We investigated which features of the substrate specificity of human immunodeficiency virus type 1 (HIV-1) integrase could be assigned to the central domain of the 288-residue HIV-1 integrase protein, composed of amino acids 50–212. This domain contains the active site and shares structural homology with a large family of polynucleotidyl transferases. Using model substrates with defined alterations in critical features we found that this domain alone is sufficient for recognition of: 1) the phylogenetically conserved CA/TG base pairs near the viral DNA end; 2) the 5′-terminal dinucleotide that is left unpaired after end processing; and 3) target DNA flanking the site of joining. Future efforts aimed at identifying specific amino acids involved in recognition of these key substrate features can now be targeted at this domain.

Integration of a double-stranded DNA copy of the retroviral genome into a host cell chromosome is essential for viral replication. Genetic and biochemical studies have shown that retroviral integration requires two viral components, integrase and the U3 and U5 att sites, which are phylogenetically conserved sequences at the ends of viral DNA. Analysis in vivo and in vitro has revealed that integrase processively catalyzes two chemical steps. In 3′-end processing, integrase cleaves two nucleotides from the 3′-ends of the double-stranded viral DNA, leaving a phylogenetically invariant CA dinucleotide sequence at the 3′-termini of the recessed viral DNA ends. In the second chemical step, integrase catalyzes the joining of the two 3′-recessed viral DNA ends to sites on opposite strands of a target DNA, separated by 5 base pairs (bps) in the case of HIV-1. The gaps that flank this product of the joining reaction are repaired by a pathway that remains uncharacterized. The chemical reactions catalyzed by integrase can be studied in vitro using model oligonucleotide substrates and purified recombinant protein.

We evaluated the specificity of the isolated core domain of HIV-1 integrase. A minimal catalytically active peptide has been identified by deletion analysis, comprised of amino acids 50–186 (1). To ensure reasonable dynamic range in our assays, we used a slightly larger, more active polypeptide, consisting of amino acids 50–212 (1). The three-dimensional structure of this domain has been solved, and it reveals striking structural homology to a larger family of polynucleotidyl transferases (2). This domain contains the phylogenetically conserved DD35E motif that defines the active site (3–8). Using model substrates with defined alterations in features known to be critical for the catalytic specificity of full-length integrase, we assessed the ability of the isolated core domain to distinguish between altered and wild-type substrates.

Integrase can carry out a concerted cleavage-ligation reaction, termed disintegration, on a model substrate that mimics the product of integration of a single viral DNA end, yielding a free cleaved viral DNA end and a ligated target DNA strand (9). A divalent metal ion is required for catalysis of disintegration as well as for end processing and joining (9–11). Disintegration substrates have proven extremely useful in studying viral DNA recognition (1, 12–14), target DNA recognition (12), catalytic requirements (12), and kinetic properties (12) of full-length and mutant integrases. The disintegration substrate allows one to analyze the catalytic properties of mutant derivatives of integrase, such as the core domain, which have defects in requirements for the forward reaction. The orientation of viral and target DNA with respect to the active site is similar for disintegration and joining. Therefore, specificity observed in the context of a disintegration substrate is likely to follow the same rules that govern specificity in the integration reaction; indeed, similar catalytic specificity for disintegration and joining has been observed (12–14).

We monitored the sensitivity of the core domain to alterations in the following substrate features: the phylogenetically conserved CA/TG bps located immediately internal to the processing site at the viral DNA end, the 5′-dinucleotide left unpaired at the viral DNA end after 3′-end processing, and target DNA flanking the joining site. The infectivity of retroviruses containing mutations in the CA/TG bps is reduced by 10−5 in vivo (15); mutations in these bps can also seriously compromise activity of model substrates in vitro (11, 16–22). The 5′-terminal dinucleotide on the unprocessed strand increases the stability of integrase-viral DNA end complexes formed in vitro and enhances processivity between the end processing and joining steps (23). Although most sites can be used at some level as integration targets, HIV-1 integrase has target site preferences; these preferences are determined both by DNA structure (24–26) and sequence (27). Domain swaps among integrases from different retroviral species indicate that the core domain, in large part, determines target site preferences as well as specificity for the species-specific features of the viral DNA end (28–30). These experiments, however, do not allow the identification of the integrase domain that recognizes the conserved features of the viral DNA end.
Substrate Specificity of the HIV-1 Integrase Core Domain

