Mapping Features of HIV-1 Integrase Near Selected Sites on Viral and Target DNA Molecules in an Active Enzyme−DNA Complex by Photo-Cross-Linking†

Timothy S. Heuer‡,§ and Patrick O. Brown*,‡,||

Department of Biochemistry, Program in Cancer Biology, and Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305-5428

Received April 4, 1997; Revised Manuscript Received June 5, 1997⊗

ABSTRACT: The virally encoded integrase protein carries out retroviral integration, and to do so, it must make specific interactions with both viral and target DNA sequences. The retroviral integrase has three domains: an amino-terminal region of about 50 amino acids that contains a zinc finger-like motif, a tightly folded, phylogenetically conserved core domain that contains the active site, and a carboxy-terminal domain that can bind DNA in a nonspecific manner. The complete roles of the amino- and carboxyl-terminal domains have not yet been determined, but they appear to participate in multimerization and nonspecific or target DNA binding, respectively. The number and identity of integrase's DNA binding sites have been difficult to determine by conventional mutagenesis studies. In this report, we describe a photo-cross-linking approach to address these issues. Our findings suggest that HIV-1 integrase contacts with conserved features of the viral DNA end are likely to be mediated by residues in two peptides within the conserved core domain. Additional cross-links were seen between viral DNA and the carboxy-terminal DNA binding domain. Numerous sites in integrase, including peptides in each of the three domains, could be cross-linked to target DNA features. Integrase is known to function as a multimer, and it remains to be determined which specific contacts are in *cis* or *trans* with respect to the active site.

Integration of retroviral DNA into the host cell genome is a prerequisite for a productive retroviral infection (Varmus & Brown, 1989). Genetic studies have demonstrated that two viral components are required for retroviral integration: (i) the protein product of the 3′ portion of the retroviral *pol* gene, integrase (Donehower & Varus, 1984; Panganiban & Temin, 1984; Schwartzberg et al., 1984), and (ii) viral DNA attachment sites for integrase contained within terminal viral DNA sequences (Colicelli & Goff, 1985). Retroviral integration consists of three discrete steps. First, endonucleolytic processing of the viral DNA removes two terminal nucleotides 3′ to a conserved CA dinucleotide. Second, joining of the recessed viral DNA 3′ ends, to staggered 5′ ends of the host cell target DNA, generates a gapped recombination intermediate. Finally, repair of the gapped intermediate creates an integrated provirus that is flanked by direct repeats of the insertion site sequence (Brown et al., 1987, 1989). The role of integrase in repair of the gapped intermediate remains to be investigated. The first two steps in the integration pathway, namely 3′ end processing and joining of viral DNA to host DNA, have been shown to require integrase. Integrase is sufficient to mediate these steps in model studies *in vitro* (Brown et al., 1989; Craigie, et al., 1990; Fujiwara & Craigie, 1989; Fujiwara & Mizuuchi, 1988; Katz et al., 1990; Roth et al., 1989).

Integrase mediates viral DNA end processing, and joining to host DNA, using a single active site (Engelman et al., 1991; van Gent et al., 1992), and a one-step in-line transesterification mechanism (Engelman et al., 1991). Pre-integration complexes isolated from virus-infected cells remain integration competent through multiple steps of purification (Bowerman et al., 1989; Brown et al., 1987), indicating that *in vivo* integrase is bound very stably to viral DNA ends. *In vitro*, the terminal nucleotides at the 5′ end of model viral DNA substrates contribute to the stability of the complex with purified integrase between the 3′ end processing and DNA joining steps (Ellison & Brown, 1994). The protein−DNA interactions through which integrase mediates stable complex formation are not yet understood. Indeed, it is not yet known whether integrase binds viral and target DNA molecules through single or multiple DNA binding sites. The simplest interpretation of kinetic data suggests that integrase might have distinct binding sites for viral and target DNA sequences (Dotan, et al., 1995; Ellison & Brown, 1994). The apparent inability of integrase to bind and discriminate efficiently between viral and nonviral DNA sequences *in vitro* (van Gent et al., 1991) has been interpreted as supporting a single-DNA binding site model. How does integrase recognize and bind viral and target DNA sequences to form an active integration complex?

Many of the viral DNA sequences that are important for processing and joining of viral DNA ends have been identified through studies using viral end substrates with altered sequences (Bushman & Craigie, 1990; Lafemina et al., 1991; Sherman et al., 1992; Vink et al., 1991a; Yoshinaga & Fujiwara, 1995) or that have been chemically modified (Bushman & Craigie, 1992). Integrase can catalyze the reverse of the integration reaction (disintegration) *in vitro* (Chow et al., 1992), and investigation of the disintegration

† This work was supported by the Howard Hughes Medical Institute and NIH Grant AI36893. T.S.H. is supported by a predoctoral training grant from the National Institutes of Health provided by the Program in Cancer Biology at Stanford University.

* Corresponding author. Phone: (415) 723-0005. Fax: (415) 723-1399. E-mail: pbrown@cmgm.stanford.edu.

‡ Department of Biochemistry.

§ Program in Cancer Biology.

|| Howard Hughes Medical Institute.

⊗ Abstract published in *Advance ACS Abstracts,* August 1, 1997.

S0006-2960(97)00782-4 CCC: $14.00

reaction has identified both viral and target DNA features recognized by integrase (Chow & Brown, 1994). Mapping the features of the integrase polypeptide that mediate important viral and target DNA interactions has proven to be more difficult.

Mutational analyses have distinguished three functionally distinct regions of integrase, two of which are characterized by amino acid motifs conserved among retroviruses and retrotransposons (Engelman & Craigie, 1992; Khan et al., 1991; Kulkosky et al., 1992). The first region comprises the amino-terminal amino acids of integrase proteins and contains a conserved zinc finger-like motif, HX_{3-7}-$HX_{23-32}CX_2C$. This region is often termed the HHCC domain, and mutations within the HHCC region cause the loss of 3′ end processing and integration activities but do not impair the ability to perform disintegration (Engelman & Craigie, 1992; Leavitt et al., 1993; Vink et al., 1993; Vincent et al., 1993), suggesting that the HHCC domain may participate in viral DNA recognition and/or binding. Integrase proteins with a disrupted HHCC domain retain the ability to recognize important features of viral DNA ends, including the conserved CA dinucleotide (Vincent et al., 1993), and indicate that this region is not the primary viral DNA recognition/binding domain. More recent studies have suggested that the HHCC region may participate in homomeric protein−protein interactions necessary for the formation of integrase oligomers (Ellison, et al., 1995; Zheng et al., 1996).

The second conserved and functionally distinct region of integrase proteins lies within a central protease-resistant core and contains the motif $DX_{39-58}DX_{35}E$. Mutation of the aspartate or glutamate residues within this triad, which is evolutionarily conserved among bacterial transposases, retrotransposases, and retroviral integrases, causes a virtual loss of all catalytic activities, strongly suggesting a direct role for these residues in catalysis (Engelman & Craigie, 1992; Leavitt et al., 1993; van Gent et al., 1992). Truncation of both the amino- and carboxyl-terminal regions of integrases has demonstrated that the central core region of integrase alone can mediate disintegration *in vitro* (Bushman et al., 1993; Vink et al., 1993) and shows that residues within the core region also participate in DNA binding. Indeed, it is the integrase core region that provides specificity for the conserved CA dinucleotide (J. Gerton and P. Brown, unpublished results). Core residues can also contribute to nonspecific DNA binding (Engelman et al., 1994).

The carboxyl-terminal region of integrase does not contain any conserved motifs that might point to its function. However, the analysis of mutant proteins truncated at the carboxyl terminus has shown that this region contains a nonspecific DNA binding domain (Engelman et al., 1994; Khan et al., 1991; Lutzke et al., 1994 Vink et al., 1993).

While the above studies have been very informative, we still do not have a clear picture of how retroviral integrase proteins recognize and bind to specific viral DNA end sequences, and with little sequence specificity to target DNA sequences. Toward this end, we have developed an *in vitro* UV cross-linking procedure that generates covalent catalytically active integrase−DNA complexes. This procedure has provided a powerful tool for studying functional integrase−DNA interactions and has allowed us to map viral and target DNA cross-links to specific regions of the HIV-1[1] integrase polypeptide.

EXPERIMENTAL PROCEDURES

Enzymes. HIV-1 integrase proteins were expressed in *Escherichia coli* using the T7 polymerase expression system, and purification of the wild type protein was as described previously (Vincent et al., 1993). Each mutant integrase protein was expressed with an amino-terminal six-histidine tag. After purification by nickel−NTA chromatography, the histidine tag was removed by proteolytic cleavage with thrombin (Cal Biochem). The storage buffer for all integrase proteins consisted of 20 mM HEPES (pH 7.5), 10 mM DTT, 1 mM EDTA, 300 mM NaCl, 10 mM CHAPS, and 10% glycerol. Endo-Glu-C protease was purchased from Worthington Enzymes Inc. T4 polynucleotide kinase was purchased from New England Biolabs, and exonuclease-free Klenow fragment of *E. coli* DNA polymerase I was from United States Biochemical Corp.

