Photo-Cross-Linking Studies Suggest a Model for the Architecture of an Active Human Immunodeficiency Virus Type 1 Integrase−DNA Complex†

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 December 2, 1997; Revised Manuscript Received February 27, 1998

ABSTRACT: The virally encoded integrase protein carries out retroviral integration, which requires specific interactions with the two ends of the viral DNA, and also with host DNA that is the target of integration. We attached a photo-cross-linking agent to specific viral and target DNA sites to identify regions of the integrase polypeptide that are in close proximity to those substrate features in the active integrase−DNA complex. The active form of integrase is a multimer. The higher-order organization of the active integration complex was therefore investigated by determining whether specific cross-links occurred to the active-site containing protomer. Both viral and target DNA cross-links to human immunodeficiency virus type 1 (HIV-1) integrase mapped predominantly to integrase protomers in trans to the active site, in a multimeric integrase complex. The results provide the basis for a model of the protein−DNA architecture of an active HIV-1 integration complex that suggests specific functions for the N-terminal, core, and C-terminal domains of retroviral integrase. One implication of this model is that the integrase multimer that mediates concerted integration of the viral DNA ends is composed of at least eight integrase protomers.

Integration of viral DNA into the host cell genome is an essential step in the retroviral life cycle (*1*). Genetic studies have demonstrated that two viral components are required for integration: (1) the integrase polypeptide that is encoded by the 3′ portion of the retroviral *pol* gene (*2*−*5*) and (2) viral DNA attachment sites for integrase contained within the terminal viral DNA sequences (*6*−*8*). The integration process consists of three discrete steps. In the first step, integrase processes the viral DNA 3′ ends by removing two terminal nucleotides 3′ to a conserved CA dinucleotide. Next, integrase joins the recessed 3′-hydroxyl group of each viral DNA end to staggered 5′ phosphates of the host cell chromosomal DNA, creating a gapped recombination intermediate. In the final step, repair of the gapped intermediate creates an integrated provirus that is flanked by direct repeats of the insertion site sequence (*9*, *10*). It is uncertain if integrase plays a role in repair of the gapped intermediate, but integrase has been shown to be sufficient to mediate the 3′-end processing and joining steps in model studies in vitro (*11*−*15*). Integrase can also catalyze the reverse of viral DNA joining (disintegration) in vitro (*16*).

The feature of the viral DNA ends that is most important for integration is a CA/TG dinucleotide pair located 3 and 4 base pairs from each viral DNA end. Sequence alteration or chemical modification of these nucleotides severely impairs end processing, integration, and also disintegration (*13*, *16*−*19*). Sequences internal to the CA dinucleotide extending up to 15 base pairs from the viral DNA end play significant but less important roles (*6*, *7*, *20*−*24*). Structural features of the viral DNA end, such as fraying of the terminal three base pairs, are also important during end processing and integration (*25*).

Mutational analyses have distinguished three functionally distinct regions of integrase, two of which are characterized by amino acid motifs conserved among retroviruses and retrotransposons (*26*−*28*). The first region comprises the amino-terminal amino acids of integrase proteins and contains a conserved zinc-finger-like motif, $HX_{3-7}HX_{23-32}CX_2C$. Mutations within this region, termed the HHCC domain, cause the loss of 3′-end processing and integration activities but do not impair the ability to perform disintegration (*26*, *29*−*31*). Integrase proteins with a disrupted HHCC domain retain the ability to recognize important features of viral DNA ends, including the conserved CA dinucleotide (*30*), indicating 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 multimers (*32*, *33*).

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, retro-

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

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

§ Department of Biochemistry.

|| Program in Cancer Biology.

⊥ Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

▽ Howard Hughes Medical Institute.

S0006-2960(97)02949-8 CCC: $15.00
Published on Web 04/28/1998

transposases, and retroviral integrases, causes a virtually complete loss of all catalytic activities, strongly suggesting a direct role for these residues in catalysis (*26*, *29*, *34*). Truncation of both the amino- and carboxy-terminal regions of integrase has demonstrated that the central core region of integrase alone can mediate disintegration in vitro (*35*, *36*), confirming that the active site resides within this domain, and also indicating that residues within the core domain participate in DNA binding. Indeed, residues within the core domain provide specificity for the conserved CA/TG dinucleotide pair, and recognition of the canonical A/T base pair depends on the universally conserved glutamic acid residue, E152 (*37*).

The carboxy-terminal region of integrase does not contain any conserved motifs that might point to its function. However, analyses of mutant proteins truncated at the carboxy-terminus have shown that this region contains a nonspecific DNA binding domain (*27*, *35*, *38*, *39*).

Integrase features that are in close proximity to important viral and target DNA features in an active integrase−DNA complex have been mapped recently by photo-cross-linking (*40*). These studies revealed that two peptides containing conserved residues within the integrase core domain are close to conserved features of the viral DNA end as well as to target DNA features in the active complex. Cross-links to viral DNA features that are proximal to the CA/TG dinucleotide pair mapped to the C-terminal DNA binding domain. The N-terminal 11 amino acid peptide of integrase was able to be cross-linked to target DNA features 5′ to the site of integration, as well as to yet-unidentified features of the viral DNA end. We have now examined the higher-order architecture of an active integration complex by mapping the cis/trans relationship of specific viral and target DNA cross-links relative to the active site. This information has been used in conjunction with coordinates that model the integrase core (*41*, *42*) and C-terminal (*43*, *44*) domain structures to develop a model for the protein−DNA architecture of an active integration complex.

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 (*30*). An expression construct for IN 1−270 was prepared by oligonucleotide substitution within a synthetic wild-type gene (J. Gerton, unpublished data) that inserted a stop codon after residue D270. A thrombin-cleavable six histidine tag was inserted at the amino terminus of this construct. An expression construct of D64R integrase was obtained from J. Gerton. This construct also contained the F185K substitution that has been shown to increase the solubility of integrase without affecting activity (*45*). The thrombin-cleavable six histidine tag at the amino terminus of the IN 1−270 construct was substituted for the existing amino terminus of the D64R, F185K expression construct. Each mutant integrase protein was expressed with the amino-terminal six histidine tag. After purification by nickel−NTA chromatography, the histidine tag was removed by proteolytic cleavage with thrombin (Calbiochem). The storage buffer for all integrase proteins consisted of 20 mM *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (HEPES), pH 7.5, 10 mM dithiothreitol (DTT), 1 mM (ethylenedinitrilo)-tetraacetic acid (EDTA), 300 mM NaCl, 10 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), and 10% glycerol. T4 polynucleotide kinase was purchased from New England Biolabs.

Oligonucleotides. DNA oligonucleotides were purchased from Operon Technologies Inc. and purified by electrophoresis through a denaturing polyacrylamide gel. The following oligonucleotide cross-linking substrates contain a phosphorothioate linkage at the single position 3′ to the base denoted in lowercase: V1HdbS (5′ aCTGCTAGTT CTAGCAGGCC CTTGGGCCGG CGCTTGCGCC); V2HdbS (5′ ACTGCTAGTT CtAGCAGGCC CTTGGGCCGG CGCTTGCGCC); V3HdbS (5′ ACTGCTAGTT CTAgCAGGCC CTTGGGCCGG CGCTTGCGCC); V4HdbS (5′ ACTgCTAGTT CTAGCAGGCC CTTGGGCCGGCGCTTGCGCC); T1HdbC1S (5′ CGCAAGCgCC); T4HdbC1S (5′ CGCaAGCGCC); T5HdbSC2 (5′ TGCTAGTTCT AGCAGGCCGC AGGTCTTGAC CTGCGgCCGG CGCTTGCG); and T6HdbSC2 (5′ TGCTAGTTCT AGCAGGCCGC AGGTCTTGAC CTGCGGCCgG CGCTTGCG).

The following oligonucleotides were not modified: HdbC1 (5′ CGCAAGCGCC) and HdbC2 (5′ TGCTAGTTCT AGCAGGCCGC AGGTCTTGAC CTGCGGCCGG CGCTTGCG).

The following oligonucleotides were used in construction of the IN 1−270 integrase expression construct: In 1−270a (5′ CGCGTCGTAA AGCTAAAATC ATCCGTGACT AATAGCCTAG GCCTGTAC) and IN1−270b (5′ AGGCCTAGGC TATTAGTCAC GGATGATTTT AGCTTTACGA CGAGGTAC).

The following oligonucleotides were used in construction of the thrombin-cleavable six histidine amino-terminal tag: His.throm1 (5′ TATGCTGGTT CCGCGTGGTT CTTTCTCAGA CGGTATCGAT AAAGCTCAGG ACGAACACGA AAAATACCAC TCTAACTGG) and His.throm2 (5′ CGCGCCAGTT AGAGTGGTATT TTTCGTGTT CGTCCTGAGC TTTATCGATA CCGTCTAGAA AAGAACCACG CGGAACCAGCA).