The crystal structure of the catalytic core domain of HIV-1 integrase has been solved to 2.5 Å (2). Crystallization was made possible by the discovery of a mutation, F185K, which enhances solubility but does not affect the in vitro biochemical properties of integrase, except to lower the $K_d$ of the constitutive dimer formed in vitro (31, 32). The crystal structure reveals an extensive hydrophobic interface between two monomers. The results presented here on substrate specificity, together with the biophysical and structural information available for the core domain, place new constraints on models of integrase-substrate interactions within the multimeric complex that catalyzes end processing and joining.

EXPERIMENTAL PROCEDURES

Cloning and Purification of HIV-1 Full-length Integrase

Full-length HIV-1 integrase was overexpressed in Escherichia coli using the T7 polymerase promoter system and purified as described previously (14).

Cloning and Purification of the Core Domain Polypeptide

A deletion construct containing amino acids 50–212 of HIV-1 integrase in the expression strain BL21 was a gift from Tim Jenkins. The gene contains a mutation, F185K, which increases the solubility of the protein. In vitro characterization of full-length integrase containing this mutation has demonstrated that its biochemical properties are virtually indistinguishable from the wild-type protein, with the exception that the mutation appears to allow the core domain protein to form a tighter dimer in solution (31). The construct encoding the core domain also codes for six histidines appended to the amino terminus. The hexahistidine-tagged 50–212 (F185K) polypeptide was expressed in the same manner as full-length integrase but purified using Ni-affinity chromatography. The thawed bacterial pellet from a 2-liter culture grown in Luria broth (LB) was washed with 20 mM HEPES, pH 7.5, 1 mM EDTA, then resuspended in 40 ml of lysis buffer containing 20 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 2 mM β-mercaptoethanol, 0.5 mM NaCl, and 2 mg/ml lysozyme. After 30 min on ice, the solution was sonicated with five 30-s pulses. After sonication, the cells were centrifuged at 18,000 rpm in an SS-34 rotor for 45 min at 4 °C. The pellet was resuspended in 50 ml of 20 mM Tris-HCl, pH 8.0, 1 mM NaCl, 2 mM β-mercaptoethanol (TNM) containing 5 mM imidazole, then processed 30 times in a Dounce homogenizer. This high salt extract was stirred at 4 °C for 1 h, then centrifuged at 28,000 rpm in an SW28 rotor for 1 h at 4 °C. The supernatant was loaded by gravity onto a 1-ml column of Ni4-nitrilotriacetic acid agarose (Ni-NTA resin, Qiagen) equilibrated in TNM containing 5 mM imidazole. The column was then washed with 20 column volumes of TNM containing 5 mM imidazole, then 20 column volumes of TNM containing 40 mM imidazole. Protein was eluted using a gradient of 40–600 mM imidazole in TNM. Fractions were pooled based on the Bradford assay and catalytic activity in disintegration assays. The pooled fractions were passed over a second 1-ml Ni-NTA column to eliminate contaminating nucleases. The pooled fractions were diluted 1:1 with 20 mM HEPES, pH 7.5, 10 mM dithiothreitol, and 10 mM CHAPS and incubated with thrombin at a concentration of 0.25 U/ml. One ml each of HCDG containing 250 mM NaCl, 500 mM NaCl, 750 mM NaCl, or 1 M NaCl was added sequentially, and 1-ml fractions were collected. The 50–212 polypeptide eluted in the 500, 750, and 1,000 mM NaCl steps. The column was then washed with 20 column volumes of TNM containing 5 mM imidazole, then 20 column volumes of TNM containing 40 mM imidazole. Unincorporated radioactive nucleotides were removed from the labeled oligonucleotide by centrifugation through 1-ml columns of Sephadex G-15 (Sigma). Preparation of the Y-mer substrate and structurally related substrates for disintegration assays was done as follows. The 5′-end-labeled oligonucleotide was added to a 3-fold molar excess of the other three oligonucleotides composing the Y-mer. Oligonucleotides in a solution containing 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 50 mM NaCl (TEN) were heated to 90 °C and allowed to cool slowly to room temperature. Ficoll loading buffer was added to achieve a final concentration of 2% Ficoll 400, 1 mM EDTA, 0.025% bromphenol blue, and 0.025% xylene cyanol, and the sample was then electrophoresed on a 10% nondenaturing polyacrylamide gel 16 h at 250 volts. The wet gel was autoradiographed, and the band corresponding to the completely annealed substrate was excised and eluted overnight in 0.5 mM ammonium acetate and 10 mM magnesium acetate. The supernatant fluid was concentrated in a Centricon-10 (Amicon) by centrifugation at 4 °C in an SS34 rotor at 6,500 rpm. The concentration of the gel slice, 2.5 ml of TEN was added, and the Centricon-10 was centrifuged another 2 h. The concentration of substrate was calculated based on the specific activity of the labeled oligonucleotide. The dumbbell and its related sequences were prepared by heating the oligonucleotide in TEN to 90 °C and then slowly cooling to room temperature. Oligonucleotide sequences used to make the U5 viral DNA end-Y-mer substrates were described in Chow et al. (9). The additional oligonucleotide sequences used to make the U3 viral DNA end-disintegration substrate were as follows.