Oligonucleotides. DNA oligonucleotides were purchased from Operon Technologies Inc. and purified by electrophoresis through a denaturing polyacrylamide gel. The following oligonucleotide cross-linking substrates contained a phosphorothioate linkage at the single position that is 3′ to the base denoted in lowercase: V1HdbS (5′-aCTGCTAGTTCTAGCAGGCCCTTGGGCCGGC-GCTTGCGCC), V2HdbS (5′-ACTGCTAGTTCtAGCAG-GCCCTTGGGCCGGCGCTTGCGCC), V3HdbS (5′-ACTGCTAGTTCTAgCAGGCCCTTGGGCCGGCGCTT-GCGCC), V4HdbS (5′-ACTgCTAGTTCTAGCAGGCCCT-TGGGCCGGCGCTTGCGCC), V5dbPS (5′-thio-pACT-GCTAGT), T5HdbS3 (5′-ACTGCTAGTTCTAGCAGGC-CCTTGGgCCGGCGCTTGCG), T5HdbS (5′-ACTGCTAG-TTCTAGCAGGCCCTTGGgCCGGCGCTTGCGCC), T6HdbSC2 (5′-TGCTAGTTCTAGCAGGCCGCAGGTCT-TGACCTGCGGCCgGCGCTTGCG), T1HdbC1S (5′-CG-CAAGCgCC), and T4HdbC1S (5′-CGCaAGCGCC).

The following oligodeoxynucleotides were not modified: V5-splint (5′-TGCTAGAACTAGCAGT), V5dbP2 (5′-TCTAGCAGGCCCTTGGGCCGGCGCTTGCGCC), HdbC1 (5′-CGCAAGCGCC), and HdbC2 (5′-TGCTAGT-TCTAGCAGGCCGCAGGTCTTGACCTGCGGCCGGC-GCTTGCG).

The following oligonucleotides contained a 5-iodo-substituted deoxycytidine at the nucleotide denoted in lowercase: VHdbI (5′-ACTGCTAGTTCTAGcAGGCCCT-TGGGCCGGCGCTTGCGCC) and THdbI (5′-CGCAAG-CGcC).

Preparation of Azidophenacyl Viral DNA End Cross-Linking Substrates. Each of the oligonucleotides V1HdbS, V2HdbS, V3HdbS, V4HdbS, and V5HdbP2 was 5′ end labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. Unincorporated radiolabeled nucleotides were removed by passing the sample through a 1 mL G-15 Sephadex spin column. Oligonucleotide V5HdbPS was synthesized with a thiophosphate at the very 5′ end of the oligonucleotide. V5HdbPS was ligated to the radiolabeled V5HdbP2 oligonucleotide to generate a full-length dumbbell substrate oligonucleotide, which was radiolabeled within the loop of the viral DNA end hairpin. Azidophenacyl was coupled to the phosphorothioate-containing oligonucleotides in a reac-

[1] Abbreviations: AZP, azidophenacyl; CHAPS, 3[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; DTT, dithiothreitol; EDTA, (ethylenedinitrilo) tetraacetic acid; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; HIV-1, human immunodeficiency virus type 1.

tion buffer consisting of 20 mM $NaHCO_3$ (pH 9.0), 40% methanol, and 10 mM azidophenacyl bromide (Sigma). The substrate concentration varied but did not exceed 1 μM. A total of 250−425 nmol of substrate was prepared for each peptide sequencing experiment in four 1 mL reaction mixtures, and the specific activity of substrates was between 50 and 3000 cpm/pmol. The reaction volume for analytical experiments was 100 μl, and the specific activity of substrates was about 5×10^5 cpm/pmol. The azidophenacyl bromide was initially dissolved in 100% methanol and subsequently added to the reaction mixture, which proceeded at 50 °C for 10 min followed by 50 min at room temperature. Excess azidophenacyl bromide was removed by phenol−chloroform extraction and ethanol precipitation of the DNA. The photoreactive substrate was resuspended in TE buffer [10 mM Tris-HCl; (pH 7.4) and 1 mM EDTA] containing 50 mM NaCl (TEN_{50}). The dumbbell substrates were annealed by heating at 80 °C for 2−3 min and then by slowly cooling to room temperature. All manipulations of the photoreactive DNA substrate, including and subsequent to azidophenacyl coupling, were performed in minimal light (residual ambient light with window shades closed and room lights turned off). Approximately 50−70% of the input DNA substrate was typically modified by azidophenacyl bromide as measured by denaturing polyacrylamide gel electrophoresis.

Preparation of Azidophenacyl Target DNA Cross-Linking Substrates. Dumbbell substrate oligonucleotide T5HdbS3 was 3′ end labeled with [α-^{32}P]dCTP (Amersham), using exonuclease-free Klenow fragment of *E. coli* DNA polymerase I at 30 °C for 60 min. To ensure that the 38-nucleotide substrate was extended to the 40-nucleotide full-length substrate, the reaction was extended for an additional 60 min in the presence of a 100-fold molar excess of unlabeled dCTP. Hybrid-dumbbell oligonucleotides T1HdbC1S and T4HdbC1S were 5′ end labeled with [γ-^{32}P]-ATP using T4 polynucleotide kinase. Unincorporated nucleotides were removed from all reaction mixtures by a G-15 sephadex spin column. Oligonucleotides T5HdbS and HdbC2 were not radiolabeled. Azidophenacyl was coupled to oligonucleotides that contained a phosphorothioate as described above. The dumbbell substrates were annealed by heating at 80 °C for 2−3 min and then by slowly cooling to room temperature. Duplex substrates (T4HdbC1S/HdbC2 and T1HdbC1S/HdbC2) were annealed by heating at 80 °C for 3 min and slowly cooling to room temperature after adding a 1.3-fold molar excess of the complementary oligonucleotide (HdbSC2).

Preparation of 5-Iodo Cross-Linking Substrates. Four to five nanomoles of each oligonucleotide VHdbI and THdbI was 5′ end labeled with 800 μCi [γ-^{32}P]ATP using 10 units of T4 polynucleotide kinase. Unincorporated radiolabeled nucleotides were removed by passing the sample through a 1 mL G-15 Sephadex spin column. Unlabeled oligonucleotide was added to the samples so that there was 200−250 nmol present. For the duplex substrate THbdI/HdbC2, a 1.3-fold molar excess of HdbC2 was included. The specific activity of each substrate was about 1000 cpm/pmol. After the NaCl concentration was brought to 50 mM, the oligonucleotides were annealed by heating each sample at 80 °C for 3 min and then slowly cooling to room temperature.

Analytical Cross-Linking. Cross-linking reactions (20 μL volume) were performed in 1.5 mL microcentrifuge tubes with the cap removed. The reaction buffer consisted of 20 mM HEPES (pH 7.5), 10 mM DTT, 20 mM NaCl, 0.02% NP-40, and either 10.0 mM $MnCl_2$, 7.5 mM $MnCl_2$, 2.5 mM $MnCl_2$, or 0.1 mM EDTA. Reaction mixtures where integrase was present in molar excess over the DNA substrate contained substrate and integrase at concentrations of 20 and 200 nM, respectively. Reaction mixtures where substrate was present in molar excess contained substrate and integrase at concentrations of 720 and 180 nM, respectively. Reactions were initiated by the addition of integrase, and mixtures were incubated in darkness for 3 min on ice. Reaction mixtures were incubated for an additional 4 min either on ice, at 30 °C, or at room temperature and were irradiated with a 4 W, 300 nm, UV light source (Fotodyne) at a distance of 6 cm. A 1 L pyrex beaker was placed between the samples and the UV lamp. Reaction mixtures subsequently incubated at 37 °C were first brought to a $MnCl_2$ concentration of 12.5 mM. Reactions were quenched at the indicated times by the addition of 4× SDS−PAGE loading buffer [200 mM Tris-HCl (pH 6.8), 8% SDS, 40% glycerol, 4% 2-mercaptoethanol, and 0.08% bromophenol blue]. Cross-linking was analyzed by electrophoresis of samples on a 10% SDS−polyacrylamide gel.

Preparative Cross-Linking and Peptide Mapping. Preparative cross-linking reactions for peptide sequencing (50−150 mL volume) were performed in parafilm trays that held a 25 mL reaction mixture, with a sample depth of 0.3 cm. The reaction buffer was as described above, containing 7.5 mM $MnCl_2$. NaCl (20 mM) was provided by the integrase storage buffer. The substrate concentration was 2 μM and that of integrase 0.5−1.0 μM. Irradiation of samples was for 6 min at a distance of 8 cm on an inverted heat block cooled to 4 °C. After cross-linking, the $MnCl_2$ concentration was increased to 12.5 mM by the addition of 0.5 M $MnCl_2$, and the reaction volume was incubated at 37 °C for 60 min in 50 mL conical tubes. Cross-linked integrase complexes were harvested by cooling on ice for 5 min, followed by centrifugation for 30 min at 2500 rpm in a Beckman GS-6R centrifuge (Ellison et al., 1995). Quantitation of the radioactive DNA indicated that precipitation by this method was approximately 80% efficient. Precipitated integrase was resuspended for purification on a 10% SDS−polyacrylamide gel. Both the gel and loading buffer contained 0.05% SDS. The gel loading buffer contained, in addition to standard components, 2.5 M urea, 8 mM CHAPS, and 80 mM EDTA. Cross-linked integrase complexes were detected by autoradiography and eluted from gel slices into water and 25 μM EDTA for 10 h at room temperature. Approximately 50−60% recovery was typically obtained at this step. Integrase was precipitated (nearly quantitatively) from the eluate at −80 °C for 2 h by a combination of 11.5% acetone and 4% trichloroacetic acid. Following centrifugation, the integrase pellet was washed two times with cold acetone and resuspended in 30 μL of 8 M urea, 0.3 M Tris-HCl (pH 7.0), and 0.4 mM EDTA. After the addition of 5 μL of 50 mM DTT, the sample was incubated at 50 °C and then at room temperature for 5 min each. Five microliters of 100 mM iodoacetamide was added, and the sample was incubated at room temperature for 15 min. Thirty microliters of water and 25 μL of 8 mM $CaCl_2$ were then added to the sample.

