Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking

Timothy M. Jenkins, Dominic Esposito,
Alan Engelman¹ and Robert Craigie²

Laboratory of Molecular Biology, NIDDK, NIH, Bethesda,
MD 20892-0560 and ¹Division of Human Retrovirology,
Dana-Farber Cancer Institute, Boston, MA 02115, USA

²Corresponding author

Analysis of the crystal structure of HIV-1 integrase reveals a cluster of lysine residues near the active site. Using site-directed mutagenesis and photo-crosslinking we find that Lys156 and Lys159 are critical for the functional interaction of integrase with viral DNA. Mutation of Lys156 or Lys159 to glutamate led to a loss of both 3' processing and strand transfer activities *in vitro* while maintaining the ability to interact with nonspecific DNA and support disintegration. However, mutation of both residues to glutamate produced a synergistic effect eliminating nearly all nonspecific DNA interaction and disintegration activity. In addition, virus containing either of these changes was replication-defective at the step of integration. Photo-crosslinking, using 5-iododeoxyuracil-substituted oligonucleotides, suggests that Lys159 interacts at the N7 position of the conserved deoxyadenosine adjacent to the scissile phosphodiester bond of viral DNA. Sequence conservation throughout retroviral integrases and certain bacterial transposases (e.g. Tn10/IS10) supports the premise that within those families of polynucleotidyl transferases, these residues are strategic for DNA interaction.

Keywords: HIV-1/integrase/photo-crosslinking/
protein–DNA interaction

Introduction

Retroviral replication requires the formation of a provirus by integration of a DNA copy of the retroviral RNA genome into a host cell chromosome. Following reverse transcription, two nucleotides are cleaved from each 3' end of the linear viral DNA adjacent to a conserved CA dinucleotide. These 3' processed ends are then available for a direct transesterification reaction, termed strand transfer, in which the 3' viral ends are covalently joined to 5' phosphates in host DNA at the site of integration. Formation of the intact provirus requires removal of the unpaired nucleotides at the 5' ends of the viral DNA and repair of the single-strand gaps between the viral and host DNA; these latter reactions are likely to be carried out by cellular enzymes. For recent reviews on retroviral integration, see Vink and Plasterk (1993) and Katz and Skalka (1994).

Human immunodeficiency virus type 1 (HIV-1) inte-

grase is able to catalyze both the 3' processing and strand transfer activities *in vitro* with oligonucleotides that mimic viral DNA ends (Sherman and Fyfe, 1990; Bushman and Craigie, 1991; LaFemina *et al.*, 1991). Integrase can also catalyze an apparent reversal of the strand transfer reaction, termed disintegration (Chow *et al.*, 1992), in which viral and target DNA segments are resolved from a branched substrate that resembles the strand transfer product. Deletion mutagenesis and limited proteolysis have revealed that integrase is composed of three functional domains. The N-terminal domain contains a conserved HHCC motif (Johnson *et al.*, 1986) which binds Zn^{2+} (Burke *et al.*, 1992; Bushman *et al.*, 1993; Haugan *et al.*, 1995; Lee and Han, 1996; Zheng *et al.*, 1996). Binding of Zn^{2+} to this domain promotes tetramerization (Zheng *et al.*, 1996) and enhances catalytic activity (Lee and Han, 1996; Zheng *et al.*, 1996). The central core domain contains the active site residues Asp64, Asp116 and Glu152, which constitute the D,D-35-E motif common to all retroviruses, retrotransposons and some bacterial transposases (Rowland and Dyke, 1990; Engelman and Craigie, 1992; Kulkosky *et al.*, 1992; Polard and Chandler, 1995). Finally, the C-terminal domain, which displays little sequence homology among retroviruses, appears to bind DNA nonspecifically (Vink *et al.*, 1993; Woerner and Marcus-Sekura, 1993; Engelman *et al.*, 1994).

Structural information is now available on all three domains of integrase. Crystal structures of the core domain of HIV-1 integrase (Dyda *et al.*, 1994) and RSV integrase (Bujacz *et al.*, 1995) have revealed that integrase is a member of a structurally related superfamily of polynucleotidyl transferases (reviewed in Yang and Steitz, 1995; Rice *et al.*, 1996) that are proposed to use acidic residues to coordinate divalent metal ions during catalysis. This suggests that the retroviral integrases share similar chemical steps of catalysis with some cellular nucleases and polymerases. NMR structures have been obtained for both the C-terminal domain and the N-terminal domain of HIV-1 integrase. The C-terminal domain has an SH3-like fold that reveals a putative DNA-binding surface (Eijkelenboom *et al.*, 1995; Lodi *et al.*, 1995). The structure of the N-terminal domain (Cai *et al.*, 1997) is very similar to that of helix–turn–helix DNA-binding domains, such as the Trp repressor. However, one of the helices that contacts DNA in the case of Trp repressor forms part of the dimer interface in the case of the N-terminal domain of HIV-1 integrase. Interestingly, the structures of all three isolated domains are dimeric. This is consistent with the finding that only constructs containing both the catalytic core and C-terminal domains form both dimers and tetramers in solution, whereas the core domain alone is dimeric (Andrake and Skalka, 1995; Jenkins *et al.*, 1996). Complementation studies (Engelman *et al.*, 1993; van Gent *et al.*, 1993; Ellison *et al.*, 1995)

have established that integrase functions as a multimer, although the number of monomers that comprise the active form of the enzyme is as yet unknown.

To catalyze the cleavage and strand transfer reactions, integrase must distinguish functionally between viral DNA ends and nonspecific DNA sequences. Numerous studies using a variety of techniques (Grandgenett *et al.*, 1978; Roth *et al.*, 1988; van Gent *et al.*, 1991; Woerner *et al.*, 1992) have failed to observe directly specific DNA binding between integrase and viral DNA. However, functionally stable complexes can be formed between integrase and viral DNA ends in the presence of divalent metal ion (Ellison and Brown, 1994). Formation of these complexes requires the intact N-terminal, core and C-terminal domains of integrase, and complexes are relatively resistant to salt, detergent and competitor DNA (Ellison and Brown, 1994; Vink *et al.*, 1994). Nonspecific DNA binding of integrase has been localized to the C-terminal region of the protein (Vink *et al.*, 1993; Woerner and Marcus-Sekura, 1993; Engelman *et al.*, 1994) and Lys264 within the C-terminal minimal DNA-binding region (220–270), has been reported to be important for both nonspecific DNA binding and enzymatic activity (Puras Lutzke *et al.*, 1994).

