were used per reaction. The reactions were stopped by adding an equal volume of 95% formamide, 20 mM EDTA (pH 8.0), 0.05% xylene cyanol, and 0.05% bromophenol blue followed by heating at 95°C for 5 min. Products were resolved by electrophoresis through 25% denaturing polyacrylamide gels at 2000 V for 5.5 hr, visualized by autoradiography, and quantitated using a Molecular Dynamics PhosphorImager. Both the ~2 products of end processing and the strand transfer products were considered in determining total end-processing activity, as strand-transfer products are derived from end-processed substrates.

Assays for the dinucleotide products of the end-processing reaction, and quantitation of the products, were carried out as for the integration assay except for the electrophoresis step, in which the products of the reaction were separated by electrophoresis through 24% denaturing polyacrylamide gels for 3 hr at 2300 V. All the radiolabeled products of end processing were considered in determining total end-processing activity.

Acknowledgments

Portions of this work were presented at the Cold Spring Harbor Retroviruses Meeting, May 25–30, 1993. This work was supported by National Institutes of Health grant AI 36989 to P.O.B. and by the Howard Hughes Medical Institute. We thank Tim Heuer, members of the Varmus laboratory, and Karl Reich for helpful discussions, and Dan Herschlag, Jennifer Gerton, Ila Singh, Louise Chang, and Tim Heuer for valuable comments. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute. B.P.S. is a trainee of the Medical Scientist Training Program of the National Institute of General Medical Sciences. The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References


Brown, P.O., B. Bowerman, H.E. Varmus, and J.M. Bishop. 1989. Retroviral integration: Structure of the initial covalent prod-


van den Ent, F.M., C. Vink, and R.H. Plasterk. 1994. DNA sub-


Disruption of the terminal base pairs of retroviral DNA during integration.

B P Scottoline, S Chow, V Ellison, et al.

Genes Dev. 1997, 11: Access the most recent version at doi:10.1101/gad.11.3.371

References

This article cites 71 articles, 36 of which can be accessed free at: http://genesdev.cshlp.org/content/11/3/371.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.