Digestion was initiated with 1 μg of endoproteinase Glu-C. Samples were incubated at 25 °C for 10−12 h. An aliquot of each digested sample was electrophoresed on a 16.5% SDS−tricine−polyacrylamide gel (Schagger & von Jagow, 1987). After complete digestion, the sample was subjected to electrophoresis on a 8 M urea/1× TBE polyacrylamide gel (7.2%), designed for the separation of nucleic acids. After gel electrophoresis, samples were electrophoretically transferred to a poly(vinylidene fluoride) (PVDF) membrane (Immobilon P^{sq}, Millipore) at 200 mA for 60 min, at 4 °C, in 1× TBE using a Bio-Rad trans blotter. The PVDF membrane was washed for 1−2 min in water and dried at room temperature. Transfer to the PVDF membrane was complete as measured by autoradiography. Slices of the membrane containing cross-linked peptides were cut out using an autoradiogram as a template. The yield of cross-linked peptide was estimated by Cerenkov counting of the dried filter fragment. The immobilized peptides were sequenced and quantitated by automated Edman degradation (W. M. Keck Foundation Resource Laboratory at Yale University). The quantitative values obtained during peptide sequencing were normalized for the amount of radioactivity in each membrane slice and are reported in Table 1.

Table 1: Abundance of Cross-Linked Complexes and Peptides

		recovery		
cross-linker	digest[a] (pmol)	peptide	amount[b] (pmol)	ratio[c]
AZP-V2	84	5	15	1:1
		6	15	
AZP-V3	60	2	50	17:1:1.3
		3	2	
		6	1.5	
AZP-V4	58	2	24	4.5:1.8:1
		3	25	
		6	6	
AZP-V5	64	2	50	3.8:1
		6	13	
5-iodo-V	15	2	1	1:1
		3	1	
AZP-T1	30	1	13	9:1:2
		3	2	
		6	4	
AZP-T4	50	3	2	1.3:1
		4	1.5	
AZP-T5	50	2	3	1:1
		6	3	
AZP-T6	25	1	12	8:1.5:1.5:1
		2	3	
		3	3	
		4	2	
5-iodo-T	10	1	7	18:1
		3	0.5	

[a] The approximate amount of gel-purified C-II complex digested with endo-Glu-C protease calculated from Cerenkov counting and the specific acitivity of the radiolabeled DNA substrate. [b] The approximate picomole values of purified cross-linked peptides obtained during peptide sequencing. [c] The ratio of cross-linked peptides reported is based upon the absolute quantitative values obtained during peptide sequencing (column 4) but have not been calculated directly from these values. In order to correct the data for undetermined factors that can influence the quantitative values obtained from sequencing different PVDF membrane slices in a single experiment, the values reported in column 4 of this table have been normalized to the amount of radioactivity in each membrane slice following gel purification. The normalization factor was calculated by determining the relative amount of radioactivity in each excised PVDF membrane slice by Cerenkov counting. This normalized ratio provides a more accurate indication of the relative frequency at which cross-linking occurred to each peptide, and in all but one case, AZP-V4 peptide 3, this value is in general agreement with the absolute value.

RESULTS

Protein-DNA Cross-Linking Substrate Design. Disintegration substrates mimic integration intermediates formed after the joining of a viral DNA end to a target DNA sequence. Because a disintegration substrate contains functionally distinct viral DNA end and target DNA, it is better suited than a model viral DNA end substrate for mapping specific viral and target DNA contacts with integrase. Two types of disintegration substrates were used in this study. In all viral DNA cross-linking experiments, a standard dumbbell substrate was used with which the reaction products were a 16-nucleotide viral DNA hairpin and a 24-nucleotide closed-circle target DNA (Figure 1A). Target DNA cross-linking was performed primarily with a hybrid dumbbell-Y-oligomer substrate, with which the reaction products were a 16-nucleotide viral DNA hairpin and a 44-nucleotide linear target DNA (Figure 1B).

Radiolabeled disintegration substrates were prepared for cross-linking by coupling a photoreactive azidophenacyl group to a single sulfur atom within the substrate. Each cross-linking substrate had a phosphorothioate inserted at a selected position in either the viral DNA end or target DNA sequence (Figure 2A,B). These locations were selected on the basis of three primary criteria. (i) Previous data suggested an important role for these DNA features in integration. (ii) The locations of the selected DNA sites were in close proximity to sites of integrase-mediated cleavage and ligation and hence likely to be positioned close to the enzyme's active site. (iii) Placement of the bulky aromatic group at the selected position did not prevent the modified DNA from being used as a substrate by integrase (data not shown).

In two substrates, the photoreactive group was a 5-iododeoxycytosine nucleotide rather than the azidophenacyl group. The 5-iododeoxycytosine was placed at the conserved cytosine four bases from the viral DNA 3′ end or in target DNA as the penultimate 3′ nucleotide.

Formation and Analysis of UV Cross-Linked Integrase-DNA Complexes That Retain Catalytic Activity. Integrase−DNA complexes were formed by irradiation of a photoreactive substrate, in standard integrase reaction mixtures using 300 nm transilluminator light. Initial experiments used an azidophenacyl-coupled dumbbell substrate. A low level of cross-linking was observed when the substrate did not contain the azidophenacyl group. Quantitation revealed that cross-linking increased 4−5-fold when the azidophenacyl group was present. Under conditions where integrase was present at a 10−20-fold molar excess over an azidophenacyl substrate, we observed 3−5 % of the substrate cross-linked (Figure 3A). A divalent cation is required by integrase for catalysis. Analysis of the irradiated samples by SDS−PAGE and autoradiography showed that two discrete integrase−DNA complexes were formed when manganese was included in the reaction mixture during irradiation at 37 °C. Only a single photoadduct was observed when manganese was omitted from the reaction (Figure 3A).

To investigate the identity of the lower-molecular weight complex (C-II) formed only in the presence of manganese, we determined if the complex formed without any cation (C-I) was a precursor of the C-II complex. C-I cross-links

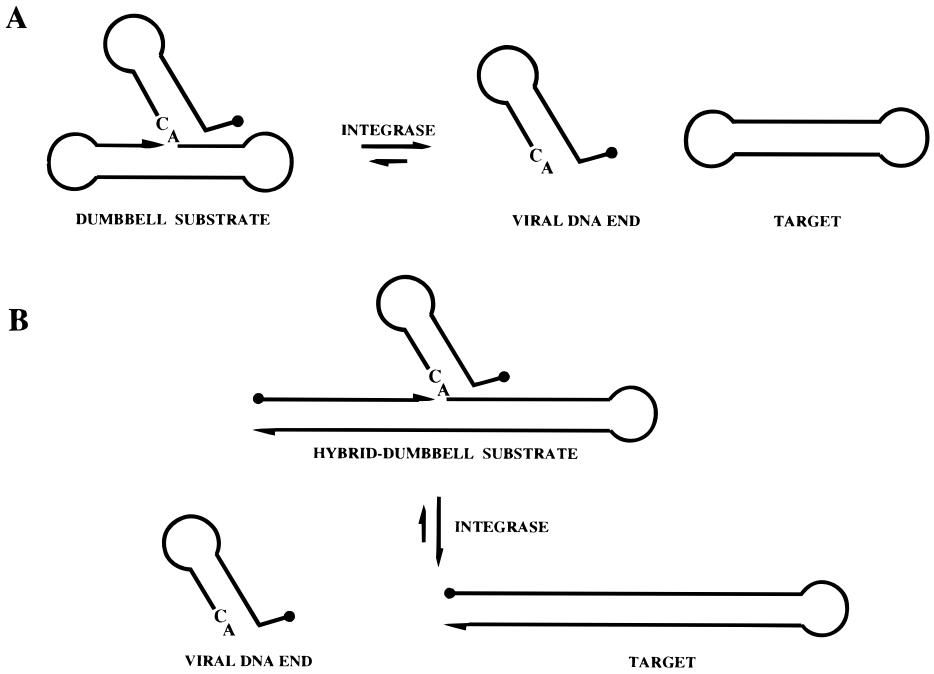

FIGURE 1: DNA substrates and reaction products. (A) The dumbbell disintegration substrate is a single oligonucleotide, 40 nucleotides long. The viral DNA sequences in the hairpin stem are identical to those of the U5 end of the HIV-1 viral DNA molecule, and the target DNA sequences are arbitrary (Chow & Brown, 1994). Disintegration reaction products are a 16-nucleotide hairpin oligonucleotide corresponding to the viral DNA end and a 24-nucleotide closed-circle target DNA. (B) The hybrid-dumbbell disintegration substrate and reaction products. This substrate facilitates ^{32}P end labeling of target DNA sequences and creates a substrate with 10 base pairs of target DNA on each side of the viral DNA junction. The hybrid-dumbbell substrate is made by annealing two oligonucleotides together. One is 50 nucleotides long and contributes the HIV-1 U5 end and a portion of the target DNA duplex; the second is 10 nucleotides long and contributes one strand of the target DNA duplex to the left of the viral DNA junction. The disintegration products of this substrate are a 16-nucleotide hairpin oligonucleotide corresponding to the viral DNA end and a 22-base pair linear duplex DNA hairpin, which represents the target DNA.