The study presented here investigated whether residues within the catalytic domain of integrase were required for functional interaction with viral DNA. Two lysine residues (Lys156 and Lys159) have been identified, near the active site, which appear to be critical for the specific interaction of integrase with viral DNA. Mutation of either residue to glutamate leads to the loss of both 3' processing and strand transfer activities. These same mutations *in vivo* produced a replication-defective virus blocked at the integration step of the viral life cycle. Photo-crosslinking studies suggest that Lys159 interacts specifically at the N7 position of the deoxyadenosine of the conserved CA dinucleotide in viral DNA. This interaction between Lys159 and viral DNA orientates viral DNA such that the scissile phosphodiester bond is positioned close to the active site residues. The significance of these findings in relation to other polynucleotidyl transfer enzymes is discussed.

Results

Basic residues near the active site of HIV-1 integrase

Deletion mutagenesis studies have demonstrated that all three functional domains of HIV-1 integrase are required for 3' processing and strand transfer activity (Drelich *et al.*, 1992; Engelman *et al.*, 1993; Vink *et al.*, 1993), with the central catalytic domain (residues 50–212) being sufficient to catalyze disintegration or polynucleotidyl transfer (Bushman *et al.*, 1993) (Figure 1). An examination of the structure of the catalytic domain of HIV-1 integrase identified a cluster of surface lysine residues (Lys156, Lys159 and Lys160), located near the active site residues, which could potentially interact with viral DNA. Figure 2 shows both a space-filled and ribbon representation of the catalytic domain, with the active site residues Asp64, Asp116 and Glu152 highlighted in red and with Lys156 and Lys159 in blue. Alignment of residues near the active site glutamate residue of the D,D-35-E motif from

Fig. 1. Structural and functional domains of HIV-1 integrase. Integrase is a three domain multimeric protein, that contains conserved amino acid sequence motifs in both the N-terminus (HHCC) and core domain (D,D-35-E). Structural analysis of the core domain (Dyda *et al.*, 1994; Bujacz *et al.*, 1995) revealed that integrase is structurally related to a diverse superfamily of polynucleotidyl transferases. The isolated C-terminal domain, which can bind DNA, adopts a SH3 domain topology in solution (Lodi *et al.*, 1995; Eijkelenboom *et al.*, 1995). The structure of the isolated N-terminal domain, bound with zinc, is similar to that of a helix–turn–helix DNA-binding domain, such as the Trp repressor (Cai *et al.*, 1997). All three domains are required for integration activity. The correct multimeric organization of these domains in the active protein is presently unknown.

retroviral integrases and a number of bacterial transposases that appear to utilize a similar D,D-35-E motif (Figure 3), revealed that the amino acid at the position corresponding to Lys159 was completely conserved as lysine. However, the amino acid at the position corresponding to Lys156 was predominantly basic but never acidic. In contrast, there was no preference for the residue position corresponding to Lys160. In view of the position of these residues in the three-dimensional structure of the catalytic domain and their clear sequence conservation, mutant proteins were expressed and purified to assess their importance for enzymatic activity and DNA interaction.

Residues Lys156 and Lys159 are critical for integrase activity

Mutant proteins were made, as described in Materials and methods, either in the context of full-length (residues 1–288) or core domain (residues 50–212) of HIV-1 integrase. Lysine (K) was replaced by glutamate (E) at either residue position 156, 159 or 160. In the context of the core domain (residues 50–212), the individual mutations K156E and K159E dramatically reduced disintegration activity to undetectable levels (Figure 4B). However, the mutation K160E had no effect on the disintegration activity of the core domain. We then assessed the interaction of core domain mutants with disintegration substrate by measuring aggregation of fluorescently labeled substrate. In the presence of Mn^{2+} , the core domain of integrase aggregated disintegration substrate DNA, which could then be removed from solution by centrifugation (data not shown). However, the mutant core proteins IN^{50–212} K156E and IN^{50–212} K159E had a markedly decreased ability to interact with and aggregate the disintegration substrate. The effects of these mutations on disintegration activity in full-length integrase (Figure 4A) showed that mutation K156E reduced disintegration activity ~4-fold, mutation

Fig. 2. Location of residues Lys156 and Lys159 in reference to the active site of HIV-1 integrase. (A) Space-fill and (B) ribbon representations of the catalytic core domain of HIV-1 integrase (F.Dyda, personal communication). Images were produced using the MIDAS display system (Ferrin *et al.*, 1988; Huang *et al.*, 1991). Active site residues Asp64, Asp116 and Glu152 are colored in red, residues Lys156 and Lys159 are colored in blue. The disordered loop (residues 139–149) is represented with a dashed line.

Fig. 3. Protein sequence alignments of retroviral integrases and bacterial transposases around the HIV-1 integrase active site residue Glu152. The locations of Glu152, Lys156 and Lys159 are highlighted. This alignment is based on previous sequence alignments of Engelman and Craigie (1992) and Bolland and Kleckner (1996). Over 30 other members of the IS3 subfamily display this same consensus (M.Chandler, personal communication), in which the seventh residue after the active site glutamate residue is a conserved lysine, while the fourth residue is predominantly basic but never a negatively charged residue.

K159E displayed wild-type levels of activity, and the double mutant IN^{1-288} K156E/K159E was reduced ~20-fold in activity. Mutant IN^{1-288} K160E exhibited no decrease in disintegration activity (data not shown). Next we examined the activity of the full-length mutant proteins in both the viral 3' processing and strand transfer reactions. IN^{1-288} K156E and IN^{1-288} K159E appeared inactive for both 3' processing (Figure 5A) and strand transfer activities (Figure 5B) even at relatively high levels of protein. Mutant IN^{1-288} K160E displayed full activity for both 3' processing and strand transfer (data not shown).