Preparation of Azidophenacyl Viral End Cross-Linking Substrates. Each oligonucleotide V1HdbS, V2HdbS, V3HdbS, and V4HdbS 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. Azidophenacyl was coupled to the phosphorothioate-containing oligonucleotides in a reaction buffer consisting of 20 mM $NaHCO_3$ (pH 9.0), 40% methanol, and 10 mM azidophenacyl bromide (Sigma). Substrate concentration varied but did not exceed 1 μM. The reaction volume 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 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 after azidophenacyl coupling, were

performed in minimal light (ambient light with window shades closed and room lights turned off).

Preparation of Azidophenacyl Target DNA Cross-Linking Substrates. Oligonucleotides T1HdbSC1, T4HdbsC1, and HdbC1 were 5′-end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. Unincorporated radionucleotides were removed from all labeling reactions by a G-15 Sephadex spin column. Oligonucleotides T5HdbSC2, T6HdbSC2, and HdbC2 were not radiolabeled. Azidophenacyl was coupled to phosphorothioate-containing oligonucleotides as described above. To form the duplex substrates (T1HdbC1S/HdbC2, T4HdbC1S/HdbC2, HdbC1/T5HdbSC2, and HdbC1/T6HdbSC2), a 1.3 molar excess of the complementary oligonucleotide was added to each radiolabeled oligonucleotide, and they were annealed by heating at 80 °C for 3 min and slowly cooling to room temperature.

Cis/Trans Cross-Linking Analysis. Viral end and target DNA contacts were examined using substrates prepared for cross-linking with azidophenacyl. 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 10.0 mM $MnCl_2$. Substrate concentration was 20 nM and the total integrase concentration was 200 nM. Reactions that contained a mixture of integrase mutant proteins contained equimolar amounts of each enzyme. Reactions were initiated by the addition of integrase and were incubated in darkness for 3 min on ice. Reactions were incubated for an additional 4 min on ice while irradiated with a 4 W, 300 nm, UV light source (Fotodyne) at a distance of 6 cm. An inverted Pyrex beaker (1 L) was placed between the samples and the UV lamp. Reactions were subsequently incubated at 37 °C for 30 min. Reactions were quenched 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 an SDS−10% polyacrylamide gel.

Molecular Modeling. PDB coordinates of model integration substrates were constructed using Insight II software (Biosym Technologies). The viral end sequences corresponded to the terminal 16 base pairs at the HIV-1 U3 and U5 ends. The target DNA sequences were arbitrary, but the core sequence corresponded to the target DNA sequence of the modified-dumbbell substrate used in the cross-linking experiments. The terminal three base pairs of the viral DNA end were frayed by moving the 5′ end nucleotides. A bend was introduced in the target DNA to approximate the bend of DNA around histones in the nucleosome. Modeling of the integration complex, using the PDB coordinates of the model integration substrates, HIV-1 integrase core (*41*, *42*), HIV-1 C-terminal domain (*44*), and ASV integrase core (*46*), was performed using Grasp software (*47*). The coordinates of the final model are available in PDB format (Supporting Information).

RESULTS AND DISCUSSION

Viral and Target DNA Features Cross-Link in Trans to Catalytic Residues. We investigated the higher-order architecture of an active retroviral integration complex by photo-

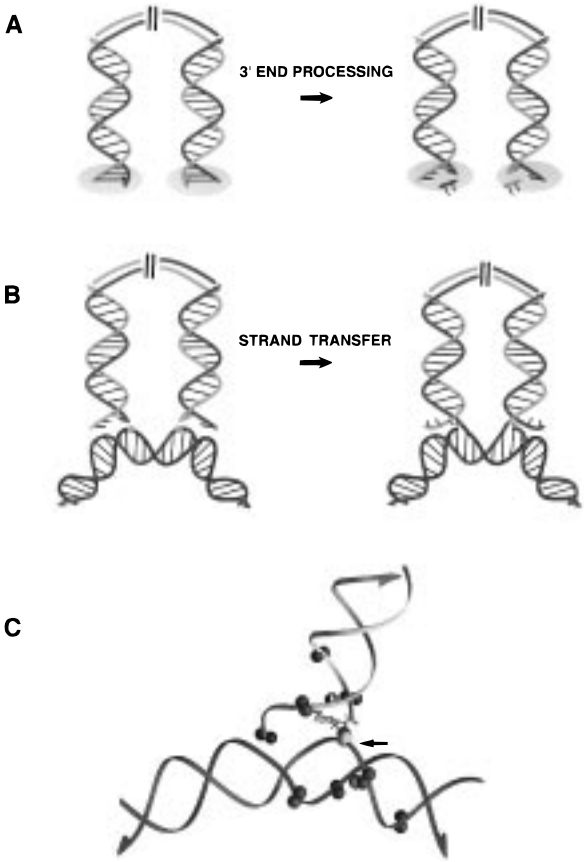

FIGURE 1: Overview of the integration reaction and cross-linking substrates. (A) Schematic depiction of 3′ end processing, the first catalytic step during integration. The end processing reaction removes two nucleotides from the 3′ end of each viral DNA strand, leaving a 3′-hydroxyl group. (B) Schematic depiction of strand transfer. During strand transfer, the recessed 3′ hydroxyl at each viral DNA end is joined to a target DNA phosphate. Joining of both viral DNA ends occurs at phosphates across the major groove of the target DNA that are staggered by five base pairs. Following strand transfer, the gapped recombination intermediate is repaired to create the integrated provirus (not shown). (C) Grasp (*47*) representation of a model integration substrate, showing the sites at which azidophenacyl cross-linkers were attached. The arrow indicates the target DNA phosphate that is attacked by the viral DNA 3′ oxygen immediately behind it. The A of the conserved CA dinucleotide at the viral DNA end is shown as a stick representation. The spheres on the viral and target DNA backbone ribbons represent the phosphate oxygens that were substituted with sulfur for attaching an azidophenacyl group. A single sphere in both the viral and target DNA major groove is labeled with an I, identifying the 5-carbon atom where iodine was substituted in 5-iododeoxycytosine cross-linking substrates.

cross-linking specific features of both viral and target DNA molecules to HIV-1 integrase. The procedure used was similar to one used previously to map integrase residues that are in close proximity to the same viral and target DNA features (*40*). Eight model DNA substrates that mimic the integration product of a single viral DNA end were used, each containing a single photoreactive azidophenacyl group placed at a selected position within the phosphodiester backbone of viral or target DNA sequences (Figure 1C). The azidophenacyl group was attached to a single phosphorothioate at a selected location in each synthetic oligonucleotide. During UV irradiation, the DNA substrates were cross-linked to a mixture of two genetically altered integrase proteins that can functionally complement each other in in vitro reactions. The cross-linked complexes were then tested for the ability

to catalyze the reverse of strand transfer (disintegration), and the identity of the protein cross-linked to the disintegration product was resolved by gel electrophoresis.

Two genetically altered integrase proteins were made to investigate the cis/trans relationship of viral and target DNA cross-links to the active site. One of the mutant integrase proteins, IN D64R, was inactivated by a point substitution at a catalytic residue. Disintegration occurred only after cross-linking of the substrates to integrase. Because the D64R protein is catalytically inactive, following cross-linking to the disintegration substrate it can be isolated cross-linked to the reaction product only if the photoreactive group on the DNA substrate was juxtaposed to a protomer of the integrase multimer that is in trans to the catalytic protomer in the active complex. The second integrase protein used for these experiments had a functional active site but was truncated at the carboxy terminus by 18 amino acids, allowing it to be resolved from the active-site mutant by gel electrophoresis. Following integrase−substrate cross-linking on ice, the reaction mixtures were incubated at 37 °C to allow for disintegration to occur. The integrase−DNA complexes were then analyzed by electrophoresis on SDS−polyacrylamide gels.

Cross-links to viral DNA features were analyzed using a "dumbbell" disintegration substrate, consisting of a single 40-mer oligonucleotide that was self-complementary. When this dumbbell substrate was radiolabeled at the 5′ end, the disintegration reaction products were a labeled 16-nucleotide viral DNA hairpin and an unlabeled 24-nucleotide closed-circle target DNA (Figure 2). The appearance of a cross-linked complex that was not observed in reactions containing either mutant variant alone, corresponding to the viral DNA disintegration product cross-linked to the catalytically inactive integrase protein, D64R, indicated that cross-linking occurred in trans to the active site. Viral DNA features at three of the four different locations analyzed were found to cross-link in trans to the active site: (1) the overhanging nucleotides at the 5′-end (AZP-V1), (2) a position 3 nucleotides from AZP-V1 in the 3′ direction (AZP-V4), and (3) a position on the complementary strand that is between 6 and 7 nucleotides from the viral DNA 3′ end (AZP-V2). The integrase features that cross-linked to azidophenacyl substrates AZP-V1 and AZP-V4 have previously been mapped to two different peptides within the integrase core domain, and those that cross-link to the AZP-V2 substrate were mapped to the DNA binding domain in the carboxy-terminal region of integrase (*40*). In reaction mixtures containing substrates AZP-V1, AZP-V2, and AZP-V4, the disintegration product was observed cross-linked to the D64R and 1−270 integrase proteins with approximately equal frequencies. Because the 1−270 protein can occupy a position in the active complex that is either in cis or trans to the active site, this observation was consistent with the equal molar concentrations of each mutant integrase protein in the reaction mixtures.