were formed by irradiation of reaction mixtures that contained 0.1 mM EDTA instead of manganese. Following UV irradiation, manganese was added to an aliquot of the sample which was then incubated at 37 °C for 40 min (Figure 3B). During this postirradiation incubation, the conversion of C-1 complexes into C-II complexes was observed. The addition of a 10-fold molar excess of unlabeled viral DNA end oligonucleotide after irradiation failed to prevent conversion of C-I to C-II complexes (data not shown). Typically, 25−30% conversion of C-I to C-II complexes was observed. We never observed greater than 50% conversion. This experiment showed that the C-II product was formed by a manganese-dependent event that occurred after cross-linking enzyme to substrate. Integrase cross-linking to the hybrid-dumbbell substrate also produced both C-I and C-II complexes. With this substrate, the C-II complexes migrated more slowly than the C-I complexes (Figure 3C). This result would be expected if in C-II complexes only the disintegration product were cross-linked to integrase.

The most likely origin of the C-II complexes was integrase-catalyzed disintegration of the cross-linked substrate. Therefore, we tested whether formation of the C-II cross-link required both a functional substrate and a functional enzyme. A substrate unable to undergo disintegration, because it lacked a 3′-hydroxyl group at the 3′ end of the target DNA, formed C-I complexes. However, these complexes were unable to be converted into C-II complexes by integrase in the presence of manganese (Figure 3D, lanes 11 and 12). Additionally, C-I cross-links made between an active substrate and a catalytically inactive integrase protein, D116N or D64R, were also unable to form C-II cross-links (lane 15). Direct cross-linking of the disintegration product alone to integrase by UV irradiation generated a complex that comigrated with the C-II complex (lane 10). We concluded that the C-II cross-linked species resulted from a disintegration event that occurred after cross-linking of the DNA substrate to integrase, in an orientation compatible with activity. Although not shown in Figure 3, formation of the C-II complex was observed to be dependent upon both the metal ion cofactor and a catalytically active integrase enzyme for each different cross-linking substrate. Additionally, the C-II complex for each different substrate had the equivalent mobility in a SDS−polyacrylamide gel.

Mapping DNA Cross-Links to Integrase Peptides. Having established a procedure that allowed functionally cross-linked integrase−DNA complexes to be isolated, we identified the cross-linked peptides in complexes that had catalyzed disintegration after substrate cross-linking. To maximize the fraction of protein molecules cross-linked to substrate, the DNA substrate was present at a 4-fold molar excess over integrase, conditions that resulted in DNA cross-links to 10−12% of the protein. The formation of C-I and C-II complexes in the presence of excess substrate was also dependent on divalent cation and required an active integrase enzyme, as observed for the conditions of enzyme excess in Figure 3 (data not shown). Cross-links were mapped only in complexes that had carried out the disintegration reaction after cross-linking. Formation of these active complexes presumably requires the same enzyme−substrate interactions under conditions of both enzyme and substrate excess. Cross-linked complexes between integrase and reaction product were gel purified and subjected to proteolysis by endo-Glu-C protease. Proteolysis typically generated two

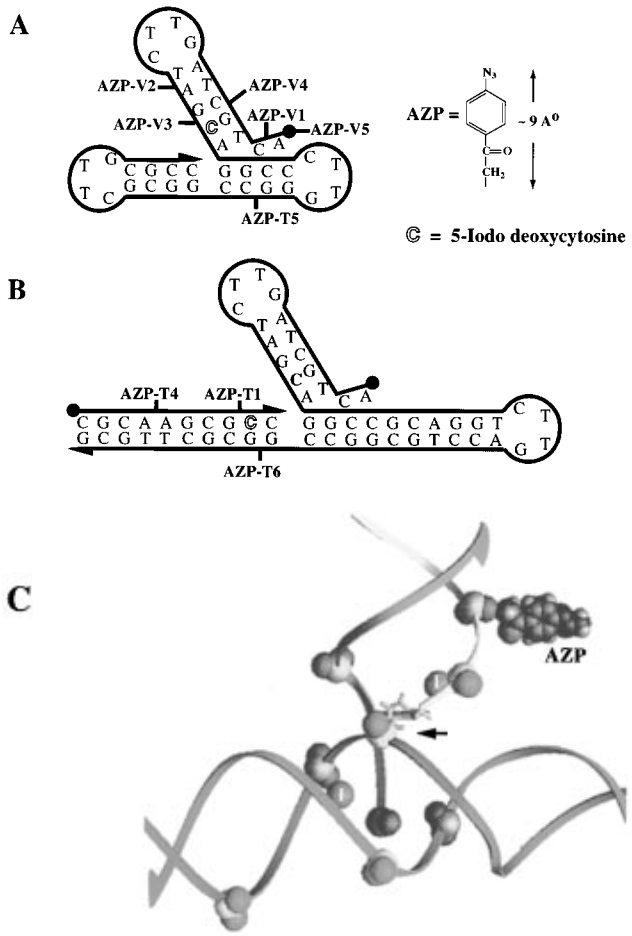

FIGURE 2: DNA cross-linking substrates and photoreactive groups. (A) The dumbbell cross-linking substrates. DNA contacts were probed by coupling a single photoreactive azidophenacyl group to each of a set of dumbbell substrates. Each substrate contained a single phosphorothioate at one of the indicated locations (AZP-V1−5, and -T5). In one substrate, an azidophenacyl group was not used, and instead, 5-iododeoxycytosine was substituted at the conserved deoxycytosine (outline text). (B) The hybrid-dumbbell cross-linking DNA substrates. Target DNA contacts were probed by coupling a single azidophenacyl to each of a set of substrates that contained a phosphorothioate at the indicated locations (AZP-T1, -T4, and -T6). In one substrate an azidophenacyl was not used, and instead a 5-iododeoxycytosine was substituted at the penultimate 3′ nucleotide (outline text). (C) The positions where photoreactive groups were placed are shown on a Grasp (Nicholls et al., 1991) representation of a viral DNA end and a target DNA duplex that are juxtaposed for integration, as in a disintegration substrate. The yellow and green ribbons correspond to the viral DNA end duplex. The red and blue ribbons correspond to the target DNA duplex. The atoms shown are colored as follows: oxygen, red; phosphate, yellow; nitrogen, blue; hydrogen, white; and carbon, gray. A stick representation of the conserved adenosine nucleotide at the viral DNA end is shown in yellow. The black arrow denotes the target DNA phosphate (yellow) where cleavage and ligation occur and the viral DNA end 3′ oxygen (red). The clustered yellow and red spheres on the DNA ribbons represent phosphates and their nonbridging oxygens that were substituted with a thio atom for attaching an azidophenacyl group. A racemic mixture of the R_p and S_p conformations was used. An azidophenacyl group (AZP) is shown juxtaposed with a phosphate oxygen at the AZP-V2 cross-linking position. The 5-carbon atom where iodine was substituted in 5-iododeoxycytosine substrates is shown in magenta and is labeled with an I. Although in this image it is viewed through the minor groove, it is exposed in the major groove.

or three discrete bands on SDS−polyacrylamide gels (Figure 4). The pattern of bands differed depending on the location of the cross-linking group. The cross-linked peptides were purified by electrophoresis through a DNA sequencing type gel and identified by amino acid sequencing. The data described below were obtained from multiple experiments that gave reproducible results. To ensure that they represented closely approximated features in enzymatically active complexes, integrase−DNA cross-links were mapped only in complexes that had catalyzed disintegration of the substrate after cross-linking.

Peptides Cross-Linked to Features of the Viral DNA End Map to the Conserved Catalytic Core and the Carboxyl-Terminal Regions of Integrase. Substrates that contained a photoreactive group near the conserved CA·TG dinucleotide pair of the viral DNA end cross-linked primarily to peptides within the conserved catalytic core region of integrase. Two peptides within the conserved core were identified, and each contained one of the universally conserved acidic residues. The AZP-V3 substrate placed the azidophenacyl group on the phosphodiester backbone immediately 5′ to the conserved cytosine of the CA·TG base pairs. The AZP-V4 substrate placed the azidophenacyl group immediately 3′ to the conserved guanosine nucleotide of these base pairs, on the opposite DNA strand (see Figure 2A). Peptide 2 (amino acids 49−69) was identified as the primary cross-link to substrates AZP-V3 and AZP-V4 (Figure 5A). A secondary cross-link to AZP-V4 was to peptide 3 (amino acids 139−152 of the catalytic core). This substrate also cross-linked at a still lower frequency to peptide 6, containing residues 271−288 of the integrase C terminus. Rare cross-links between the AZP-V3 substrate (CA strand X-link) and peptides 3 and 6 were detected; however, the frequency of cross-links at these positions was over 15-fold lower than those to peptide 2 (Table 1).