IN^{1-288} K156E and IN^{1-288} K159E are not deficient in dimer-tetramer self-association

Purified integrases of various retroviruses have been reported to exist as monomers, dimers and tetramers in solution (Grandgenett *et al.*, 1978; Sherman and Fyfe, 1990; Jones *et al.*, 1992; Vincent *et al.*, 1993). We have previously demonstrated that the soluble mutant IN^{1-288} F185K/C280S exists in an equilibrium between dimeric and tetrameric species in solution, and that the core and C-terminal domains are necessary and sufficient for tetramerization (Jenkins *et al.*, 1996). We therefore examined the effects of both K156E and K159E mutations on the multimeric state of integrase by gel filtration. Gel

Fig. 4. Disintegration activities of HIV-1 integrase. The migration position of the 15mer substrate (S) and the 30mer product (P) are as indicated. Proteins were assayed at 37°C for 1 h. (A) Disintegration activity of full-length (1–288) mutants, at protein concentrations 0.04, 0.08, 0.16 and 1.28 μ M. (B) Disintegration activity of core domain (50–212) mutants, at protein concentrations 0.32, 0.64 and 1.28 μ M.

filtration, using a Superdex 200 column, was performed with the double mutant IN^{1-288} K156E/K159E loaded at a concentration of 0.5 mg/ml. The resulting single peak migrated at the position expected for a dimer (data not shown). When IN^{1-288} K156E/K159E was loaded at the higher concentration of 5 mg/ml, a single peak was again obtained, but this peak now migrated at the position expected for a tetramer of integrase (data not shown). Gel filtration of both IN^{1-288} K156E and IN^{1-288} K159E also revealed the same dimer-tetramer self-association in solution, which is indistinguishable from that of the soluble mutant IN^{1-288} F185K/C280S without any mutations in the lysine residues (data not shown). Near UV and far UV CD spectra from mutants IN^{1-288} K156E, IN^{1-288} K159E and IN^{1-288} K156E/K159E were indistinguishable from those obtained from IN^{1-288} F185K/C280S (data not shown). These gel filtration and CD studies therefore confirmed that mutation of Lys156 and/or Lys159 did not grossly change the secondary structure or multimeric state of integrase in solution.

Residues Lys156 and Lys159 are involved in interactions with viral DNA

UV crosslinking, at 254 nm, to both specific viral U5 DNA ends and nonspecific DNA was measured for the full-length mutant proteins IN^{1-288} K156E, IN^{1-288} K159E and IN^{1-288} K156E/K159E (Figure 6A). Similar levels of UV crosslinking were obtained with full-length integrase (IN^{1-288}) and mutant proteins IN^{1-288} K156E and IN^{1-288}

Fig. 5. Catalytic activities of full-length (1–288) HIV-1 integrase mutants. Proteins were assayed at 37°C for 1 h at protein concentrations of 0.09, 0.47 and 0.94 μ M and 80 nM DNA. (A) 3' Processing of integrase mutants; migration positions of 21mer substrate (S) and 19mer product (P) are as indicated. (B) Strand transfer activity of integrase mutants; migration positions of 19mer substrate (S) and strand transfer products (P) are as indicated.

K159E for both nonspecific and viral DNA. However, UV crosslinking to both viral DNA and nonspecific DNA was reduced 10-fold with mutant protein IN^{1–288} K156E/K159E.

Photo-crosslinking studies utilized oligonucleotides prepared with 5-iododeoxyuracil (IdU) substituting for the deoxyadenosine within the conserved CA dinucleotide adjacent to the scissile bond. The 3' end of this oligonucleotide was also biotin tagged to facilitate removal of unprocessed substrate DNA after photo-crosslinking. These modified substrates supported both 3' processing and strand transfer activities at 40% relative to wild-type ends (data not shown). Nonspecific oligonucleotides were also prepared containing the 5-IdU and 3' biotin at locations equivalent to that of the modified viral U5 end oligonucleotide. Radiolabeled specific and nonspecific duplex oligonucleotides were incubated with integrase under assay conditions and irradiated as indicated. Crosslinked photoproducts were separated from the free substrate DNA by electrophoresis and then analyzed by autoradiography. Efficient photo-crosslinking was readily detected between integrase and specific U5 viral DNA but not with nonspecific DNA (Figure 6B). A second nonspecific oligonucleotide substrate, varying at every position from the first nonspecific substrate DNA, also failed to crosslink to integrase (data not shown). To identify the location of the DNA photo-crosslink, proteolytic peptides were generated from the crosslinked integrase–DNA complex and the radiolabeled peptides sequenced. The majority of radioactive label was found to be confined to a single peptide (residues 158–198). Figure 6C, top panel, represents the N-terminal sequence of the fragment in the absence of photo-crosslinking; the

Fig. 6. UV crosslinking and specific photo-crosslinking of viral DNA to HIV-1 integrase. (A) UV crosslinking of full-length (1–288) HIV-1 integrase mutants to specific and nonspecific DNA. One μ M of each protein was incubated with 5 nM radiolabeled DNA for 30 min at 4°C, and then crosslinked by irradiating for 20 min at 254 nm. Substrate DNA (S) was either viral U5 DNA (specific) or a sequence which fails to match the viral U5 oligonucleotide at any position (nonspecific). The migration position of the substrate (S) and UV-crosslinked products (P) are as indicated. (B) Photo-crosslinking of specific and nonspecific 5-IdU-DNA substrates. One μ M of each protein was incubated with 5 nM radiolabeled 5-IdU-DNA for 90 min at 4°C, and irradiated for 30 min at 302 nm. Substrate DNA (S) was either viral 5-IdU-labeled U5 DNA (specific) or a 5-IdU-labeled sequence which failed to match the viral U5 oligonucleotide at any position (nonspecific). The migration positions of the substrate DNA (S) and photo-crosslinked products (P) are as indicated. (C) PTH-amino acid analysis of photo-crosslinked peptide samples. Gel-separated peptides were identified by Coomassie staining or autoradiography, excised and subjected to standard N-terminal sequencing on an Applied Biosystems 471A sequencer. PTH-amino acid recoveries were quantified using NIH Image software. Absolute recoveries in pmol were determined by comparison with known amounts of PTH-amino acid standards. Peptides analyzed were uncrosslinked peptide 158–198 (upper panel) and photo-crosslinked peptide 158–198 (lower panel). Each column corresponds to a single cycle of Edman degradation; the PTH-amino acid detected at each cycle is identified with the single-letter amino acid code above each bar. The amino acid position on the horizontal axis refers to the location of the residue in the context of full-length (1–288) HIV-1 integrase.

lower panel displays the N-terminal sequence of the same peptide now photo-crosslinked with labeled U5 viral DNA. The PTH-Lys159 derivative was conspicuously absent from the sequence of the radiolabeled peptide fragment. Furthermore, mutant IN¹⁻²⁸⁸ K156E was able to be photo-crosslinked efficiently to the specific U5 oligonucleotide, but mutant IN¹⁻²⁸⁸ K159E displayed only nonspecific levels of photo-crosslinking (Figure 6B). This identified Lys159 as being specifically crosslinked to U5 viral DNA at the position of the 5-IdU in the cleaved strand. Significantly, when 5-IdU was placed on the noncleaved strand, directly opposite the deoxyadenosine of the conserved CA dinucleotide, only nonspecific levels of crosslinking were observed (data not shown).