Cross-links to target DNA features were analyzed by using a modified version of the dumbbell substrate, which was prepared by annealing two oligonucleotides together. A 50-mer oligonucleotide contributed the complete viral DNA end portion and most of the target DNA portion. The top strand of the target DNA left arm, as diagrammed in Figure 3, was provided by a second 10-mer oligonucleotide. When the

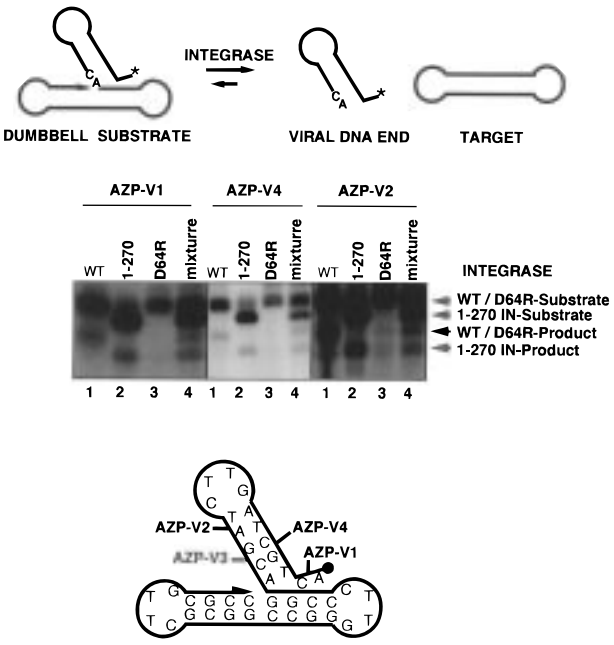

FIGURE 2: Cis/trans analysis of integrase−DNA cross-links to viral DNA substrates: Viral DNA cross-links that occur in trans to the active site. The dumbbell disintegration substrate and reaction products are shown at the top. The substrate is a single oligonucleotide, 40 nucleotides long, that was radiolabeled at the 5′ end with ^{32}P. 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 (*61*). Disintegration reaction products are a 16-nucleotide hairpin oligonucleotide corresponding to the viral DNA end and a 24-nucleotide closed-circle target DNA. 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 locations indicated at the bottom (AZP-V1−4). Cross-linking was performed and the reactions were subsequently assayed for disintegration as described in Experimental Procedures. All reactions were analyzed by electrophoresis through an SDS−10% polyacrylamide gel. Each set of lanes numbered 1−4 corresponds to a single cross-linking substrate: AZP-V1, AZP-V2, and AZP-V4. The integrase component of each reaction, indicated above each lane 1−4, was as follows: (1) wild type, (2) a C-terminal truncation of 18 amino acids, (3) catalytically inactive D64R, and (4) equimolar mixture of the truncated and inactive enzymes. The integrase proteins present in each reaction mixture are indicated above each lane. The identity of each cross-linked complex is indicated to the right. Cross-linking that occurs in trans to the active site was detected in lanes 4. The appearance of a complex that contains the viral DNA disintegration reaction product cross-linked to the catalytically inactive protein, D64R, was formed by integrase complementation after cross-linking (black arrowhead). Trans cross-linking was not detected in a substrate containing azidophenacyl at the AZP-V3 position (data not shown).

substrate was radiolabeled at the 5′ end of the 10-mer oligonucleotide, the disintegration reaction products obtained from this substrate were an unlabeled 16-nucleotide viral DNA hairpin and a labeled 44-nucleotide linear target DNA (Figure 3). Because in two of the substrates, AZP-T5 and AZP-T6, the photoreactive azidophenacyl group was in the unlabeled substrate oligonucleotide, cross-linking was detected to the disintegration reaction product only and not to the labeled substrate oligonucleotide. The appearance of a cross-linked complex that was not observed in reactions containing either mutant variant alone, corresponding to the target DNA disintegration product cross-linked to the catalytically inactive integrase protein, D64R, indicated that cross-linking occurred in trans to the active site. Target DNA

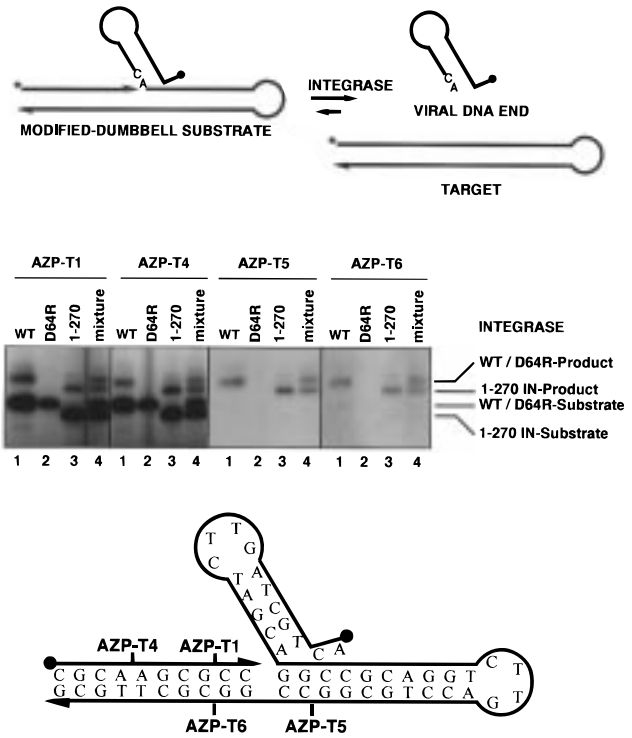

FIGURE 3: Cis/trans analysis of integrase−DNA cross-links to target DNA substrates. The modified-dumbbell substrate and reaction products are shown at the top. This substrate is made by annealing two oligonucleotides together: (1) a 50-mer contributes the HIV-1 U5 end and a portion of target DNA duplex, and (2) a 10-mer contributes the top strand of the target DNA duplex to the left of the viral DNA junction. The 10-mer oligonucleotide was 5′ end labeled with ^{32}P. The disintegration products of this substrate were a 16-nucleotide hairpin corresponding to the viral DNA end and a 22-base-pair linear duplex DNA hairpin, representing the target DNA. Cross-linking and disintegration reactions were performed as described above and in Experimental Procedures. Reaction components and products were labeled as in part A. Cross-linking in trans was again detected by the appearance of a complex that contains the disintegration reaction product cross-linked to the catalytically inactive protein, D64R, formed by integrase complementation after cross-linking (black line).

features at each of four positions analyzed were found to cross-link in trans to the active-site protomer. Three substrates placed the photoreactive azidophenacyl group on the phosphodiester backbone within 2 nucleotides of the viral DNA junction, in either the 3′ or 5′ direction (AZP-T1, -T5, and -T6). The other substrate placed it 6 nucleotides 5′ to the viral DNA junction (AZP-T4). Integrase features that cross-linked to sites AZP-T1 and AZP-T6 were previously mapped predominantly to the amino terminus of integrase, while those that cross-linked to sites AZP-T4, AZP-T5, and AZP-T6 were mapped to the catalytic core and carboxy-terminal domains (*40*).

Our previous experiments, which mapped the peptides to which cross-linking occurred, were carried out under conditions in which the DNA substrate was present in excess over integrase, to maximize the fraction of protein molecules cross-linked to the substrate. The cis/trans mapping experiments described here were carried out under conditions of enzyme excess. The formation of active complexes, whether in the presence of excess substrate or excess enzyme, was dependent upon the metal-ion cofactor and a catalytically active integrase enzyme (*40*, *48*). In both sets of experiments, cross-links were mapped only in complexes that were able to carry out the disintegration reaction after substrate cross-linking. Our interpretation relies on the reasonable assumption that formation of the active complexes requires the same enzyme−substrate interactions under conditions of either enzyme or substrate excess.

Molecular Modeling of an Active Integrase−DNA Complex. The atomic coordinates of structural models of the integrase catalytic core (amino acids 50−212) (*41*, *42*, *46*) and C-terminal domain (amino acids 220−270) (*44*) were used with molecular modeling software [Insight II (Biosym Technologies) and Grasp (*47*)] to develop a working model of an active integration complex. The primary constraints for modeling were integrase−DNA cross-links. We previously mapped viral and target DNA cross-links to integrase peptides (*40*), and in the present work we determined whether these cross-links occur in cis or trans to the active site. Additional constraints on the model were the necessary requirement that the active site be near the phosphodiester bond joining the viral and target DNA molecules, and that active-site residue E152 be near the conserved deoxyadenosine at the viral DNA end (*37*). Because the most important features of integrase−substrate recognition and catalysis are conserved among all retroviruses, and also because the core domain structures for both HIV-1 and avian sarcoma virus (ASV) are highly conserved (*46*), a final constraint imposed upon the model was that an analogous model be compatible with the structural coordinates for the ASV integrase core domain. Preferred integration substrates have specific viral and target DNA features that differ from standard B-form DNA, and the model DNA substrate used for molecular modeling incorporated these preferences: (1) the terminal 3 base pairs at the viral DNA end were frayed (*25*), and (2) the target DNA was bent to approximate the bend of DNA around nucleosomes (*49*, *50*).