A non-azido substrate that was analogous to the AZP-V3 substrate contained 5-iododeoxycytosine in place of the conserved cytosine, and it also cross-linked to peptides 2 and 3 of the integrase core domain. Cross-links between the 5-iodo substrate and integrase occurred at equal frequency to each of these core peptides, and not detectably to the C-terminal domain.

Cross-links to sequences flanking the conserved CA dinucleotide were mapped using azidophenacyl-derivatized substrates. The viral DNA 5′ end (substrate AZP-V5) cross-linked to peptide 2 of the integrase core domain, and with a 4-fold lower frequency to peptide 6 of the C terminus. Finally, substrate AZP-V2 contained the azidophenacyl group between six and seven nucleotides from the viral DNA 3′ end. With this substrate, cross-links were observed to the carboxyl-terminus of integrase only. Contiguous peptides 5 and 6, comprising amino acids 247−270 and 271−288, respectively, were found to cross-link with approximately equal frequencies to this substrate.

Photoreactive Groups on Target DNA Features Cross-Link to Core and Amino- and Carboxyl-Terminal Domains. Previous studies have shown that target DNA features on the DNA arm that is to the left of the viral DNA junction (as diagrammed in Figures 1−3) are more visible to integrase than features that are on the right side of the junction (Chow & Brown, 1994). We therefore designed most of our target DNA cross-linking substrates with the photoreactive azido group in the left arm of the target DNA. A substrate which placed azidophenacyl between the second and third nucleotides from the unjoined 3′ end of the target DNA (AZP-T1) was cross-linked principally to peptide 1, at the very amino terminus of integrase (amino acids 1−11). This substrate also cross-linked to peptides 3 and 6, in the core

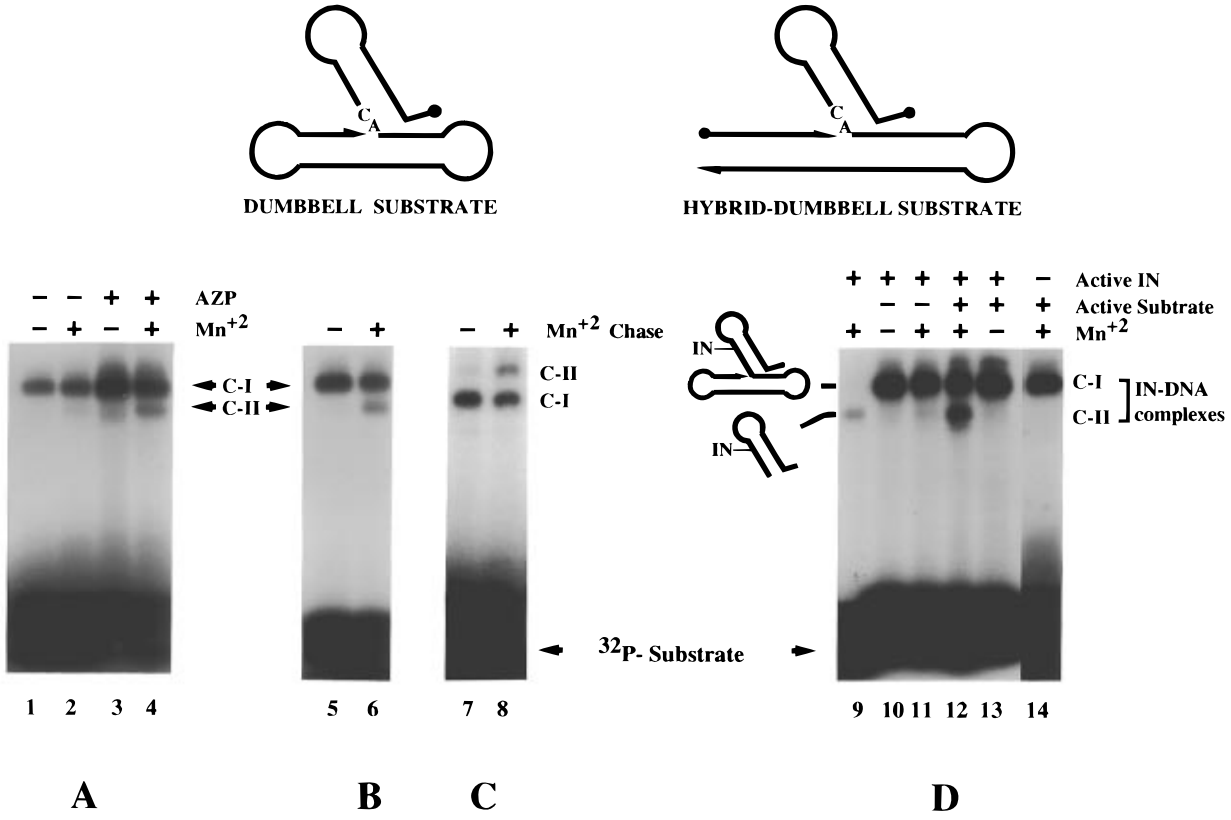

FIGURE 3: Cross-linked integrase−DNA complexes are manganese-dependent and catalytically active. (A) The state of cross-linked species depends upon Mn^{2+}. The reaction mixtures contain 20 nM AZP-V2 dumbbell substrate (^{32}P-5′-end-labeled) and 200 nM HIV-1 integrase and were analyzed by electrophoresis through a 10% SDS−polyacrylamide gel (PAGE). Cross-linking was performed as described in Experimental Procedures and used substrates with (lanes 3 and 4) or without (lanes 1 and 2) a photoreactive azidophenacyl group, and also in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of Mn^{2+}. Reactions in the presence of manganese produced two different cross-linked products, whereas those without manganese produced only a single product (C-I complex). The small amount of material in the light band that migrates very close to the C-II complexes is occasionally observed and is presumably due to nonenzymatic radiolysis of the DNA at the phosphorothioate. (B) Conversion of C-I complexes to C-II complexes by a postirradiation incubation at 37 °C in the presence of 12.5 mM $MnCl_2$. Reactions were analyzed by 10% SDS−PAGE. In lane 5, formation of C-I complexes by UV irradiation of AZP-V2 dumbbell substrate with integrase and 2.5 mM $MnCl_2$ at 5 °C is shown. In lane 6, an aliquot from the reaction in lane 5 was incubated for 40 min postirradiation at 37 °C in the presence of 12.5 mM $MnCl_2$. C-II complexes were produced while C-I complexes were depleted. (C) Cross-linking reaction products with 20 nM AZP-T4 hybrid-dumbbell substrate and 200 nM HIV-1 integrase were analyzed by 10% SDS−PAGE. The substrate was ^{32}P-end-labeled on the 10-mer target DNA oligonucleotide. In lane 7, C-I complexes were formed by UV irradiation with 2.5 mM $MnCl_2$ during UV irradiation at 5 °C. Because there was a low level of $MnCl_2$ present during UV irradiation, a very small amount of C-II complexes that were made during the incubation with UV light can be seen. In lane 8, C-I complexes were converted to C-II complexes by a postirradiation incubation at 37 °C with 12.5 mM $MnCl_2$. Note that, when the hybrid-dumbbell substrate was used, C-I complexes migrated more rapidly than the C-II complexes. (D) Formation of C-II complexes required Mn^{2+}, an active substrate, and the active enzyme. The reaction mixtures contained 20 nM AZP-V1 dumbbell substrate that was ^{32}P-5′-end-labeled and 200 nM HIV-1 integrase. UV irradiation was for 8 min at room temperature, and cross-linking was analyzed by 10% SDS−PAGE. In lane 9, synthetic oligonucleotide corresponding to the viral DNA disintegration product was directly cross-linked to integrase. In lanes 10 and 11, C-II complexes were not detected when cross-linking was to an inactive AZP-V1 dumbbell substrate that contained a terminal 3′ dideoxynucleotide. Lane 10 contained products cross-linked from a reaction stopped immediately after UV irradiation, and lane 11 contained cross-linked products isolated following a 60 min postirradiation incubation at 37 °C with 12.5 mM $MnCl_2$. The small amount of material observed close to the C-II complexes is generated by nonenzymatic hydrolysis of the substrate. In lanes 12 and 13, C-II complexes were detected when cross-linking was to an active AZP-V1 dumbbell substrate that had a terminal 3′ deoxynucleotide. Lane 12 contained cross-linked products isolated following a 60 min postirradiation incubation at 37 °C with 12.5 mM $MnCl_2$, and lane 13 contained products cross-linked from a reaction stopped immediately after UV irradiation. In lane 14, C-II complexes were not detected when cross-linking was to an inactive integrase protein that contained a D116N point substitution. The reaction mixture was incubated for 60 min after UV irradiation at 37 °C, with 12.5 mM $MnCl_2$.

and carboxyl-terminal domains, respectively (Figure 5B and Table 1). When the azido group was moved four nucleotides in the 5′ direction on the same strand (AZP-T4), cross-links were observed to peptide 3 (core domain) and peptide 4 (amino acids 213−234 of the C terminus) with approximately equal frequency.