Mutations K156E and K159E block integration in infected cells

The K156E and K159E mutations greatly reduced the *in vitro* 3' processing and strand transfer activities of recombinant HIV-1 integrase. Based on these observations, the consequences of each of these changes were tested on the growth of HIV-1. Each mutation was introduced into the integrase-coding region of pNL4-3 (Adachi *et al.*, 1986), and a virus containing both the K156E and K159E changes was also analyzed. Virus stocks were produced by transfecting HeLa cells. Transfection of CD4⁺ cells bypasses early events in the viral life cycle which may be altered by mutations in integrase, such as reverse transcription and/or integration. The growth kinetics of each virus were then assessed following infection of CD4⁺ T-cell lines. Wild-type HIV-1 showed a peak of [³²P]reverse transcriptase ([³²P]RT) activity 10 days after infection (Figure 7A). Each mutation had a dramatic effect on the growth of HIV-1, as neither of the single mutant viruses showed any sign of replication over the 18-day observation period (Figure 7A). The same was true for the K156E/K159E double mutant virus. Each of the mutant plasmids yielded levels of [³²P]RT activity similar to the wild-type plasmid following transfection of HeLa cells (data not shown), indicating that the K156E and K159E substitutions did not grossly affect HIV-1 assembly and release (Engelman *et al.*, 1997). Following infection of 12D7-CEM T-cells with wild-type or mutant virus, viral DNA synthesis was monitored by PCR. This analysis revealed the expected 170 bp DNA product in cells infected with wild-type HIV-1 (Figure 7B, lane 1). Cells infected with the K156E, K159E and K156E/K159E mutant viruses also displayed this product, indicating that these mutations did not affect reverse transcription in infected cells (lanes 2–4). In fact, the level of PCR product was greater for cells infected with the mutant viruses than for wild-type infected cells (compare lanes 2–4 with lane 1). This phenotype has been previously described for HIV-1 integrase mutants carrying amino acid substitutions of the Asp and Glu residues of the D,D-35-E motif (Engelman *et al.*, 1995; Wiskerchen and Muessing, 1995; see Figure 7B, lane 5). Thus, the phenotype of HIV-1 harboring the K156E and K159E substitutions is similar to that of active site mutants, underscoring that the replication defect caused by these changes is at the integration step of the viral life cycle.

Fig. 7. The K156E and K159E changes block HIV-1 integration in infected cells. (A) Jurkat CD4⁺ T-cells were infected with either wild-type or the indicated mutant viruses. Cultures were split at 2-day intervals and aliquots were assayed for [³²P]RT activity. (B) 12D7-CEM T-cells were infected with the indicated viruses. The D116N mutant virus contains the substitution of Asn for integrase active site residue Asp116 (Engelman *et al.*, 1995). Cell lysates were subjected to PCR using primers which amplify the junction between the LTRs in circular viral DNA. The arrow indicates the 170 bp PCR product. The lysate used in lane 6 was from mock-infected cells. Molecular mass standards in bp are indicated on the right.

Discussion

Specific and nonspecific DNA interactions with integrase

Integrase is able to distinguish functionally between viral DNA ends and nonspecific DNA. However, integrase appears to bind equally well to both nonspecific DNA and specific viral DNA. Nonspecific DNA binding to integrase does not require the N-terminal domain (Schauer and Billich, 1992) or the presence of intact active site residues within the core domain (Engelman *et al.*, 1994), rather nonspecific DNA binding has been localized to residues 220–270 within integrase (Puras Lutzke *et al.*, 1994). The three-dimensional solution structure of this dimeric minimal DNA-binding region (Eijkelenboom *et al.*, 1995; Lodi *et al.*, 1995), reveals a putative DNA-binding cleft located within a saddle-shaped groove near the dimer interface. The surface of this cleft contains a number of basic residues that have been suggested to provide a DNA-binding region. One of the residues in this region, Lys264, has previously been implicated in nonspecific binding of DNA to integrase (Puras Lutzke *et al.*, 1994). Although most studies have focused upon the C-terminal domain of integrase, the isolated catalytic core domain of integrase has been reported to UV crosslink to the disintegration substrate in the presence of Mn²⁺ (Engelman *et al.*, 1994),

and show a preference for the conserved CA dinucleotide in an intramolecular cleavage–ligation reaction (Kulkosky *et al.*, 1995). However, no residues have yet been identified within the catalytic domain of integrase that are involved in DNA interaction or contribute to viral DNA specificity. Previously, it has been suggested that integrase may possess two distinct DNA-binding sites; a stable functionally defined viral DNA-binding site, and a functionally defined target or nonspecific DNA-binding site that can exchange more rapidly (Ellison and Brown, 1994). Here, we demonstrate that viral ends interact specifically with the catalytic domain of HIV-1 integrase and identify some of the functionally important residues in the integrase–viral DNA complex.

Conservation of basic residues adjacent to the active site residue Glu152

Retroviral protein sequence alignments near the active site residue Glu152 (E) of the D,D-35-E motif reveal conservation of residues corresponding to HIV-1 integrase residues Lys156 and Lys159 (Figure 3). The seventh residue after Glu152 is an absolutely conserved lysine, while the fourth residue is predominantly basic and never negatively charged. This consensus is seen not only in retroviruses, but also among a number of bacterial transposases that appear to possess a D,D-35-E motif (Figure 3). However, in retrotransposons, although the fourth residue is again predominantly basic, the seventh residue after Glu152 appears to lack a conserved lysine residue (Kulkosky *et al.*, 1992). In light of the results reported here, the D,D-35-E motif may be considered to be part of the extended motif D,D-35-E-7-K for retroviral integrases and certain families of bacterial transposases.

Functional interaction between HIV-1 integrase and DNA involves both Lys156 and Lys159

All three catalytic activities of integrase: 3' processing, strand transfer and disintegration, can be abolished by mutation of any of the active site residues Asp64, Asp116 or Glu152. Other mutations that exhibit deleterious effects on both the 3' processing and strand transfer activities, such as mutations within the conserved HHCC motif (Engelman and Craigie, 1992; van Gent *et al.*, 1992) or within the C-terminus (Puras Lutzke *et al.*, 1994), do not inhibit disintegration activity. The requirements for disintegration are therefore less stringent than for either 3' processing or strand transfer. In the context of the core domain, the mutations K156E or K159E completely inhibited disintegration (Figure 4B), and therefore exhibited a phenotype that resembles active site mutants. However, unlike active site mutants, disintegration activity was restored when the same individual mutations were made in the context of the full-length integrase (Figure 4A). These results are consistent with the involvement of residues K156 and K159 in the core domain, along with residues in the C-terminal domain, in DNA interaction.