The Overall Architecture. In the solved crystal structures, both the HIV-1 and the ASV integrase core domains were dimers (type I dimer) with an extensive hydrophobic interface (*42*, *46*). Our model retained this feature. Both the spacing of the active sites within the type I integrase dimer, and also the location of peptides to which cross-links occurred, made a model that mediated concerted integration of both viral DNA ends by a single type I integrase dimer implausible. In this type of model, the 35 Å distance between the active sites would require severe stretching and unwinding of the target DNA because integration occurs at target DNA phosphates that are separated by only 5 base-pairs (~17 Å in standard B-form DNA). Even if such dramatic deformation of the target DNA were allowed, the face of the core domain that is expected to interact with the viral DNA, and also to which cross-linking occurred, would not be in close proximity to the viral DNA in this type of model.

Nearly all of the integrase−substrate cross-links occurred in trans to the active site, and in the model that best fit the cross-linking constraints, viral DNA cross-links could be distributed among core domains of four different integrase protomers: 1A, 1B, 2A, and 2B (Figures 4 and 5). In this model, core domains 1A and 2A are directly involved in catalysis, and each is able to recognize the CA/TG base pairs at a viral DNA end in cis to 3′-end processing and stand transfer. Core domains 1B and 2B are not directly involved in catalysis, providing interactions primarily with target DNA features. With the same constraints, the ASV core domain

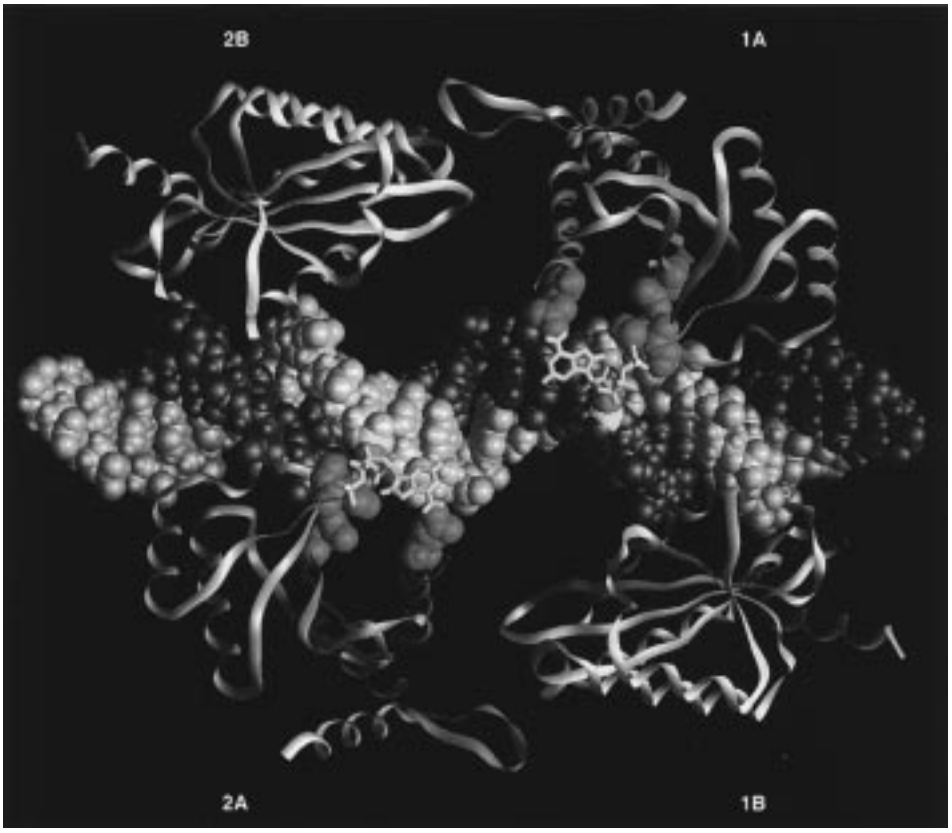

FIGURE 4: Overview of the core-domain architecture of the integration complex model. Ribbon diagrams depicting the secondary structure of the integrase α-carbon backbone are shown in pink and yellow. A protomers that catalyze integration are colored in pink and are labeled 1A and 2A. B protomers that do not participate directly in catalysis are colored in yellow and are labeled 1B and 2B. The active site residues of the A protomers (pink ribbons) are shown in green CPK representation. The target DNA duplex is also shown as a CPK model (light and dark gray). The reactive phosphates where integration occurs are colored yellow. The 3′ oxygen (attacking nucleophile) of each viral DNA end is colored red, and the conserved deoxyadenosine nucleotide at the viral DNA end is depicted with stick bonds in gold.

could be modeled as a very similar complex (not shown). Because each core domain protomer is in a type I dimer, the minimal integrase multimer needed to catalyze concerted integration of both viral DNA ends is an octamer. The four integrase protomers that make contacts with the DNA substrates (shown in Figures 4 and 5) will be referred to as protomers 1A, 1B, 2A, and 2B. The second protomer of each type I dimer will be referred to as 1A′, 1B′, 2A′, and 2B′, respectively. The A/A′ and B/B′ dimers will hereafter be referred to as A and B dimers, respectively.

An important feature of the model imposed by the cross-linking constraints is that the A (active-site) and B (non-active-site) integrase dimers make asymmetric DNA interactions. The B integrase dimers are rotated approximately 45° relative to the A dimers, resulting in a more extensive interface with target DNA features (Figure 5). The interaction between the A and B core domains on the same face of the target DNA is limited to a small surface area involving approximately 5 loosely packed residues from each protomer. Because the interactions between the 1A and 2B or the 2A and 1B core domains are not extensive, juxtaposition of the two viral DNA ends presumably requires additional interactions between the A and B protomers. The relative positions of the C-terminal domain of the A protomer and the core domain of the B protomer suggests that residues within the last 18 amino acids of this C-terminal domain, which are disordered in the NMR structure, are likely to be in position to interact with the core domain of the B protomer. Consistent with this proposed interaction, experimental evidence indicates that the core and C-terminal domains of integrase are required for the formation of integrase tetramers in vitro (*45*).

Viral DNA cross-links to the carboxy-terminal domain were in trans to the active-site protomer and provided constraints for incorporating this region of integrase into the model. The portions of the core and C-terminal domains of the integrase polypeptide for which we have structural models from X-ray crystallographic and NMR data are separated by 11 amino acids. The C-terminal domain is a dimer in solution (*43*, *44*), and each C-terminal domain is positioned in our model so that the amino-terminal residues of this region are in close proximity to the carboxy-terminal residues of the core domain of the same protomer. This places the C-terminal domain on top of the core domain, as diagrammed in Figure 5, extending toward the viral DNA (Figure 6). Viral DNA interactions with the C-terminal domain help to stabilize the critical interactions between viral DNA and the core domain, which, in this model, are limited to the five terminal base pairs of the viral DNA.

The amino-terminal residues of integrase cross-linked to specific target DNA features and also to unidentified features at the viral DNA end (*40*). Very recently a structural model has been solved for the N-terminal domain of HIV-1 integrase (*51*, *52*); however, because the coordinates of the structure are not yet available, it was not formally included in the molecular modeling. Cross-linking data, along with physical constraints imposed by the positions of the core and carboxy-terminal domains of integrase, suggest that an