**Disintegration Reaction Conditions**

All disintegration reactions were performed in solution containing 20 mM HEPES, pH 7.5, 10 mM dithiothreitol, and 10 mM MnCl2, and 0.05% Nonidet P-40. The NaCl concentration was kept at or below 10 mM except when noted in the figure legend. Reactions were stopped by the addition of an equal volume of formamide loading buffer (95% formamide, 50 mM EDTA, 0.1% bromphenol blue, and 0.1% xylene cyanol). Reactions were heated to 90 °C for 2–3 min before loading onto a 20% denaturing polyacrylamide gel. Quantitation of products was carried out with a Molecular Dynamics PhosphorImager. All reactions (except turnover experiments) were performed with 150 nm integrase and 20 nm substrate and were incubated for 30 min at 37 °C. For turnover experiments, 100 or 200 nm integrase was incubated with 1,000 nM Y-mer at 37 °C. Aliquots were removed from the reaction at the indicated time intervals and stopped by the addition of an equal volume of formamide loading buffer. The concentrations of integrase cited throughout this report refer to photoreceptors.

RESULTS

The Catalytic Core Domain Recognizes the Gross Structural Features of the Viral DNA End—Y-mer disintegration sub-
strates with altered viral DNA portions were constructed to investigate the sensitivity of the HIV-1 catalytic core domain to gross structural alterations in the viral DNA. Alterations made to disintegration substrates included: 1) replacement of the viral DNA portion with single-stranded viral DNA (Fig. 1, lanes 7–9); 2) replacement of the viral DNA with a single adenine nucleotide (Fig. 1, lanes 4–6); 3) changing the phylogenetically conserved CA/TG bps of the viral DNA end to TC/GA (Fig. 1, lanes 10–12). A substrate in which the CA/TG bps were instead changed to GT/AC was also tested, and the results were qualitatively similar to those obtained for the CA/TG to TC/GA variant (data not shown). The core domain was clearly able to distinguish these grossly altered viral DNA ends from a wild-type viral DNA end (Fig. 1, lanes 1–3). Under these reaction conditions, the core domain was more sensitive to these gross changes in the viral DNA end than was full-length integrase.