Further evidence for the importance of residues K156 and K159 in the interactions between integrase and DNA comes from examination of the ability of these mutants to crosslink to DNA and/or aggregate with DNA. Neither the K156E nor the K159E mutant core domains aggregated the disintegration DNA substrate (data not shown). Full-length mutants containing only K156E or K159E were

able to UV crosslink to DNA (Figure 6A), support disintegration activity (Figure 4A) and aggregate with DNA (data not shown). However, introduction of the double mutation (K156/159E) into full-length integrase dramatically reduced UV crosslinking to DNA (Figure 6A), essentially abolished disintegration activity (Figure 4A) and led to a marked reduction in aggregation with DNA (data not shown). The disintegration activity of these full-length proteins therefore correlates strikingly with their ability to interact with DNA. Although full-length mutants containing either K156E or K159E permit nonspecific DNA interaction and disintegration activity, they are not able to support 3' processing or strand transfer activity (Figure 5A and B).

Additional evidence for the importance of Lys156 and Lys159 comes from recent studies with the Tn10/IS10 transposase. Sequence alignments of IS10 transposase and HIV-1 integrase show that integrase residues 156 and 159 are equivalent to Arg296 and Lys299 of Tn10/IS10 transposase (Figure 3). The mutation R296Q inhibits excision of Tn10 (Kennedy and Haniford, 1996). *In vitro* studies of this mutant demonstrated that cleavage activity was abolished and strand transfer activity was greatly reduced. Lys299 of IS10 has also been identified as being critical for transposition. The mutation K299A decreases strand transfer activity, in part due to an inability of the mutant protein to interact with and capture target DNA (Bolland and Kleckner, 1996).

Mutation K156E or K159E results in replication-defective virus

Numerous integrase mutations have been described which disrupt steps in the HIV-1 life cycle other than at integration. For example, some mutations affect virus assembly as the virus leaves cells (Shin *et al.*, 1994; Ansari-Lari *et al.*, 1995; Engelman *et al.*, 1995; Bukovsky and Goettlinger, 1996; Jenkins *et al.*, 1996), while others primarily display reverse transcription defects (Shin *et al.*, 1994; Taddeo *et al.*, 1994; Masuda *et al.*, 1995; Leavitt *et al.*, 1996; Engelman *et al.*, 1997). Mutation of either Lys156 or Lys159 to glutamate renders HIV-1 replication-defective (Figure 7A). Significantly, the phenotype of the mutant viruses closely resembles that of active site mutants. They are proficient for particle assembly, entry and reverse transcription, but defective at the integration step of the life cycle. This result is consistent with the involvement of these residues in viral DNA interaction. Substitutions of alanine for Lys156 or Lys159 (Wiskerchen and Muesing, 1995) or glutamine Lys159 (Shin *et al.*, 1994) have been reported not to affect viral replication overtly in cell culture; although substitution of proline at position 159 resulted in delayed replication kinetics (Cannon *et al.*, 1994). Previous *in vitro* data with different retroviral integrases have suggested that Lys159 may be important for integrase activity (Kulkosky *et al.*, 1992; van Gent *et al.*, 1992). Therefore, both *in vivo* and *in vitro* data suggest that, whereas acidic residues at position 156 or 159 abolish integrase activity, some more conservative substitutions may be tolerated.

Proposed roles for Lys156 and Lys159 within HIV-1 integrase

Despite numerous studies, conclusive evidence for specific interaction between integrase and viral DNA substrate at

the level of DNA binding has been lacking. Yet, it is clear that integrase is able to functionally recognize the viral end DNA substrate when activity is assayed, albeit with rather limited DNA sequence requirements. We decided to use structure-based mutagenesis and photo-crosslinking of halogenated base-containing DNA in an attempt to identify specific and functionally important contacts between integrase and viral DNA, before the steps of chemical catalysis. Halogenated nucleotides have previously been used successfully to probe protein–DNA contacts (Willis *et al.*, 1993). Crosslinking to the halogenated base does not require the presence of a specific amino acid; instead any amino acid can in theory be crosslinked so long as it lies in very close proximity to the modified base (Deitz *et al.*, 1987). In this study we have demonstrated that 5-IdU substituted for the deoxyadenosine of the conserved CA dinucleotide of viral DNA specifically crosslinks with Lys159 within the catalytic domain of HIV-1 integrase. The substitution of 5-IdU for deoxyadenosine introduces a mismatch into the viral DNA substrate. However, recent studies (Scottoline *et al.*, 1997) indicate that fraying or disruption of the terminal 3 bp occurs after the initial binding of viral DNA to integrase, and that mismatches in any of these terminal base pairs actually increase the K_{cat} for 3' processing.

This specific crosslink of viral DNA to Lys159 delineates the orientation of the viral DNA end as it is processed by HIV-1 integrase. The structure of DNA reveals that the C5 position of a pyrimidine (deoxyuracil) is directly equivalent to the N7 position of a purine (deoxyadenosine). Therefore, we suggest that Lys159 interacts with the N7 position of the deoxyadenosine in the conserved CA dinucleotide of viral DNA. Figure 8 shows Lys159 in relation to the active site residues Asp64, Asp116 and Glu152. It demonstrates that Lys159 is perfectly positioned to interact with and orientate viral DNA, such that the scissile phosphodiester bond is placed close to the active site residues. Substitution of a glutamate for lysine at position 159 would therefore be expected to reduce dramatically the ability of integrase to interact specifically with viral DNA ends and therefore adversely affect both the 3' processing and strand transfer activities of integrase (Figure 5A and B). In addition, Lys156 also appears to be an important residue for DNA interaction. Mutation of either Lys156 or Lys159 drastically affects DNA-dependent aggregation and disintegration in the context of the core domain of integrase. However, in full-length integrase, mutation of both Lys156 and Lys159 was required to produce a reduction in both disintegration activity and interaction with DNA.