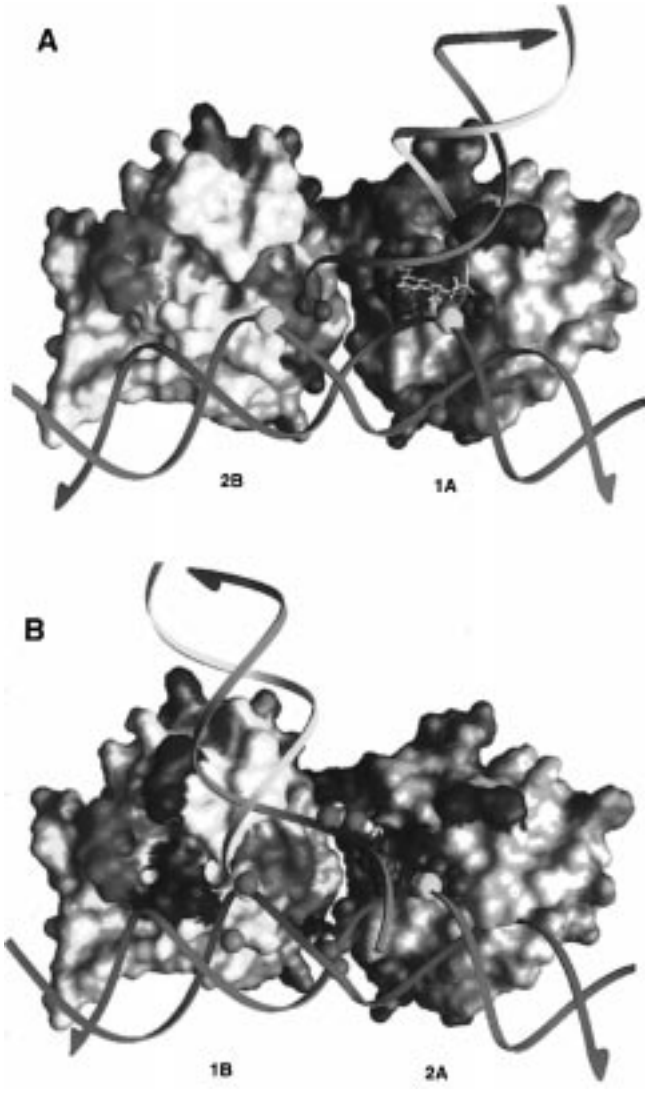

FIGURE 5: Modeling viral and target DNA cross-links to the integrase core domain structure. (A) Viral and target DNA cross-links to the A (active-site) and B (non-active-site) core domains bound to one face of the target DNA (type II dimer). Modeling was performed using Grasp (*47*). The yellow and green backbone ribbons correspond to the viral DNA end, and the blue and red ribbons correspond to target DNA. A stick representation of the phylogenetically conserved deoxyadenosine nucleotide at the viral DNA 3′ end is in gold. The target DNA phosphates where viral DNA joining occurs are shown as yellow spheres on the backbone ribbon, and the attacking 3′ oxygen of the shown viral DNA end (red) is immediately behind the reactive phosphate. The active-site protomer is shaded pink and labeled 1A. The non-active-site protomer is shaded yellow and labeled 2B. Amino acids comprising the active site, D64, D116, and E152, are colored green. Ordered residues in cross-linked peptides 1 and 2 are colored in blue and purple: (1) residues 54−63 and 65−69 and (2) residues 139−142 and 149−151. DNA phosphate oxygens substituted with sulfur, and where a photoreactive azidophenacyl group was attached, are shown as red spheres on the backbone ribbons. The orange sphere near the integrase active site represents the 5-carbon atom of the canonical deoxycytosine substituted with iodine in a 5-iodo cross-linking substrate. The core domain was a dimer in the crystal structure; however, for simplicity, only the monomers that make contact with DNA are shown. It is likely that the corresponding dimers assemble to form the active complex. The olive-colored residues provide additional landmarks for orientation of each protomer. (B) Viral and target DNA cross-links to the A and B core domains bound to the target DNA face opposite of that in panel A. The coloring of all DNAs and amino acids is the same as in panel A. The only exception is in the non-active-site protomer (labeled 1B, yellow), where E152 is colored purple instead of green because it was the C-terminal amino acid of the corresponding peptide. The active-site protomer (labeled 2A, pink) mediates joining of the second viral DNA end, which for simplicity is not shown but can be accommodated in the model without steric clashes. The azidophenacyl group has been included at two cross-linking positions that might appear to be far from the cross-linked peptides and illustrates that the appropriate peptides are indeed within reach of the photoreactive group.

amino-terminal domain (amino acids 1−50) occupies a site near target DNA features 5′ to the site of joining and viral DNA features distal to the CA/TG base pairs. This location could bring the amino-terminal domain of the non-active-site protomers (1B and 2B) close to the core domain of the corresponding active-site protomers (1A and 2A) situated across the viral and target DNA helices, forming bridges across the target DNA between the 1A and 1B and the 2A and 2B integrase protomers, respectively. Intermolecular bridges connecting transposase monomers bound to different DNA substrate features are also features of the intertwined architecture proposed for the MuA transposase−DNA complex (*53*, *54*).

Viral DNA−Integrase Interactions. The terminal 5 base pairs of the viral DNA end are placed close to the core domain. Conserved amino acids in active-site core domains 1A and 2A are in a position that would allow them to participate in recognizing the conserved CA/TG base pairs at the viral DNA ends. Notably, a conserved catalytic residue, E152, is required for normal recognition of the conserved adenine three nucleotides from each viral DNA 3′ end (*55*). The model places a carboxyl oxygen of this glutamic acid side chain close to N6 of the conserved adenine. While proximity of E152 and this nucleotide was a constraint during development of our model, the model suggests a specific explanation for the role of this residue in viral DNA recognition; namely, that there is a direct interaction of the adenine base and glutamic acid side chain. Recent experimental evidence suggests that residue K159 is also in close proximity to this canonical adenine (*56*). The model also places K159 close to the viral DNA 3′ end; however, this residue is closest to the penultimate nucleotide, a guanine residue that is part of the dinucleotide removed during 3′-end processing. The model also places the conserved C/G base pair close to a 7 amino acid peptide, residues 142−149, that was disordered in the core domain crystal structure.

The two-nucleotide overhang at the viral DNA 5′ end contributes to formation of a stable complex with retroviral integrase (*57*). As suggested by cross-linking results, these bases are close to residues in two core peptides in the model (Figure 5A): (1) amino acids at the amino terminus of peptide 49−69 and (2) amino acids at the carboxy terminus of peptide 139−152. An integrase protein that has an amino acid substitution within peptide 139−152, Q148L, has an impaired ability to recognize the 5′ dinucleotide (*55*).

The C-terminal domain can bind nonspecifically to DNA, and its role in integration has been uncertain. The carboxy-terminal domain of integrase (amino acids 220−270) is placed close to viral DNA base pairs 6−13 from the viral DNA end, and is thus in a position to stabilize viral DNA interactions with the core domain (Figure 6). The residues that are closest to the viral DNA are contributed in trans to the active-site protomer (core domain A) by the C-terminal domain of the A′ protomer. These residues include basic amino acids that are predicted to interact with the phosphodiester backbone. The model suggests that amino acids S230, R263, and K264 are likely to make hydrogen bonds with the backbone phosphates or bases. Residues R263 and K264 are within peptide 247−269, which cross-linked to viral DNA features 6−7 nucleotides from the 3′ end. Our model predicts that mutation of these residues would impair