The Catalytic Core Domain Can Recognize the Conserved CA/TG Dinucleotide Pair at the Viral DNA End—To determine whether the core domain could specifically recognize the conserved CA/TG bps of the viral DNA end, we measured the catalytic activity of the core domain on Y-mer disintegration substrates with base substitutions in the phylogenetically conserved CA/TG bps of the wild-type substrate (Fig. 2A) or C/G bp (Fig. 2B). Each substrate contained one of three types of alterations: 1) two complementary mutant bases substituted for the wild-type bp; 2) two noncomplementary mutant bases substituted for the wild-type bp; or 3) one mutant base mispaired with one wild-type base substituted for the wild-type bp. This approach allowed the effects of structure (matched or mismatched) to be distinguished from the effects of specific base substitutions. The amounts of product formed in reactions with wild-type substrate are shown in the two far right lanes in Fig. 2, indicated by A/T in panel A and C/G in panel B. The pair of bases substituted for the AT bp (panel A) or the C/G bp (panel B) is specified below the column. For this analysis, 150 nM integrase and 20 nM substrate were incubated at 37 °C for 30 min. At least two nucleotide changes were tested at each of the four positions in the viral CA/TG bps.

Changing the wild-type A/T bp to the complementary mutant bp G/C or T/A resulted in a 21- or 14-fold decrease, respectively, in product formed from a disintegration substrate by the core domain. Similarly, changing the C/G bp to G/C or T/A led to a 20- or 15-fold reduction, respectively, in product formed from a disintegration substrate by the core domain. Thus, the core domain can recognize these conserved bases in the viral DNA end. Indeed, the core domain was more sensitive to alterations of these bps than was full-length integrase, which produced roughly 2-fold less product from substrates with the wild-type AT bp changed to G/C or T/A and roughly 5-fold less product from disintegration substrates in which the phylogenetically conserved C/G bp was changed to G/C or T/A.

Similar reductions in the disintegration activity of the core domain were observed when these bps were replaced by noncomplementary alternative base pairs. Replacing the wild-type AT bp with T/C or G/A resulted in a 4- or 20-fold reduction, respectively, in product formed from a disintegration substrate. Replacing the wild-type C/G bp with G/A or T/C led to a 50-fold and 6-fold reduction, respectively, in product formed from a disintegration substrate by the core domain. Although full-length integrase also showed reduced activity on these mutant substrates, the effect of the mutations was less severe (2–3-fold less product than on a wild-type substrate).
Substrates in which a wild-type base was paired with a noncomplementary base, generating a mismatch, were, in general, worse substrates than the wild-type substrate for both full-length integrase and the core domain. Once again, the viral DNA end mutations had a more severe effect on the activity of the core domain than on full-length integrase. The core domain produced 4–100-fold less product from mismatched substrates in which the A, C, or G bases were altered and mispaired with a wild-type base (the A/T bp was changed to G/T and T/T, the C/G bp was changed to G/G, T/G, C/C, and C/A) than on a wild-type substrate. For full-length integrase, identical mutations in the substrate resulted in up to an 8-fold decrease in product. One subset of mismatched mutant substrates actually proved to be slightly better substrates than wild-type substrates. The core domain and full-length integrase both generated slightly more product from substrates containing a wild-type A base mispaired with a C (panel A, A/C) or A base (panel A, A/A) than from a wild-type substrate (panel A, A/T).

It has been demonstrated previously that incorporating mismatches at the A/T position in a viral DNA end substrate can facilitate end processing in vitro, strongly suggesting that fraying of the terminal bps of the viral DNA precedes cleavage (22). Mismatches at this position resulting from substitutions for the thymine residue in disintegration substrates lower the energetic barrier to fraying, enhancing disintegration of these substrates by both the core domain and full-length integrase. In contrast, mismatches at the A/T bp caused by substitutions for the adenosine residue did not enhance disintegration. Presumably, the favorable effect of the fraying did not outweigh the deleterious effect of the lack of base-specific interactions with the adenosine residue.