At present, the multimeric organization of the active integrase–viral DNA complex is unknown. Presumably, other important viral DNA contacts exist within the catalytic core domain, and possibly within other domains of integrase, that determine viral DNA specificity. Nevertheless, it appears that the specific interaction of viral DNA with HIV-1 integrase requires the absence of a negatively charged residue at either position 156 or 159 to enable both 3' processing and strand transfer activities. This organization of strategically placed basic residues appears not only common to retroviruses but also to large families of bacterial transposases that appear to utilize a D,D,E motif to coordinate divalent metal ion for catalysis

Fig. 8. A detailed view of Lys159 and viral DNA in relation to the active site residues of HIV-1 integrase. Active site residues Asp64, Asp116 and Glu152 are colored red, while Lys159 is colored blue and viral DNA colored green. Lys159 is shown to be in close association with the N7 position of the deoxyadenosine of the conserved CA dinucleotide of viral DNA. This positions the viral DNA such that the scissile bond adjacent to the 3' phosphate is located near to the active site residues. This image was produced using the MIDAS display system (Ferrin *et al.*, 1988; Huang *et al.*, 1991).

of the chemical steps of phosphoryl transfer. Although high resolution structures are available for all three domains of HIV-1 integrase, this information reveals little about the architecture of the active complex with DNA substrate. Crosslinking provides an approach to impose constraints on the possible organization of the complex. In this report we have identified close contacts between Lys159 and the conserved nucleotide adjacent to the scissile bond. This result provides insights into the positioning of the viral DNA substrate with respect to the active site. Low resolution structural studies on active complexes between integrase and DNA, such as those reported here, provide a scaffold into which the high resolution structural data on the individual domains of integrase can eventually be modeled.

Materials and methods

Preparation of plasmids

Plasmids that encode the point mutations K156E, K159E and K160E were constructed in the context of both the soluble catalytic core domain (50–212) F185K mutant (Jenkins *et al.*, 1995) and the soluble full-length (1–288) F185K/C280S mutant (Jenkins *et al.*, 1996) of HIV-1 integrase. Site-directed mutations were generated using an ‘overlapping’ two-step PCR protocol as described (Higuchi *et al.*, 1988; Jenkins *et al.*, 1995). K156E, K159E and K156/159E mutations were also introduced into the integrase-coding region of pNL4-3, as previously described (Engelman *et al.*, 1995). Mutant sequences were confirmed by DNA sequencing (Sanger *et al.*, 1977).

Preparation of proteins

Proteins were expressed in *Escherichia coli* and purified from soluble extracts. All proteins contained a histidine tag (HT) at the N-terminus of the protein to facilitate purification. A thrombin cleavage site was present in the HT sequence to allow removal of the polyhistidine motif following purification. Core domain (50–212) mutants were purified from a 0.5 M NaCl lysis buffer [0.5 M NaCl, 20 mM hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES pH 7.5), 2 mM β -mercaptoethanol (β -ME), 0.3 mg/ml lysozyme, 5 mM imidazole] using an Ni-affinity (Chelating Sepharose Fast Flow, Pharmacia) column (5 $\times$ 7 cm), as described (Jenkins *et al.*, 1995). Full-length (1–288) mutants were purified from a 1 M NaCl lysis buffer (1 M NaCl, 20 mM HEPES pH 7.5, 2 mM β -ME, 0.3 mg/ml lysozyme, 5 mM imidazole) using a Ni-affinity column (5 $\times$ 7 cm) and then subsequently further purified using a Mono S HR 10/10 column (Pharmacia) as described (Jenkins *et al.*, 1996). The final concentrations of all purified proteins were determined using calculated extinction coefficients based on their amino acid composition (Wetlauffer, 1962).

Cells and viruses

HeLa cells were grown in Dulbecco's modified Eagle's medium supplemented to contain 10% fetal calf serum (FCS), 100 U penicillin G sodium and 0.1 mg streptomycin sulfate per ml. Cells (0.7×10^6) were seeded in 10 cm tissue culture plates 24 h prior to transfection. Virus stocks were prepared by transfecting cells with 20 μ g of plasmid in the presence of calcium phosphate (Sambrook *et al.*, 1989). Culture supernatants were tested for Mg²⁺-dependent [³²P]RT activity 48 h post-transfection as previously described (Engelman *et al.*, 1995). Supernatants were filtered through 0.45 μ m nitrocellulose and equal [³²P]RT c.p.m. of wild-type and mutant HIV-1 were used in subsequent infections.

The CEM-12D7 (Ross *et al.*, 1991) and Jurkat (Weiss *et al.*, 1984) CD4⁺ T-cell lines were grown in RPMI 1640 supplemented to contain 10% FCS, 100 U penicillin G sodium and 0.1 mg streptomycin sulfate per ml (RPMI-10). Cells (5.0×10^6) were infected with 5.0×10^5 [³²P]RT c.p.m. in 0.5 ml of RPMI-10 for 1 h at 37°C, prior to adding 5 ml of RPMI-10. Cultures were fed at 2-day intervals and aliquots of the medium were assayed for [³²P]RT activity. Each virus was tested in at least two independent experiments.

Analysis of viral DNA synthesis

Viral DNA synthesis was analyzed by PCR using primers which amplify sequences from the two long terminal repeat (LTR)-containing circle formed in the nuclei of infected cells. CEM-12D7 cells were infected as described above, and cells were lysed 18 h post-infection as previously described (Engelman *et al.*, 1995). PCR and electrophoresis conditions were as previously described (Engelman *et al.*, 1995).

Integrase activity assays

DNA substrates for disintegration, 3' processing and strand transfer activities were prepared as previously described (Engelman and Craigie, 1992; Jenkins *et al.*, 1996). Disintegration reaction mixtures contained 25 mM 3-morpholinopropanesulfonic acid (MOPS) pH 7.2, 10 mM β -ME, 10% (w/v) glycerol, 100 μ g/ml bovine serum albumin (BSA), 30 mM NaCl, 7.5 mM MnCl₂, 25 nM DNA substrate and 0.04–1.28 μ M integrase in a final 16 μ l volume. 3' Processing and strand transfer reactions mixtures contained 25 mM MOPS pH 7.2, 10 mM DTT, 5% (v/v) polyethylene glycol-8000, 5% (v/v) dimethylsulfoxide, 0.05% Nonidet P-40, 30 mM NaCl, 10 mM MgCl₂, 80 nM DNA substrate and 0.04–0.94 μ M integrase in a final 16 μ l volume. All samples were incubated for 1 h at 37°C and stopped with 16 μ l sequencing gel loading dye (95% formamide, 10 mM EDTA, 0.03% bromophenol blue and 0.03% xylene cyanol) and electrophoresed on 15% polyacrylamide-urea gels. Quantitation of reaction products was carried out using a PhosphorImager and ImageQuant 4.2 (Molecular Dynamics).