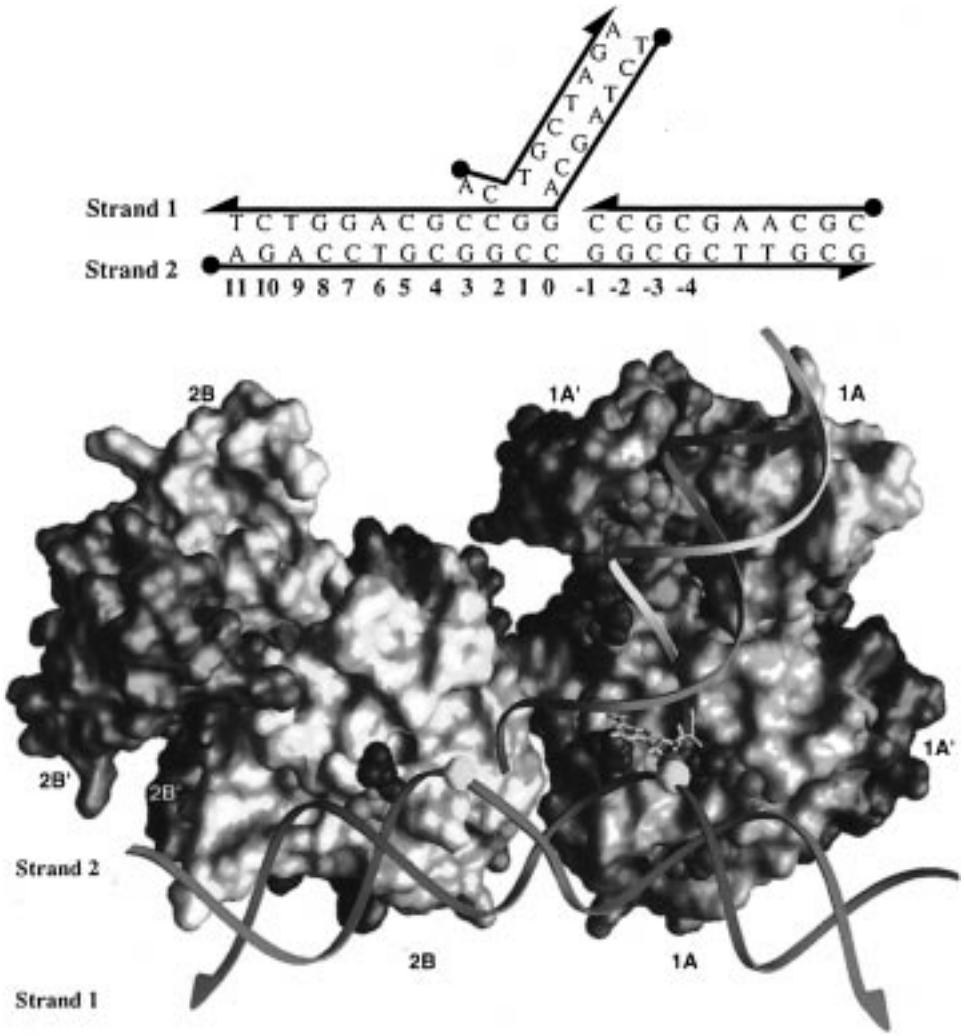

FIGURE 6: Model of the core and C-terminal integrase domains bound to an integration intermediate containing a single viral DNA end. (Upper panel) Schematic diagram of the model DNA substrate, indicating the position of numbered target DNA nucleotides as referred to in the text. (Lower panel) Viral and target DNA duplexes and the integrase core domains are colored as in Figure 4. The target DNA phosphate to which the second viral DNA end joins (not shown for simplicity) is staggered by 5 base pairs from that to which the illustrated viral DNA end is joined; it is indicated with a yellow sphere on the red DNA strand. In this illustration, the second monomer of each core domain dimer is shown for completeness (gray). The second monomer of the A dimer is labeled 1A′, and that of the B dimer, 2B′. The C-terminal domain dimer is above the core domain, as oriented in this diagram. The C-terminal domains are labeled according to the core domain to which they belong: 1A, 1A′, 2B, and 2B′. The active-site amino acids, D64, D116, and E152, are green. Highlighted core domain amino acids that make close encounters with viral and target DNA features are as follows: H67, blue; N117, orange; E152 (monomer 2B), dark purple; and S153, light purple. Highlighted C-terminal domain amino acids that make close encounters with viral DNA features are as follows: R263, dark blue (lower residue); K264, cyan; and S230, light blue. The olive-colored residues provide additional landmarks for orientation of each dimer.

the DNA binding activity of the C-terminal domain and as a consequence inhibit viral DNA 3′-end processing and strand transfer. Although mutations in S230 have not been investigated, proteins with mutations in both R263 and K264 have been described (*39*). Amino acid substitutions at these residues inhibited the ability of the isolated C-terminal domain to bind DNA, and had a corresponding deleterious effect on 3′-end processing and strand transfer activities of the full-length integrase protein. The proposed role for the C-terminal domain in stabilizing viral DNA binding to the core domain is also consistent with the effects of monoclonal antibodies on the catalytic activities of integrase in vitro (*58*).

The C-terminal domains of the B integrase dimers are not close to either viral or target DNA features. We suggest that the unoccupied space between the C-terminal domains of the B dimers and the DNA substrates could accommodate the N-terminal domain of the B protomers.

Target DNA−Integrase Interactions. The core domains of the A and B protomers, but none of the C-terminal domains, contact target DNA features in our model. The A and B core domains are close to 6−7 nucleotides of target DNA. All contacts between integrase residues and target DNA features in the model are with the phosphodiester backbone, rather than base-specific contacts. For example, residues H67 and N117 of the A protomers are positioned to allow hydrogen bonding to phosphate oxygens of nucleotides −3 and −4 of target DNA strand 2 (Figure 6).

Target DNA features are close to different amino acids in the A and B protomers. Residues E152 and S153 of the B protomer contact the target DNA (Figure 6). The position of residue E152 allows hydrogen-bonding contacts with nucleotide 6 of target DNA strand 2, and S153 is positioned to allow hydrogen-bonding to nucleotide 7 of the same strand (Figure 6). Additional residues near the carboxy terminus

of peptide 139−152, which are disordered in the crystal structure, would be expected to be close to nucleotides 9 and 10 of strand 1. Basic residues approach target DNA features at nucleotide 8 of strand 2 (K160) and nucleotides 11 and 12 of strand 1 (KRK186−188).

Contacts between the core domain and target DNA features are consistent with experiments that map the target-site specificity of chimeric integrase proteins to the integrase core (*59*, *60*). Additionally, the phosphates predicted to contact integrase are consistent with the experimental results of adduct-interference studies (*17*, *61*). Moreover, a mutant integrase protein with an amino acid substitution for N117 has been described to be preferentially defective in the strand transfer step of integration (*34*). Contacts between integrase and target DNA bases are not observed in our model. This feature of the model may partially account for the observations that structural, and not sequence, features of the target DNA are the primary determinants of target-site selection for retroviral integrases (*49*, *50*, *62*−*65*).

Implications and Significance. Both viral and target DNA cross-links to HIV-1 integrase have been found to occur predominantly in trans to the active site as part of a multimeric integrase complex. The sites within the integrase polypeptide to which these cross-links occur have previously been mapped (*40*). The cis/trans and peptide mapping information were both obtained from active integrase−DNA complexes, and together they provide strong constraints for developing a working model of the interactions that integrase makes with its viral and target DNA substrates in an active integration complex. Although atomic resolution structural models for the integrase catalytic core (*41*, *42*, *46*) N-terminal domain (*51*, *52*), and C-terminal domain (*43*, *44*) are available, there are not yet structure models for an integrase−DNA cocrystal, the entire integrase polypeptide, or the active higher-order integrase multimer.

The model that we have developed for the integration complex implies that an integrase octamer is required to mediate concerted integration of both viral DNA ends. Experimentally, the multimeric state of integrase in active integration complexes has not yet been established. Our previous experiments mapped the sites of cross-linking to the integrase polypeptide under conditions where the disintegration substrate was present at a 4-fold molar excess relative to integrase. Cross-links were observed to only about 10−12% of the protein molecules under these conditions (*40*). Although the active fraction of integrase molecules in the reactions was unknown, this cross-linking stoichiometry is consistent with the formation of an integrase multimer on the disintegration substrate that contains 10 or fewer integrase protomers. HIV integrase has been observed in vitro to exist in both a monomer−dimer (*66*) and a monomer−dimer−tetramer equilibrium (*45*). Similar experimental observations have been made with Rous sarcoma virus integrase (*67*). Tetramerization of HIV integrase can be promoted by a zinc-dependent interaction that requires the zinc-finger-like motif in the amino-terminal domain (*32*).

A divalent cation-dependent interaction that involves the amino-terminal HHCC domain and a feature near residue C56 in the core domain of integrase is required for integrase activity and can lead to formation of large catalytically active aggregates of integrase (*33*). Propagation of the A and B dimer subunits in our model along the viral DNA helical axes, which is consistent with the phasing of the viral DNA helix relative to the spacing of the integrase protomers, would form higher-order integrase multimers. Specifically, interaction between the core and C-terminal domains of integrase protomers that are stacked upon each other on the same viral DNA end could be stabilized in trans to the core domain by the N-terminal domain of protomers bound to the opposite face of the same viral DNA end. A functional role for this higher-order multimer, formation of which would be promoted by viral DNA sequences more than 14 base pairs from the DNA end, is suggested by experimental evidence showing that the efficiency with which model viral DNA substrates are used can be increased by extending the length of the substrate beyond 15 base pairs from the end (*20*, *68*). It has been estimated that virions typically contain about 50−100 molecules of integrase. Thus an octamer, or a higher-order species, is easily within the physical limit of the in vivo integration complex.

The asymmetry between the A and B core domains is an important feature of the model. A consequence of this asymmetry is that the C-terminal domains of B protomers were not close to either viral or target DNA features. This, however, is consistent with observations that not all integrase protomers in an active complex require a functional active site, amino-terminal, or carboxy-terminal domain (*33*, *69*, *70*). In the model more than one face of the integrase core domain binds target DNA, suggesting that the core domain has more than one face with which it can bind target DNA.

Previously described examples of asymmetric protein−DNA interactions in crystal structure models include that of a transcription factor complex (*71*), the IHF−DNA complex (*72*), and the $\gamma\delta$ resolvase−DNA complex (*73*). Additionally, an asymmetric model has been proposed for the Tn3 resolvase−DNA complex (*74*). In the $\gamma\delta$ resolvase example, which may be most pertinent here, the N- and C-terminal domains of a resolvase dimer are related by different dyad axes, neither of which coincides with the dyad axis relating the DNA binding sites, and importantly, interactions between resolvase and the DNA are different in two protomers. Finally, having multiple surfaces that can bind DNA, and the formation of an asymmetric protein−DNA complex, might benefit organisms such as retroviruses that have a relatively high mutation rate, by buffering the deleterious effects of many point mutations.