In all cases (with the exception of substrates containing a wild-type A base mispaired with a mutant base), the mutant disintegration substrates were poorer substrates than the wild-type substrate for the core domain protein, indicating that features of integrase capable of recognizing the phylogenetically invariant CATG bps are contained in the core domain. Furthermore, the core domain was more sensitive to alterations in the conserved A/T and C/G bps than was full-length integrase, despite their similar activity on wild-type substrates (either a U5 or U3 viral DNA end; data shown for the U5 end only). The hypersensitivity of the core domain to mutations in these bps may be a result of the loss of the nonspecific DNA binding activity of the COOH terminus (33–35), making the core domain more dependent on the remaining specific contacts with the CATG bps.

The Isolated Core Domain Turns Over Faster Than Full-length Protein on a Disintegration Substrate—The disintegration activity of the core domain of HIV-1 integrase has been reported to be lower than that of full-length integrase (1). We attribute this discrepancy with our results to the different reaction conditions used. NaCl and MnCl₂ titrations revealed that the isolated core domain was extremely sensitive to the ionic milieu of the reaction. Concentrations of NaCl or MnCl₂ above 10 mM caused a precipitous drop in the activity of the core domain (data not shown). Full-length HIV-1 integrase, in contrast, could catalyze disintegration under reaction conditions containing up to 100 mM NaCl or 100 mM MnCl₂. Electrostatic interactions thus appear to play a critical role in the residual DNA binding activity of the core domain. When disintegration reactions were carried out in solutions containing 10 mM NaCl or less, the core domain often produced slightly more product than full-length integrase, under standard conditions of enzyme excess. This observation prompted the comparison of turnover in disintegration reactions catalyzed by the two proteins.

The core domain can turn over faster than full-length integrase in reactions using a wild-type Y-mer disintegration substrate (Fig. 3); the $k_{cat}$ of the core domain was 1.5 h⁻¹, whereas the $k_{cat}$ of full-length integrase was 0.26 h⁻¹, under reaction conditions favorable to the activity of the core domain. Aliquots were taken at indicated time intervals from reactions containing 1,000 nM wild-type disintegration substrate and 200 nM protein. Each point plotted in Figs. 3 and 4 represents the average of two independent determinations. The apparent difference in activity may be an underestimate because the concentration of NaCl in these reactions (16 mM) was sufficient to inhibit partially the activity of the core domain but not that of full-length integrase. Qualitatively similar results were obtained in a similar experiment carried out with 100 nM protein and 1,000 nM substrate (data not shown). (The disintegration substrate was stored in 50 mM NaCl to prevent dissociation of the strands; the large volume of substrate added in these reactions resulted in the relatively “high” concentration of NaCl.)

The Core Domain Can Recognize the Unpaired Dinucleotide at the Viral DNA 5′-End—The unpaired dinucleotide, left at each 5′-end of viral DNA after end processing, is important for stability of the complex between integrase and viral DNA (23). The unpaired dinucleotide apparently allows the enzyme to remain stably associated with the viral DNA end after processing, such that end processing and joining can be executed processively. We investigated whether the core domain could take advantage of this 5′-dinucleotide to stabilize its interaction with viral DNA. A comparison of turnover on a Y-mer disintegration substrate without the 5′-dinucleotide versus a substrate with the 5′-dinucleotide revealed that turnover was faster when substrates lacked the 5′-dinucleotide for both full-length integrase (Fig. 4A) and the core domain (Fig. 4B). The $k_{cat}$ of the core domain in reactions using a synthetic disintegration substrate that lacked this dinucleotide was 7.5 h⁻¹ compared with 1.5 h⁻¹ for the reaction with a substrate that
absence (indicated by 1)

Substrates were identical except for the presence (indicated by 2) of 200 nM core domain or full-length integrase at 37 °C. Aliquots were removed at 10, 20, 30, 45, 60, 90, 120, 180, 240, and 300 min. Duplicate samples were quantitated using a PhosphorImager and averaged. The NaCl concentration was maintained at 5 bps), we found that a target DNA portion of 10, 14, 18, 22, or 30 bps (while the viral DNA portion was kept at 19 bps (Fig. 5, lane 13 with the same efficiency as full-length integrase (Fig. 5, lane 14), but its activity relative to that of full-length integrase was reduced when the length of the target DNA portion was shortened to 10 bps while that of the viral DNA portion was kept at 19 bps (Fig. 5, lanes 16 and 17).