Cross-links mediated by 5-iododeoxycytosine at a position two nucleotides from the unjoined 3′ end were similar to those observed with the analogous azidophenacyl substrate, AZP-T1. The primary cross-link was to peptide 1, with a low level of cross-linking to peptide 3, and not detectably to peptide 6.

Cross-links to the target DNA strand complementary to the strand containing the unjoined 3′ hydroxyl were mapped as well. Integrase cross-links to target DNA features on the DNA arm to the left of the viral DNA junction (AZP-T6) identified contacts to peptides 1−4. Cross-linking to the amino terminus was 8-fold more frequent than that to the core or carboxyl-terminal peptides. Cross-links to a target DNA feature on the DNA arm that is to the right of the viral DNA junction were mapped. Substrate AZP-T5 placed the azido group on the phosphodiester backbone between the second and third base pairs from the viral DNA junction. This substrate cross-linked to peptide 2 within the conserved

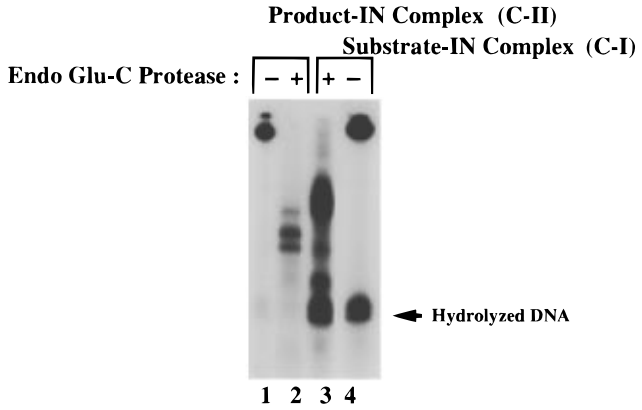

FIGURE 4: Endo-Glu-C proteolysis of cross-linked HIV-1 integrase−DNA complexes. Preperatively cross-linked complexes were made using the AZP-V2 dumbbell substrate, subjected to proteolysis with endo-Glu-C protease, and then analyzed by electrophoresis through a 16.5% SDS−tricine−polyacrylamide gel as described in Experimental Procedures: lane 1, a mock proteolysis of catalytic C-II complexes; lane 2, a 15 h proteolysis of catalytic C-II complexes with endo-Glu-C protease at 25 °C; lane 3, a 15 h proteolysis of C-I complexes with endo-Glu-C protease at 25 °C; and lane 4, a mock proteolysis of C-I complexes. The faster-migrating material in lane 1 is the hydrolyzed DNA substrate. This material is also present in the digested samples and was usually less abundant in product complexes than in the substrate complexes. The reason for the greater abundance of hydrolyzed DNA in substrate complexes is unclear.

core and to peptide 6 within the carboxyl terminus of integrase. With the exception of peptide 4, each peptide that cross-linked to target DNA sequences was also able to be cross-linked to viral DNA end sequences.

Peptides That Can Cross-Link in the Absence of Azidophenacyl or 5-Iodo Groups. We observed that more than one peptide was able to be cross-linked at each substrate position examined. In order to determine if any of these contacts represented cross-links not mediated by the specifically positioned photoreactive group, we identified peptides that could be cross-linked to a dumbbell substrate without a specific photoreactive group. Although the efficiency with which integrase−product complexes were formed in the absence of a photoreactive group was low, the peptides mediating these nonspecific cross-links were mapped. Two such peptides were identified. Peptide 1 was the most abundant of these and accounted for two thirds of the cross-linked peptides. Peptide 2 accounted for the remaining one-third of the peptides that cross-linked to the unmodified substrate. The DNA cross-links to peptide 1 that were detected using azido and iodo substrates were 3−4-fold more abundant than the corresponding products obtained in experiments using the non-photoreactive substrate. Cross-links to peptide 2 reported in Table 1 were 10-fold more abundant when using azido or iodo substrates rather than the unmodified substrate. We therefore believe that those cross-links represented specific contacts mediated by the azido or iodo group in the DNA substrate. We often detected low-level cross-links between peptide 1 and the DNA portion of substrates containing azidophenacyl or 5-iodo groups. When the abundance of peptide 1 cross-linked to these substrates was less than 5% of the total cross-linked peptides, and not significantly above that detected with the unmodified substrate, it was not reported as a specific cross-link above. The repeated occurrence of these cross-links to the modified viral DNA end suggests, however, that in the active complex peptide 1 is in close proximity to a yet unidentified feature of the viral DNA.

DISCUSSION

To establish a productive infection, retroviruses must integrate a DNA copy of their genomic RNA into the host cell's genome, a process that requires the viral gene product integrase. The amino acids, or even the localized regions of integrase that mediate important interactions with features of the viral DNA end, or with target DNA, have not been identified previously. By cross-linking integrase to photoreactive DNA substrates, we have identified specific HIV-1 integrase peptides in close proximity to both viral DNA end and target DNA features.

Specificity and Efficiency of Integrase−DNA Cross-Linking. Because the integrase−DNA complexes we analyzed were required to be catalytically active after substrate cross-linking, the protein−DNA cross-links mapped in this report reflect functional integrase−substrate interactions. Peptide mapping, using 10 DNA substrates that differed in the identity and position of a photoreactive group, identified six different integrase peptides that were close to viral and target DNA features. In most cases, more than one peptide was within the radius of the photoreactive group. A unique mixture and ratio of peptides was observed with each different substrate. The composition of the different peptide mixtures was reproducible for each substrate, indicating that the cross-links were mediated by the specifically placed photoreactive group in the DNA substrate. Additionally, the viral DNA substrate features to which core peptides 2 and 3 were primary cross-links are in close proximity to each other, which, in the simplest model that assumes these cross-links occur in *cis* relative to each other, is in agreement with the spacing of peptides 2 and 3 in the X-ray crystal structure (Figure 6).

Two UV-photoreactive groups with different cross-linking radii were used in this study, and when used at nearby positions, similar results were obtained with both probes. Most experiments used an azidophenacyl cross-linker that has a radius of 9 Å from its point of attachment to the phosphodiester backbone. To mediate attachment of the azidophenacyl group to the DNA, a nonbridging phosphorothioate was inserted into the oligonucleotides during chemical synthesis, thereby creating a racemic mixture of the R_p and S_p conformations. Two experiments used 5-iododeoxycytosine. Iodine has a van der Waal's radius of 2.15 Å and can probe for contacts close to the nucleotide base. The azidophenacyl substrates that we tested were found to cross-link with a greater average efficiency than the 5-iodo substrates. This may reflect the larger effective radius and greater reactivity of the photoactivated arylnitrene. We were unable to identify the specific amino acid residues to which cross-linking occurred. In some experiments, this may have been due to incomplete sequencing of the entire peptide. More often, however, it was likely due to the large radius and promiscuous reactivity of the photoactivated arylnitrene, as well as being due to the fact that a racemic mixture of phosphorothioate stereoisomers provided the attachment sites, all of which could allow cross-linking to occur at multiple sites. The small amounts of peptide recovered for sequencing, which limits the reliability with which small differences in yield per cycle can be measured, were also a factor. The low level of cross-linking obtained with the 5-iodo substrates, in particular, prevented the reliable identification of specific cross-linked residues with these