The catalytic core domain of integrase is sufficient to catalyze disintegration in vitro, but the critical in vivo reactions, 3′-end processing and strand transfer, require the entire polypeptide. The N-terminal domain contains a zinc-finger-like motif that binds zinc, and is thought to interact with the core domain to mediate the formation of an integrase multimer required for 3′-end processing and strand transfer (*32*, *33*). Integrase deletion derivatives lacking the N-terminal domain require substrates with at least 11 base pairs of target DNA on each side of the viral DNA junction to catalyze disintegration, whereas integrase proteins containing the N-terminal domain require only 5 base pairs upstream of the strand transfer site (*30*, *37*, *61*). These observations suggest that integrase−viral DNA interactions are weakened in N-terminal mutants and that target DNA interactions can stabilize formation of an active complex. The N-terminal domain has not been thought to interact directly with specific features of the viral DNA end, and the nature of the active

integrase multimer has been uncertain. Our model suggests an explanation for these observations.

A bridge between the A and B integrase protomers, which we propose to be mediated by the N-terminal domain, provides a role for the N-terminal domain. In the absence of the N-terminal domain, the active disintegration complex is proposed to be an "open" complex, formation of which requires target-DNA-promoted 1A/1A′−2B/2B′ tetramers (Figure 7A). Propagation of the 1A−2B protomer interaction along the target DNA helical axis could form an integrase oligomer, in which binding of the active-site monomer to features at the viral DNA end is stabilized. In the presence of the N-terminal domain, the active complex is proposed to be a "closed" complex, formation of which requires interaction between A and B dimers situated on opposite faces of the viral and target DNA helices (1A−1B and 2A−2B) (Figure 7B). In this closed complex, binding of the A protomer (active site) to the viral DNA end is stabilized by interaction of the A protomer core domain with the N-terminal domain of the B protomer bound at the same viral DNA end. This interaction, which requires the N-terminal domain of the B protomer and not that of the A protomer, and is therefore unlikely to be reciprocal, is suggested by experimental evidence showing that the N-terminal domain interacts in trans with the core domain to protect residue C56 from alkylation by *N*-ethylmaleimide (NEM) (*33*).

Several predictions arise from details of the model, providing a basis to assess its validity. (1) Residues in peptide 142−149, which were not ordered in the X-ray crystal structure, are predicted to make specific interactions with the phylogenetically conserved C/G base pair of the viral DNA end. P145, which is conserved among all retroviral integrases except for that of ASV, where it is replaced by serine; P142; Y143; Q146; and S147, also highly conserved, are candidates for mediating this recognition. The active conformation of the active site has been proposed to be stabilized by an interaction of E152, which is on the fringe of this disordered peptide, with the conserved deoxyadenosine at the viral DNA end (*37*). Interaction between E152 and this canonical deoxyadenosine may also promote interaction between residues within the flexible peptide and the conserved C/G base pair, further stabilizing the active conformation of the active site. Such a model predicts that recognition of both conserved base pairs occurs in a cooperative manner. Flexible peptides containing an active-site residue have emerged as a common feature of DNA recombinases that bind DNA in a nonspecific manner yet cleave DNA at conserved nucleotide sequences. Examples include MuA transposase (*75*), λ integrase (*76*), and $\gamma\delta$ resolvase (*73*). Stabilization of the active site in a catalytic conformation, by interaction of residues on the flexible peptide with important DNA nucleotides, offers a general mechanism to provide reaction specificity. (2) The minimal multimer for concerted integration of both viral DNA ends is an integrase octamer. (3) An active integration complex should be assembled by a specific three-way complementation of integrase variants, no pair of which is sufficient to form an active multimer (*33*, *69*, *70*): (i) the core domain of integrase only (amino acids 50−212 of HIV-1 integrase), (ii) the amino-terminal and core domains with an inactivating mutation in the active site (amino acids 1−212 of HIV-1 integrase), and (iii) the core and C-terminal domain (amino acids 50−288 of HIV-1 integrase) with an inactivating mutation again in the active site, and that has also been treated with the alkylating agent *N*-ethylmaleimide to inhibit interaction with the amino-terminal domain. (4) The relative orientation of the two viral DNA ends can be tested by cross-

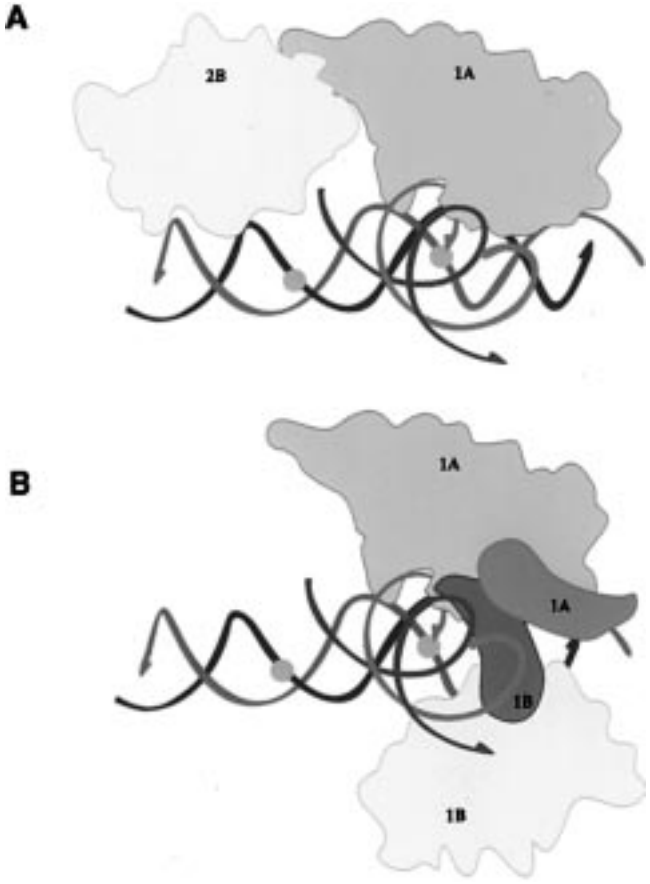

FIGURE 7: Model of "open" and "closed" integrase complexes, dependent on the N-terminal HHCC domain, that form different active integrase multimers. (A) In the absence of the HHCC domain, integrase is proposed to form only an "open" complex, requiring extensive target DNA interactions to stabilize an integrase oligomer that forms along one face of the target DNA. Weak protein−protein interactions between core dimers stabilize cooperative binding to the target DNA. While only two protomers are shown, propagation of this subunit interaction along the target DNA helical axis could form a higher-order oligomer. Viral DNA strands are colored in shades of blue, and target DNA strands are colored in shades of gray. The target DNA phosphate atoms where viral DNA joining occurs are indicated by light orange circles. Integrase core domains are colored as in previous figures. For simplicity, the second protomer of each type I integrase dimer is not shown. (B) In the presence of the N-terminal HHCC domain, integrase is proposed to form a "closed" complex with the DNA substrates. A possible position of the HHCC domain of protomer 1B is indicated in green. In this position, the HHCC domain of protomer 1B interacts with the core domain of protomer 1A on the opposite face of the target DNA and is close to target DNA features 5′ to the site of integration as well as to features at the viral DNA end. In this closed complex, target DNA interactions occur primarily with features less than 5 base pairs in the 5′ direction from the site of integration. The HHCC domain of protomer 1A is dispensable for integrase activity; its possible position in the complex is indicated in dark pink. The HHCC domains of protomers 1A and 1B are shown to interact; however, HHCC dimerization is not required for interaction of the 1A and 1B protomers. Moreover, interaction between the HHCC and core domains of monomers 1A and 1B is not likely to be reciprocal, as the HHCC domain of protomer 1A is not required for catalytic activity (*33*, *69*, *70*). For simplicity, the second protomer of each type I integrase dimer is not shown.

linking, by fluorescence energy transfer experiments, or by using tethered viral DNA ends. (5) Different residues of the A and B core domains should affect target-site selection; for example, H67 and N117 of the A protomer and S153 and K160 of the B protomer. This can be tested by examining the target-site specificity of complementing pairs of integrase variants that contain different combinations of mutations in the active-site residues and the residues predicted to affect target-site specificity. (6) Mutagenesis of residues P142−S147 and N117 or H67 on the different faces of the core domain that participate in viral and target DNA binding, respectively, should generate mutant proteins that are affected differentially for 3′-end processing and strand transfer activities.

The general architecture of our model is compatible with the structural coordinates of the avian sarcoma virus integrase core domain, which is very similar to the HIV-1 core domain (*46*). General features of the model may also be applicable to transposases such as MuA that share many biochemical and structural features with retroviral integrases. For example, domain sharing occurs among different MuA recombinase protomers (*77*, *78*), and the active complex is an intricate higher-order assembly where many protein−DNA interactions occur in trans to the active-site protomer (*53*, *54*, *79*). Biochemical studies have also identified clear differences in the architecture of the retroviral integrase and MuA complexes. In particular, both of the corresponding end processing and strand transfer activities of MuA are catalyzed only when residues in two different domains of the polypeptide assemble from different transposase protomers to form the fully functional active site (*77*−*80*). Additionally, the active MuA−DNA complex is assembled from MuA protomers that bind to three defined subsites on each Mu DNA end (*81*, *82*).