By using disintegration substrates with target DNA portions of 10, 14, 18, 22, or 30 bps (while the viral DNA portion was maintained at 5 bps), we found that a target DNA portion of 25813

FIG. 4. Turnover rate in disintegration reactions catalyzed by full-length HIV-1 integrase or the core domain is increased in the absence of the 5′-dinucleotide of the viral DNA end. 1,000 nM Y-mer disintegration substrate was incubated with 200 nM core domain or full-length integrase at 37 °C. Aliquots were removed at 10, 20, 30, 45, 60, 90, 120, 180, 240, and 300 min. Duplicate samples were quantitated using a PhosphorImager and averaged. The NaCl concentration in the reactions was 16 mM. The x axis represents time in minutes, and the y axis represents the percent of substrate converted to product. Qualitatively similar results were obtained with 100 nM integrase. Substrates were identical except for the presence (indicated by +) or absence (indicated by −) of the 5′-dinucleotide. Panel A: ●, core domain, Y+; ○, core domain, Y−. Panel B: ○, full-length integrase, Y+; ●, full-length integrase, Y−.

carried the terminal dinucleotide. The $k_{cat}$ of full-length integrase in reactions using a substrate without the 5′-dinucleotide was 0.86 h$^{-1}$ compared with 0.26 h$^{-1}$ for reactions that used a substrate with this dinucleotide. These results are consistent with a model in which this substrate feature contributes interactions with the core domain which stabilize the complex between integrase and viral DNA.

The Core Domain Makes Extensive Interactions with Target DNA—The core domain can catalyze disintegration of a Y-mer substrate but not a dumbbell substrate. The standard Y-mer substrate models a 19-bp terminal viral DNA sequence joined to a 30-bp segment of target DNA. The standard dumbbell substrate is smaller, consisting of one oligonucleotide folded on itself to form a dumbbell-type structure. The substrates in the reactions analyzed in lanes 19–21 contained the standard wild-type Y-mer disintegration substrate. The slower mobility intense bands in lanes 1–18 correspond to substrate, and the faster mobility bands correspond to product. In lanes 19–21, the faster mobility intense bands correspond to substrate, and the slower mobility intense bands correspond to product. Oligonucleotides are described further under “Experimental Procedures.”

FIG. 5. Disintegration by the core domain and full-length integrase is influenced by the lengths of flanking target DNA. 150 nM core domain (lanes 1, 4, 7, 10, 13, 16, 19), 150 nM full-length integrase (lanes 2, 5, 8, 11, 14, 17, 20), or no integrase (lanes 3, 6, 9, 12, 15, 18, 21) was incubated with 20 nM substrate for 30 min at 37 °C. Reactions with substrates with 5 bps of viral DNA were analyzed in lanes 1–15. Reactions with substrates with 19 bps of viral DNA were analyzed in lanes 16–21. The substrates in the reactions analyzed in lanes 1–3 and 16–18 had 10 bps of target DNA. The substrates in the reactions analyzed in lanes 4–6, 7–9, and 10–12 had 14, 18, or 22 bps of target DNA, respectively. The substrates in the reactions analyzed in lanes 13–15 and 19–21 had 30 bps of target DNA. The substrates in the reactions analyzed in lanes 1–18 were each composed of a single oligonucleotide that folded on itself to form a dumbbell-type structure. The substrates in the reactions analyzed in lanes 19–21 contained the standard wild-type Y-mer disintegration substrate. The slower mobility intense bands in lanes 1–18 correspond to substrate, and the faster mobility bands correspond to product. In lanes 19–21, the faster mobility intense bands correspond to substrate, and the slower mobility intense bands correspond to product. Oligonucleotides are described further under “Experimental Procedures.”