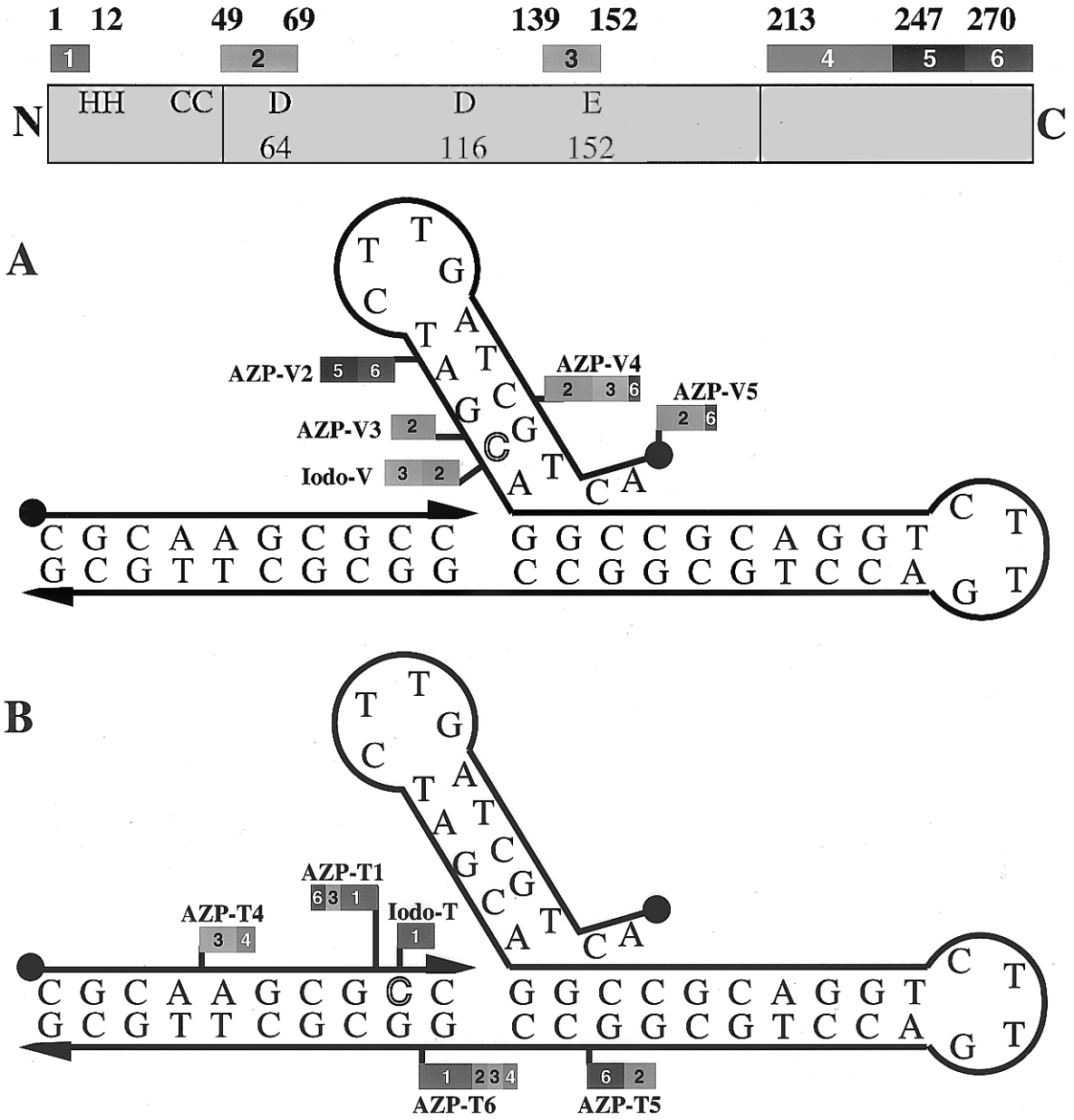

FIGURE 5: Summary of peptides that were found cross-linked to catalytic integrase−DNA complexes. Six integrase peptides, mapping to amino acids 1−11, 49−69, 139−152, 213−246, 247−270, and 271−288, were found to be cross-linked to different substrates. All of the data were obtained by mapping cross-links in complexes that had catalyzed disintegration after UV cross-linking. Each peptide is identified by a number and a color (top). Rectangular boxes flagged to the substrate (below) illustrate where on the viral and target DNA sequences each peptide was found to be cross-linked. Because two or three peptides were found to be cross-linked to each site, the approximate relative abundance of the different peptides cross-linked at each site is indicated by the size of the corresponding colored rectangles (see Table 1). The positions of 5-iododeoxynucleotides are indicated by outline text. (A) Peptides that cross-linked to features of the viral DNA end. (B) Peptides that cross-linked to features of target DNA.

substrates. Although we found the 5-iododeoxycytosine substrates to cross-link with low efficiency to integrase, 5-iodouracil has been shown previously to be an efficient cross-linking group in RNA and DNA nucleoprotein complexes (Willis et al., 1993).

Viral DNA End Cross-Links to HIV-1 Integrase. Residues within the conserved catalytic core domain and carboxyl-terminal domain cross-linked to features of the viral DNA end. It is not surprising to find conserved residues within the core domain of integrase in close proximity to the conserved CA·TG base pairs at the viral DNA end. These base pairs are conserved among all retroviral integrases and bacterial transposases. The CA dinucleotide is located three and four nucleotides from the 3′ end of the viral DNA, and 3′ end processing occurs 3′ to this A. Two core peptides of 20 and 14 amino acids were found to cross-link on both DNA strands near the C·G base pair. Peptide 2 (residues 49−69) and peptide 3 (residues 139−152) each contain a single acidic residue of the catalytic $DD_{35}E$ motif. Substitution at any of the three canonical amino acids of this motif with a nonacidic residue abolishes virtually all catalytic activities of integrase proteins *in vitro* (3′ end processing, integration, and disintegration). It has been postulated that this acidic triad might coordinate a metal ion required for catalysis (Bujacz et al., 1995; Drelich et al., 1992; Engelman & Craigie, 1992; Kulkosky et al., 1992). Finding peptides containing residues D64 and E152 close to both strands of the CA·TG base pairs might reflect only the necessity to have the nucleophile and leaving groups of the viral DNA end within the active site. However, it might also reflect a function for these residues in the recognition of the conserved base pairs (J. Gerton and P. Brown, unpublished results). Additional amino acids

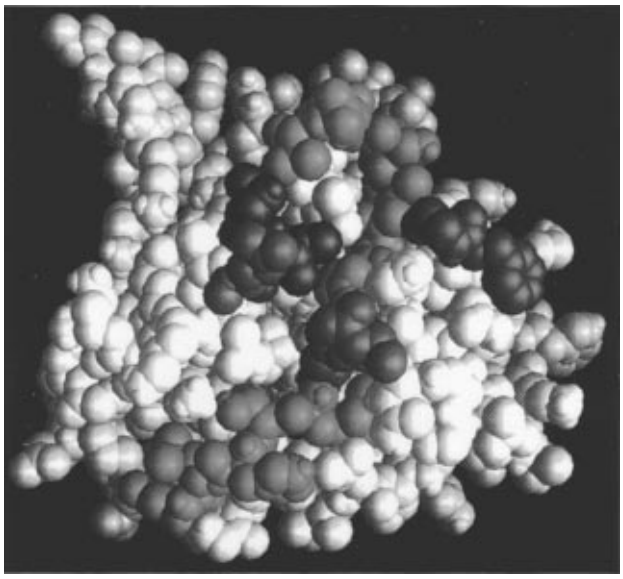

FIGURE 6: Location of cross-linked core peptides on the crystal structure of the HIV-1 integrase core domain, amino acids 54−208 (Dyda et al., 1994; F. Dyda, personal communication). A single integrase monomer is shown, oriented so that the dyad axis of symmetry of the dimer interface is vertical. The active site residues, D64, D116, and E152, are colored green. The ordered residues of peptide 2, amino acids 54−63 and 65−69, are colored blue, and those of peptide 3, amino acids 139−141 and 149−151, are colored purple.

highly conserved among retroviral integrases are found in peptide 2 (W61, Q62, T66, H67, and G69) and peptide 3 (P142, Y143, P145, Q146, S147, and G149). These amino acids are close to the active site residues in crystal structures of both HIV-1 (Figure 6) (Dyda et al., 1994; F. Dyda, personal communication) and ASV (Bujacz et al., 1995) integrase core proteins. It seems likely that some of these residues function in the recognition of viral DNA ends. When some of the above amino acids have been targeted by single-amino acid substitutions, they have showed reduced levels of 3′ end processing and joining, altered nucleophile preference, and hypersensitivity to mutations at the conserved CA·TG base pairs (Engelman & Craigie, 1992; Kulkosky et al., 1992; van Gent et al., 1993; J. Gerton and P. Brown, unpublished results). These phenotypes are consistent with those residues being close to viral DNA end features. However, a role for any of these residues in making either specific base interactions or important nonspecific interactions with viral DNA has not been demonstrated.

The most carboxyl-terminal peptide of integrase (peptide 6, amino acids 271−288) was found to cross-link with low frequency to sites near the conserved CA·GT base pairs. The residues of this peptide are not highly conserved among retroviral integrases, and thus, it seems unlikely that these residues participate in recognition of the conserved bases. The 5-iododeoxycytosine substrate would be expected to cross-link more specifically to residues that contact this base in the major groove than the azidophenacyl substrates. Because the 5-iododeoxycytosine substrate failed to cross-link peptide 6, interactions with the phosphodiester backbone of the viral DNA end seem to be a more likely role for this portion of integrase.

3′ end processing of the viral DNA end results in a two-nucleotide 5′ overhang, and these two nucleotides are known to participate in a stable metal-ion-dependent interaction with integrase (Ellison & Brown, 1994). The 5′ end of the viral DNA end could be cross-linked to peptide 2 of the core domain. A secondary cross-link to peptide 6 in the carboxyl-terminal domain was 5-fold less frequent. Interactions between the viral DNA 5′ end and residues in peptide 2 may thus be involved in maintaining the stability of the integrase−viral DNA end complex.