In conclusion, results of cross-linking experiments have helped to provide a framework for interpreting diverse experimental observations bearing on the structure of the active complex formed between integrase, viral DNA, and target DNA. The resulting physical model generates specific testable hypotheses regarding substrate recognition, architecture, and assembly of this essential intermediate in retroviral integration.

ACKNOWLEDGMENT

We thank Fred Dyda and Bob Craigie for providing unpublished data. We thank Jennifer Gerton, Joe DeRisi, all other members of the Brown laboratory, Dan Herschlag, Bill Weis, and Gilbert Chu for many helpful discussions. We also thank Rachel Crowley, Richard Sutton, and Linda McAllister for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Coordinates in PDB format of the final HIV-1 integration complex model. Access information is given on any current masthead page.

REFERENCES

1. Varmus, H. E., and Brown, P. O. (1989) *Retroviruses*, pp 53−108, American Society for Microbiology Press, Washington, DC.
2. Donehower, L. A., and Varmus, H. E. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6461−6465.
3. Panganiban, A. T., and Temin, H. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7885−7889.
4. Schwartzberg, P., Colicelli, J., and Goff, S. P. (1984) *Cell 37*, 1043−1052.
5. Hippenmeyer, P. J., and Grandgenett, D. P. (1985) *J. Biol. Chem. 260*, 8250−8256.
6. Colicelli, J., and Goff, S. P. (1985) *Cell 42*, 573−580.
7. Colicelli, J., and Goff, S. P. (1988) *J. Mol. Biol. 199*, 47−59.
8. Panganiban, A. T., and Temin, H. M. (1983) *Nature 306*, 155−160.
9. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987) *Cell 49*, 347−356.
10. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2525−2529.
11. Fujiwara, T., and Mizuuchi, K. (1988) *Cell 54*, 497−504.
12. Fujiwara, T., and Craigie, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3065−3069.
13. Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell 62*, 829−837.
14. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) *Cell 63*, 87−95.
15. Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989) *Cell 58*, 47−54.
16. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) *Science 255*, 723−726.
17. Bushman, F. D., and Craigie, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3458−3462.
18. Sherman, P. A., and Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5119−5123.
19. Vink, C., van Gent, D. C., Elgersma, Y., and Plasterk, R. H. A. (1991) *J. Virol. 65*, 4636−4644.
20. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol. 63*, 5319−5327.
21. Bushman, F. D., and Craigie, R. (1990) *J. Virol. 64*, 5645−5648.
22. Leavitt, A. D., Rose, R. B., and Varmus, H. E. (1992) *J. Virol. 66*, 2359−2368.
23. Van Den Ent, F. M. I., Vink, C., and Plasterk, R. H. A. (1994) *J. Virol. 68*, 7825−7832.
24. Vink, C., van Gent, D. C., and Plasterk, R. H. A. (1990) *J. Virol. 64*, 5219−5222.
25. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) *Genes Dev. 11*, 371−382.
26. Engelman, A., and Craigie, R. (1992) *J. Virol. 66*, 6361−6369.
27. Khan, E., Mack, J. P. G., Katz, R. A., Kulkosky, J., and Skalka, A. M. (1991) *Nucleic Acids Res. 19*, 851−860.
28. Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) *Mol. Cell. Biol. 12*, 2331−2338.
29. Leavitt, A. D., Shiue, L., and Varmus, H. E. (1993) *J. Biol. Chem. 268*, 2113−2119.
30. Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1993) *J. Virol. 67*, 425−437.
31. van Gent, D. C., Oude Groeneger, A. A. M., and Plasterk, R. H. A. (1993) *Nucleic Acids Res. 21*, 3373−3377.
32. Zheng, R., Jenkins, T. M., and Craigie, R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13659−13664.
33. Ellison, V., Gerton, J., Vincent, K. A., and Brown, P. O. (1995) *J. Biol. Chem. 270*, 3320−3326.
34. van Gent, C. C., Uode Groeneger, A. A. M., and Plasterk, R. H. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9598−9602.
35. Vink, C., Oude Groeneger, A. A. M., and Plasterk, R. H. A. (1993) *Nucleic Acids Res. 21*, 1419−1425.
36. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3428−3432.
37. Gerton, J., and Brown, P. O. (1997) *J. Biol. Chem. 272*, 25809−25815.
38. Engelman, A., Hickman, A. B., and Craigie, R. (1994) *J. Virol. 68*, 5911−5917.
39. Lutzke, R. A., Vink, C., and Plasterk, R. H. A. (1994) *Nucleic Acids Res. 22*, 4125−4131.

40. Heuer, T. S., and Brown, P. O. (1997) *Biochemistry 36*, 10655−10665.
41. Dyda, F., and Craigie, R., personal communication.
42. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science 266*, 1981−1986.
43. Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hard, K. (1995) *Nat. Struct. Biol. 2*, 807−810.
44. Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995) *Biochemistry 34*, 9826−9833.
45. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem. 271*, 7712−7718.
46. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodower, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol. 253*, 333−346.
47. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.
48. Heuer, T. S., and Brown, P. O., unpublished data.
49. Pruss, D., Bushman, F. D., and Wolffe, A. P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5913−5917.
50. Pruss, D., Reeves, R., Bushman, F. D., and Wolffe, A. P. (1994) *J. Biol. Chem. 269*, 25031−25041.
51. Eijkelenboom, A. P., van den Ent, F. M., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T. D., Plasterk, R. H., Kaptein, R., and Boelens, R. (1997) *Curr. Biol. 7*, 739−746.
52. Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol. 4*, 567−577.
53. Aldez, H., Schuster, E., and Baker, T. A. (1996) *Cell 85*, 257−269.
54. Savilahti, H., and Mizuuchi, K. (1996) *Cell 85*, 271−280.
55. Gerton, J., Morris, S., Olson, M., DeRisi, J., and Brown, P. O. (1998) *J. Virol.* (in press).
56. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J. 16*, 6849−6859.
57. Ellison, V., and Brown, P. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7316−7320.
58. Nilsen, B. M., Haugan, I. R., Berg, K., Olsen, L., Brown, P. O., and Helland, D. E. (1996) *J. Virol. 70*, 1580−1587.
59. Katzman, M., and Sudol, M. (1995) *J. Virol. 69*, 5687−5696.
60. Shibagaki, Y., and Chow, S. A. (1997) *J. Biol. Chem. 272*, 8361−8369.
61. Chow, S. A., and Brown, P. O. (1994) *J. Virol. 68*, 3896−3907.
62. Pryciak, P. M., Sil, A., and Varmus, H. E. (1992a) *EMBO J. 11*, 291−303.
63. Bor, Y.-C., Bushman, F. D., and Orgel, L. E. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10334−10338.
64. Muller, H.-P., and Varmus, H. E. (1994) *EMBO J. 13*, 4704−4714.
65. Pryciak, P. M., and Varmus, H. E. (1992) *Cell 69*, 769−780.
66. Hickman, A. B., Palmer, I., Engelman, A., Craigie, R., and Wingfield, P. (1994) *J. Biol. Chem. 269*, 29279−29287.
67. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M. (1992) *J. Biol. Chem. 267*, 16037−16040.
68. Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1339−1343.
69. Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J. 12*, 3269−3275.
70. van Gent, D. C., Vink, C., Oude Groeneger, A. A. M., and Plasterk, R. H. A. (1993) *EMBO J. 12*, 3261−3267.
71. Marmorstein, R., and Harrison, S. C. (1994) *Genes Dev. 8*, 2504−2512.
72. Rice, P., Craigie, R., and Davies, D. R. (1996) *Curr. Opin. Struct. Biol. 6*, 76−83.
73. Yang, W., and Steitz, T. A. (1995) *Cell 82*, 193−207.
74. Blake, D. G., Boocock, M. R., Sherratt, D. J., and Stark, W. M. (1995) *Curr. Biol. 5*, 1036−1046.
75. Rice, P., and Mizuuchi, K. (1995) *Cell 82*, 209−220.
76. Kwon, H. J., Tirumalai, R., Landy, A., and Ellenburger, T. (1997) *Science 276*, 126−131.
77. Baker, T. A., Mizuuchi, M., Savilahti, H., and Mizuuchi, K. (1993) *Cell 74*, 723−733.
78. Yang, J. Y., Kim, K., Jayaram, M., and Harshey, R. M. (1995) *EMBO J. 14*, 2374−2384.
79. Yang, J. Y., Jayaram, M., and Harshey, R. M. (1996) *Cell 85*, 447−455.
80. Wu, Z., and Chaconas, G. (1995) *EMBO J. 14*, 3835−3843.
81. Kuo, C. F., Zou, A. H., Jayaram, M., Getzoff, E., and Harshey, R. (1991) *EMBO J. 10*, 1585−1591.
82. Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9031−9035.

BI972949C