The Core Domain Makes Extensive Interactions with Target DNA—The core domain can catalyze disintegration of a Y-mer substrate but not a dumbbell substrate. The standard Y-mer substrate models a 19-bp terminal viral DNA sequence joined to a 30-bp segment of target DNA. The standard dumbbell substrate is smaller, consisting of one oligonucleotide folded on itself to form a similar structure, in which the viral DNA end is itself to form a similar structure, in which the viral DNA end is
more than 22 bps was needed for maximal activity of the core domain (Fig. 5, compare lanes 1, 4, 7, and 10 with lane 13 (30-bp target DNA), full-length integrase in lanes 2, 5, 8, 11, and 14). Disintegration of the standard Y-mer substrate by the core domain (Fig. 5, lane 19) and full-length integrase (Fig. 5, lane 20) under identical reaction conditions is shown for comparison. This result indicates that the core domain was able to make nonspecific contacts with target DNA and that these contacts extended more than 11 bps from the site of joining.

Full-length integrase needs 5 bps of target DNA 5' to the site of joining in the dumbbell substrate, but the target DNA 3' to the site of joining can be shortened to 2 bps without compromising disintegration activity (12). Corresponding disintegration substrates with asymmetric target DNA were made with 15 bps on one arm of target DNA and 2 bps on the other arm (and 5 bps of viral DNA). Full-length integrase could mediate disintegration of a substrate with a short arm 3' to the joining site (Fig. 6, lane 5) but not a substrate with a short arm 5' to the joining site, as expected (Fig. 6, lane 2). The core domain could not disintegrate either "one-armed" substrate to a detectable level (Fig. 6, lanes 1 and 4). The removal of substrate parts with which integrase interacts, whether specific interactions with the CA/TG bps or relatively nonspecific interactions with the target DNA, appears to be generally much less well tolerated by the core domain than by full-length integrase. This may reflect a deficiency in DNA binding on the part of the core domain. Consistent with this type of defect, the core domain was unable to form a stable complex with a disintegration substrate (data not shown).

**DISCUSSION**

A CA/TG dinucleotide pair is invariably found 2 (or rarely 3) bases from each end of the unintegrated retroviral DNA molecule. Mutations in these bps result in a greater than 10^-5-fold reduction in integration (15); the proviruses arising from remaining "integration" events do not have the hallmark features of proviruses generated by integrase (36). In *in vitro*, alterations in these bps can reduce end processing activity to nearly undetectable levels (11, 16–22). The experiments described in this report show that the disintegration activity of the core domain is inhibited by mutations in either of these bps, clearly demonstrating that the specificity of full-length integrase for these bps is conferred by the core domain. The elevated sensitivity of the core domain compared with full-length integrase to alterations in the sequence of the viral DNA end probably results from the loss of the nonspecific DNA binding interactions normally provided by the COOH-terminal domain (33–35), making the core domain more dependent on the remaining specific protein-DNA contacts with the CA/TG bps.

It has been previously suggested that integrase might play a role *in vivo* in repairing the gaps that flank the viral DNA after the 3' -ends are joined to a host chromosome (9, 37). Kulkosky et al. (37) found evidence that integrase had sequence specificity in an *in vitro* reaction in which the two unpaired 5'-nucleotides on the viral DNA end were cleaved. As this reaction involves a bond on the strand complementary to the strand that is processed and joined to host DNA by integrase, the sequence specificity observed in this reaction reflects a different mode of viral DNA binding than that involved in the end processing and DNA joining reactions. The core domain is much more sensitive to mutations in the conserved CA/TG bps in the context of the disintegration substrate than in the context of the 5'-dinucleotide cleavage substrate studied by Kulkosky et al. (37), further suggesting that a different set of interactions provides specificity for the two substrates. Although the *in vivo* relevance of 5'-dinucleotide cleavage remains to be determined, viral and target DNA are similarly oriented with respect to the active site in disintegration and DNA joining, and these two reactions show parallel sequence requirements (12–14), arguing that similar protein-DNA contacts mediate specificity in both cases. Thus, the results of experiments with disintegration substrates are likely to reflect the same interactions that define specificity in the joining reaction.