Determination of the multimeric state and secondary structure of integrase

The multimeric state of point mutations K156E, K159E and K160E in both the core domain and full-length (1–288) integrase was analyzed by gel filtration. Purified proteins at 0.3–5 mg/ml were applied to a calibrated Superdex 200 PC 3.2/30 column (Pharmacia) on a Pharmacia Smart™ System. The column was equilibrated with 25 mM HEPES pH 7.5, 1 M NaCl, 5 mM DTT, 1 mM EDTA and 10% (w/v) glycerol, and operated at a flow rate of 50 μ l/min at 4°C. CD measurements were carried out using a Jasco J-720 spectrometer (Jasco, Tokyo). Far UV spectra (200–260 nm) were measured at protein concentrations of ~0.3 mg/ml in a 0.1 mm path length quartz cuvette. Near UV spectra (250–320 nm) were

measured at protein concentrations of ~3–5 mg/ml in a 0.5 mm path length quartz cuvette. All spectra measurements were carried out at room temperature.

DNA-dependent aggregation studies

Fluorescent disintegration substrate was prepared from 5'-carboxy-fluorescein (FAM)-labeled oligonucleotide AE157 (5'-GAAAGCGACCGCGCC) purchased from Keck Oligonucleotide Synthesis Facility (Yale University). After purification through a 15% polyacrylamide-urea gel, AE157 was annealed with equimolar amounts of purified oligonucleotides AE146, AE117 and AE156, as previously described (Engelman and Craigie, 1992) to generate the disintegration substrate. Fluorescence measurements were obtained using a Fluorolog-2 Spectrofluorometer (SPEX Industries, Inc.), with an excitation maximum of 493 nm and an emission scanning spectrum of 505–550 nm. Emission maxima between 517 and 520 nm were obtained for FAM-labeled DNA. Reaction mixtures contained 25 mM MOPS pH 7.2, 10 mM β -ME, 10% (w/v) glycerol, 100 μ g/ml BSA, 30 mM NaCl, 1 mM MnCl₂, 5 nM DNA substrate and 50–200 nM core domain (50–212) protein in a final 3 ml volume. Reaction samples were mixed and the total fluorescent DNA substrate measured. Samples were then centrifuged at 16 000 *g* for 10 min at 4°C and the amount of soluble fluorescent DNA measured, and expressed as a percentage of total substrate.

Oligonucleotides for specific photo-crosslinking

Oligonucleotide DE4 (5'-GTGTGGAAATCTCTAGCIdUGT), which mimics the viral U5 end, was synthesized containing 5-IdU at the indicated position. DE4 also carried a biotin tag at the 3' end from synthesis on a BiotinTEG support (Glen Research, Sterling, VA). Gel-purified DE4 was labeled at the 5' end with [γ -³²P]ATP (3000 Ci/mmol; NEN) using T4 polynucleotide kinase (Pharmacia) and annealed with AE 117 as described (Engelman and Craigie, 1995). Nonspecific oligonucleotides were also synthesized with 3' biotin labels and 5-IdU incorporated at the identical positions to oligonucleotide DE4, but with no other sequence identities.

Specific photo-crosslinking and isolation of crosslinked peptides

Reaction mixtures contained the same components as the 3' processing reaction, with 5 nM DNA substrate and 0.94 μ M integrase in a final 16 μ l volume. Samples were incubated for 90 min at 37°C, transferred to a 96-well microtiter plate (Dynatech Laboratories, Inc.) and irradiated at 302 nm with a flux of 2 mW/cm² for 30 min. Samples were then centrifuged at 16 000 *g* for 10 min, and the pellet resuspended in buffer containing 25 mM HEPES pH 7.5, 5 mM EDTA, 10 mM DTT, 1 M NaCl and 6 M urea, and boiled for 5 min. Samples were then precipitated with 10% trichloroacetic acid (TCA), washed with acetone, dried and resuspended in buffer containing 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 5 mM DTT and 0.1 M NaCl. Biotinylated material was removed by adsorption to avidin-coated magnetic beads (Streptavidin Magnosphere, Promega, Inc.) according to the manufacturer's instructions. Sodium chloride concentration was increased to 1 M, SDS was added to a final concentration of 0.01%, and the samples boiled for 10 min. Glu-C (V8) protease (Boehringer Mannheim) was then added at 0.2 mg/mg of integrase and incubated at 37°C for 12 h. Samples were then precipitated with 10% TCA, washed with acetone, dried, resuspended in NuPage loading dye and separated on a 15% Tris-Tricine denaturing gel (Novex, Inc.). Peptides were transferred to a PVDF membrane (Trans-Blot, Bio-Rad Laboratories) using a Novex blotting apparatus. Radioactive peptides were identified by autoradiography, these bands were then excised and the peptide sequences determined using an ABI 471 protein sequencer (ABI Instruments, Inc.).

Oligonucleotides for UV crosslinking without IdU

UV crosslinking reaction mixtures contained the same components as the 3' processing reaction, with 80 nM DNA substrate and 0.04–0.94 μ M integrase in a final 16 μ l volume. Reaction mixtures were incubated on ice for 20 min and then 10 μ l aliquots were transferred to a 96-well microtiter plate (Dynatech Laboratories, Inc.). Samples were exposed to UV radiation at 254 nm, with a flux of 2.4 mW/cm² for 20 min at 4°C, boiled for 10 min with NuPage loading dye and electrophoresed on a 4–12% acrylamide NuPage gel (Novex, Inc.). UV crosslinking was quantified using a PhosphorImager and ImageQuant 4.2 (Molecular Dynamics).

Acknowledgements

We thank Kiyoshi Mizuuchi and Martin Gellert for careful reading of the manuscript. Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081). This work was supported by the National Institutes of Health Intramural AIDS Targeted Antiviral Program (R.C.) and NIH grant AI39394 (A.E.).