The carboxyl-terminal domain of integrase was the only domain that we found cross-linked to the viral DNA end at a position internal to the conserved CA·TG base pairs. Peptides 5 and 6 cross-linked to this position. Thus, conserved residues within the catalytic core of the enzyme make the primary cross-links with viral DNA end features in the immediate vicinity of the conserved CA·GT base pairs, and also extending out to the very end of the viral DNA. In contrast, residues located within the carboxyl terminus cross-link to less conserved features that are internal with respect to conserved base pairs at the viral DNA end. The C-terminal region of the integrase protein has been shown to contain a minimal domain from amino acids 220 to 270 that can fold independently and bind DNA in a sequence-nonspecific manner (Engelman & Craigie, 1992; Lodi et al., 1995; Lutzke et al., 1994; Vink et al., 1993). The role of this domain has been uncertain, and its ability to bind DNA nonspecifically has been viewed generally as support for a role in target DNA binding. The results reported in this study, however, provide physical evidence that, in a catalytically active integrase−DNA complex, this domain can be in close proximity to viral DNA end features that are proximal to the conserved CA·TG base pairs. Because neither the nucleotides at this position nor the amino acids of peptides 5 and 6 are highly conserved, interactions made by these residues are likely to include nonspecific contacts with the phosphodiester backbone as well as, perhaps, contacts with nonconserved bases. Peptide 2 contains several basic residues that might provide electrostatic interactions with the phosphodiester backbone.

Target DNA−Integrase Cross-Links. Target DNA cross-links were found to map within each of the amino-terminal, conserved core, and carboxyl-terminal regions of the protein. The most surprising cross-links that we identified were those to peptide 1, which comprises the first 11 amino acids of integrase. Although a protein that is deleted for the first five amino acids is severely impaired in its ability to carry out 3′ end processing, or to integrate viral DNA end substrates *in vitro*, the reason for this defect has been unknown, and it has not been associated with a defect in nonspecific DNA binding (Khan et al., 1991; Vink et al., 1993). We found peptide 1 to cross-link specifically to target DNA sequences within three nucleotides to the unjoined 3′ end of the target DNA. In addition, it was found to cross-link nonspecifically to viral DNA end features.

The core domain of HIV-1 integrase cross-links to target DNA sequences as well. Although the cross-links to the amino terminus were not expected, cross-links to the core domain were not surprising. Recent studies of chimeric integrase proteins have demonstrated that target site selection is largely determined by the core domain of retroviral integrases (Katzman & Sudol, 1995; Shibagaki & Chow 1997). Peptides 2 and 3 were found to cross-link to target sequences. Therefore, residues that determine target site selection may lie within these peptides.

Carboxyl-terminal residues close to target sequences map to peptides 4 and 6. Peptide 4 was found to cross-link only target DNA features. Peptide 6 was cross-linked within three

nucleotides on both sides of the recombination junction and was identified as a secondary cross-link at positions within the viral DNA end as well. In experiments that used the AZP-T1 substrate, peptide 6 cross-linked less frequently than peptide 1. From this, we can infer either that the amino and carboxyl termini of HIV-1 integrase are close to each other in the three-dimensional structure or that these regions are brought close in a multimeric complex. The sites of cross-linking to the integrase polypeptide were mapped under conditions where the DNA substrate was present at a 4-fold molar excess relative to integrase. Under these conditions, cross-links were observed to only about 10−12% of the protein molecules. Although the active fraction of integrase molecules in the reactions was unknown, this cross-linking stoichiometry is consistent with the formation of a higher-order integrase multimer. With the data presented here, we cannot address the higher-order architecture of the integration complex. Recently, it has been shown that the Mu transposasome forms a higher-order complex, in which donor cleavage and strand transfer are catalyzed by the MuA protomer that is bound in *trans* to the partner Mu DNA end (Aldez et al., 1996; Savilahti & Mizuuchi, 1996). The *cis* and/or *trans* arrangement of the integrase−DNA cross-links, relative to the active site that catalyzes disintegration, will provide important constraints on both the stoichiometry and positioning of integrase protomers in a model of an integration complex. We are currently determining whether each reported cross-link is in *cis* or in *trans* to catalytic residues in an active integration complex. This information should facilitate our effort to develop a detailed model of integrase interactions with both target and viral DNA end sequences.

ACKNOWLEDGMENT

We thank the W. M. Keck Foundation Resource Laboratory at Yale University for providing amino acid sequencing and technical advice on the peptide mapping. We thank Fred Dyda and Bob Craigie for providing unpublished data. We thank Jennifer Gerton and Dan Herschlag for many helpful discussions. We also thank Sam Chow for critical reading of the manuscript.

REFERENCES

Aldez, H., Schuster, E., & Baker, T. A. (1996) *Cell 85*, 257−269.
Bowerman, B., Brown, P. O., Bishop, J. M., & Varmus, H. E. (1989) *Genes Dev. 3*, 469−478.
Brown, P. O., Bowerman, B., Varmus, H. E., & Bishop, J. M. (1987) *Cell 49*, 347−356.
Brown, P. O., Bowerman, B., Varmus, H. E., & Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2525−2529.
Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodower, A., Merkel, G., Katz, R. A., & Skalka, A. M. (1995) *J. Mol. Biol. 253*, 333−346.
Bushman, F. D., & Craigie, R. (1990) *J. Virol. 64*, 5645−5648.
Bushman, F. D., & Craigie, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3458−3462.
Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., & Craigie, R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3428−3432.
Chow, S. A., & Brown, P. O. (1994) *J. Virol. 68*, 3896−3907.
Chow, S. A., Vincent, K. A., Ellison, V., & Brown, P. O. (1992) *Science 255*, 723−726.
Colicelli, J., & Goff, S. P. (1985) *Cell 42*, 573−580.
Craigie, R., Fujiwara, T., & Bushman, F. (1990) *Cell 62*, 829−837.
Donehower, L. A., & Varmus, H. E. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6461−6465.
Dotan, I., Scottoline, B. P., Heuer, T. S., & Brown, P. O. (1995) *J. Virol. 69*, 456−468.
Drelich, M., Wilhelm, R., & Mous, J. (1992) *Virology 188*, 459−468.
Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., & Davies, D. R. (1994) *Science 266*, 1981−1986.
Ellison, V., & Brown, P. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7316−7320.
Ellison, V., Gerton, J., Vincent, K. A., & Brown, P. O. (1995) *J. Biol. Chem. 270*, 3320−3326.
Engelman, A., & Craigie, R. (1992) *J. Virol. 66*, 6361−6369.
Engelman, A., Mizuuchi, K., & Craigie, R. (1991) *Cell 67*, 1211−1221.
Engelman, A., Hickman, A. B., & Craigie, R. (1994) *J. Virol. 68*, 5911−5917.
Fujiwara, T., & Mizuuchi, K. (1988) *Cell 54*, 497−504.
Fujiwara, T., & Craigie, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3065−3069.
Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., & Skalka, A. M. (1990) *Cell 63*, 87−95.
Katzman, M., & Sudol, M. (1995) *J. Virol. 69*, 5687−5696.
Khan, E., Mack, J. P. G., Katz, R. A., Kulkosky, J., & Skalka, A. M. (1991) *Nucleic Acids Res. 19*, 851−860.
Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., & Skalka, A. M. (1992) *Mol. Cell. Biol. 12*, 2331−2338.
Lafemina, R. L., Callahan, P. L., & Cordingley, M. G. (1991) *J. Virol. 65*, 5624−5630.
Leavitt, A. D., Shiue, L., & Varmus, H. E. (1993) *J. Biol. Chem. 268*, 2113−2119.
Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., & Gronenborn, A. M. (1995) *Biochemistry 34*, 9826−9833.
Lutzke, R. A., Vink, C., & Plasterk, R. H. A. (1994) *Nucleic Acids Res. 22*, 4125−4131.
Nicholls, A., Sharp, K. A., & Honig, B. (1991) *Proteins 11*, 281−296.
Panganiban, A. T., & Temin, H. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7885−7889.
Roth, M. J., Schwartzberg, P. L., & Goff, S. (1989) *Cell 58*, 47−54.
Savilahti, H., & Mizuuchi, K. (1996) *Cell 85*, 271−280.
Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Schwartzberg, P., Colicelli, J., & Goff, S. P. (1984) *Cell 37*, 1043−1052.
Sherman, P. A., Dickson, M. L., & Fyfe, J. A. (1992) *J. Virol. 66*, 3593−3601.
Shibagaki, Y., & Chow, S. A. (1997) *J. Biol. Chem. 272*, 8361−8369.
van Gent, C. C., Uode Groeneger, A. A. M., & Plasterk, R. H. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9598−9602.
van Gent, D. C., Elgersma, Y., Bolk, M. W. J., Vink, C., & Plasterk, R. H. A. (1991) *Nucleic Acids Res. 19*, 3821−3827.
van Gent, D. C., Uode Groeneger, A. A. M., & Plasterk, R. H. A. (1993) *Nucleic Acids Res. 21*, 3373−3377.
Varmus, H. E., & Brown, P. O. (1989) *Retroviruses*, pp 53−108, American Society for Microbiology Press, Washington, DC.
Vincent, K. A., Ellison, V., Chow, S. A., & Brown, P. O. (1993) *J. Virol. 67*, 425−437.
Vink, C., van Gent, D. C., Elgersma, Y., & Plasterk, R. H. A. (1991) *J. Virol. 65*, 4636−4644.
Vink, C., Oude Groeneger, A. A. M., & Plasterk, R. H. A. (1993) *Nucleic Acids Res. 21*, 1419−1425.
Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R., & Koch, T. H. (1993) *Science 262*, 1255−1257.
Yoshinaga, T., & Fujiwara, T. (1995) *J. Virol. 69*, 3233−3236.
Zheng, R., Jenkins, T. M., & Craigie, R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13659−13664.

BI970782H