The principles underlying substrate recognition by the phylogenetically conserved core domain of HIV-1 integrase may apply to a broader superfamiliy of polynucleotidyl transferases (38). Other members of this family of enzymes, including the bacteriophage Mu transposase (39), do not share extensive homology at the amino acid level but have very similar folded structures, and contain a DD35E motif that appears to define residues responsible for catalysis (38). Mu transposase, like HIV-1 integrase, possesses a cleavage activity that is specific for a CA/TG dinucleotide sequence located at the ends of the element, and as well, substrate distortion near the site of cleavage enhances the cleavage activity of both enzymes (22, 40). As the core domain of HIV-1 integrase appears to provide the specificity for the CA/TG sequence and structure, and this specificity is likely to be determined by phylogenetically conserved residues near the active site residues,³ it will be interesting to discover whether similar interactions determine the corresponding specificity of other transposases.

Although the core domain was able to recognize key substrate features, its DNA binding activity was weaker than that of full-length integrase (33). The isolated core domain was much less tolerant of mutations in the conserved CA/TG bps at the viral DNA end than was full-length integrase. Moreover, the activity of this domain was highly dependent upon nonspecific contacts with the target DNA, presumably reflecting its inability to exploit DNA binding interactions normally provided by the COOH-terminal domain (33–35) or by other subunits in a higher order integrase multimer (41–43). The ex-
Substrate Specificity of the HIV-1 Integrase Core Domain

extreme sensitivity of the isolated core domain to the ionic milieu of the reaction may reflect an increased dependence on electrostatic interactions for DNA binding by the core domain. Consistent with the notion that the overall DNA binding activity of the core domain is impaired relative to full-length integrase, the core domain was unable to form a stable complex with a disintegration substrate (data not shown). Moreover, the ability of the core domain to cross-link to a disintegration substrate has been reported to be entirely dependent on the presence of a divalent metal ion (33).

Under multiple turnover conditions, the steady-state rates of disintegration measured for both full-length integrase and the core domain were higher for substrates lacking the 5′-dinucleotide. The results of the turnover experiments are most easily explained by differences in product dissociation. If interactions with the 5′-dinucleotide act to stabilize enzyme-viral DNA complexes, and product dissociation is a rate-limiting step for turnover, then the presence of this substrate feature would be expected to slow dissociation of the enzyme from the viral DNA end product of the reaction, reducing the steady-state rate.

These results thus provide further evidence that this substrate feature plays a key role in stabilizing the interaction between integrase and viral DNA and demonstrate that this stabilizing interaction is mediated, at least in part, by the core domain. The core domain may turn over faster than full-length integrase because its deficit in DNA binding activity allows it to dissociate more rapidly from the products of the reaction. Although the 5′-dinucleotide is important for stable association between the viral DNA end and integrase, this feature must not be critical for initial binding (23).

Although most sites in a DNA target can be used to some extent for integration, they are not all used equally; bent DNA has been shown to be a preferred integration target (27, 44), as have sites flanking the exposed major groove on nucleosomes, particularly at the sites of the most severe bending (24–26). Because we were unable to detect any joining products with the core domain using a sensitive polymerase chain reaction-based assay (24), we could not examine directly the target site preferences of the isolated core domain. Nonetheless, our observation that the activity of the core domain was dependent on target DNA flanking the junction with viral DNA attachment is consistent with a role for this domain in target DNA recognition. Studies of chimeric proteins in which the amino and carboxyl termini of one integrase were appended to the core domain of another integrase have demonstrated that target site preferences are, in large part, determined by the core domain (28–30).

The critical role of integration in the retroviral life cycle makes integrase an excellent target for antiviral therapy. Substrate specificity is a critical feature of the in vivo integration reaction. Understanding the basis of the substrate specificity of integrase may contribute to the development of new therapeutic approaches to HIV infection. For example, an agent that caused relaxation of the normal specificity of the cleavage activity of integrase could result in irreparable damage to viral replication intermediates (45). The localization of interactions critical to the substrate specificity of integrase now focuses our attention on the core domain as we seek to understand these interactions and perhaps develop ways to thwart them.