References

- Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R., Rabson,A. and Martin,M.A. (1986) Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.*, **59**, 284–291.
- Andrake,M.D. and Skalka,A.M. (1995) Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. *J. Biol. Chem.*, **270**, 29299–29306.
- Ansari-Lari,M.A., Donehower,L.A. and Gibbs,R.A. (1995) Analysis of human immunodeficiency virus type 1 integrase mutants. *Virology*, **211**, 332–335.
- Bolland,S. and Kleckner,N. (1996) The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell*, **84**, 223–233.
- Bujacz,G., Jaskolski,M., Alexandratos,J., Wlodawer,A., Merkel,G., Katz,R.A. and Skalka,A.M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.*, **253**, 333–346.
- Bukovsky,A. and Goettlinger,H. (1996) Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J. Virol.*, **70**, 6820–6825.
- Burke,C.J., Sanyal,G., Bruner,M.W., Ryan,J.A., LaFemina,R.L., Robbins,H.L. and Zeff,A.S. (1992) Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.*, **267**, 9639–9644.
- Bushman,F.D. and Craigie,R. (1991) Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. *Proc. Natl Acad. Sci. USA*, **88**, 1339–1343.
- Bushman,F.D., Engelman,A., Palmer,I., Wingfield,P. and Craigie,R. (1993) Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.
- Cai,M., Zheng,R., Caffrey,M., Craigie,R., Clore,G.M. and Gronenborn,A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.*, **4**, 567–577.
- Cannon,P.M., Wilson,W., Byles,E., Kingsman,S.M. and Kingsman,A.J. (1994) Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. *J. Virol.*, **68**, 4768–4775.
- Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723–726.
- Deitz,T.M., von Trebra,R.J., Swanson,B.J. and Koch,T.H. (1987) Photochemical coupling of 5-bromouracil (BU) to a peptide linkage. A model for BU-DNA protein photocrosslinking. *J. Am. Chem. Soc.*, **109**, 1793–1797.
- Drelich,M., Wilhelm,R. and Mous,J. (1992) Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology*, **188**, 459–468.
- Dyda,F., Hickman,A.B., Jenkins,T.M., Engelman,A., Craigie,R. and Davies,D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.
- Eijkelenboom,A.P.A.M., Puras Lutzke,R.A., Boelens,R., Plasterk,R.H.A., Kaptein,R. and Hard,K. (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Struct. Biol.*, **2**, 807–810.
- Ellison,V. and Brown,P.O. (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc. Natl Acad. Sci. USA*, **91**, 7316–7320.
- Ellison,V., Gerton,J., Vincent,K.A. and Brown,P.O. (1995) An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J. Biol. Chem.*, **270**, 3320–3326.
- Engelman,A. and Craigie,R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.*, **66**, 6361–6369.
- Engelman,A. and Craigie,R. (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J. Virol.*, **69**, 5908–5911.
- Engelman,A., Bushman,F.D. and Craigie,R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.*, **12**, 3269–3275.
- Engelman,A., Hickman,A.B. and Craigie,R. (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.*, **68**, 5911–5917.
- Engelman,A., Englund,G., Orenstein,J.M., Martin,M.A. and Craigie,R. (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J. Virol.*, **69**, 2729–2736.
- Engelman,A., Liu,Y., Chen,H., Farzan,M. and Dyda,F. (1997) Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. *J. Virol.*, **71**, 3507–3514.
- Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Langridge,R. (1988) The Midas display system. *J. Mol. Graphics*, **6**, 13–27.
- Grandgenett,D.P., Vora,A.C. and Schiff,R.D. (1978) A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology*, **89**, 119–132.
- Haugan,I.R., Nilsen,B.M., Worland,S., Olsen,L. and Helland,D.E. (1995) Characterization of the DNA-binding activity of HIV-1 integrase using a filter binding assay. *Biochem. Biophys. Res. Commun.*, **217**, 802–810.
- Higuchi,R., Krummel,B. and Saiki,R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.*, **16**, 7351–7367.
- Huang,C.C., Petersen,E.F., Klein,T.E., Ferrin,T.E. and Langridge,R. (1991) Conic: a fast renderer for space-filling molecules with shadows. *J. Mol. Graphics*, **9**, 230–236.
- Jenkins,T.M., Hickman,A.B., Dyda,F., Ghirlando,R., Davies,D.R. and Craigie,R. (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl Acad. Sci. USA*, **92**, 6057–6061.
- Jenkins,T.M., Engelman,A., Ghirlando,R. and Craigie,R. (1996) A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.*, **271**, 7712–7718.
- Johnson,M.S., McClure,M.A., Feng,D.F., Gray,J. and Doolittle,R.F. (1986) Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc. Natl Acad. Sci. USA*, **83**, 7648–7652.
- Jones,K.S., Coleman,J., Merkel,G.W., Laue,T.M. and Skalka,A.M. (1992) Retroviral integrase functions as a multimer and can turn over catalytically. *J. Biol. Chem.*, **267**, 16037–16040.
- Katz,R.A. and Skalka,A.M. (1994) The retroviral enzymes. *Annu. Rev. Biochem.*, **63**, 133–173.
- Kennedy,A.K. and Haniford,D.B. (1996) Isolation and characterization of IS10 transposase separation of function mutants: identification of amino acid residues in transposase that are important for active site function and the stability of transposition intermediates. *J. Mol. Biol.*, **256**, 533–547.
- Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P. and Skalka,A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.*, **12**, 2331–2338.
- Kulkosky,J., Katz,R.A., Merkel,G. and Skalka,A.M. (1995) Activities and substrate specificity of the evolutionary conserved central domain of retroviral integrase. *Virology*, **206**, 448–456.
- LaFemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.*, **65**, 5624–5630.
- Leavitt,A.D., Robles,G., Alesandro,N. and Varmus,H.E. (1996) Human immunodeficiency virus type 1 integrase mutants retain *in vitro* integrase activity yet fail to integrate viral DNA efficiently during infection. *J. Virol.*, **70**, 721–728.
- Lee,S.P. and Han,M.K. (1996) Zinc stimulates Mg²⁺-dependent 3'-processing activity of human immunodeficiency virus type 1 integrase *in vitro*. *Biochemistry*, **35**, 3837–3844.
- Lodi,P.J., Ernst,J.A., Kuszewski,J., Hickman,A.B., Engelman,A., Craigie,R., Clore,G.M. and Gronenborn,A.M. (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry*, **34**, 9826–9833.

- Masuda,T., Planelles,V., Krogstad,P. and Chen,I.S. (1995) Genetic analysis of human immunodeficiency virus type 1 integrase and the U3 att site: unusual phenotype of mutants in the zinc finger-like domain. *J. Virol.*, **69**, 6687–6696.
- Polard,P. and Chandler,M. (1995) Bacterial transposases and retroviral integrases. *Mol. Microbiol.*, **15**, 13–23.
- Puras Lutzke,R.A., Vink,C. and Plasterk,R.H.A. (1994) Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.*, **22**, 4125–4131.
- Rice,P., Craigie,R. and Davies,D.R. (1996) Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.*, **6**, 76–83.
- Ross,E.K., Buckler-White,A.J., Rabson,A.B., Englund,G. and Martin,M.A. (1991) Contribution of NF-kappa B and Sp1 binding motifs to the replicative capacity of human immunodeficiency virus type 1: distinct patterns of viral growth are determined by T-cell types. *J. Virol.*, **65**, 4350–4358.
- Roth,M.J., Tanese,N. and Goff,S.P. (1988) Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. *J. Mol. Biol.*, **203**, 131–139.
- Rowland,S.J. and Dyke,K.G. (1990) Tn552, a novel transposable element from *Staphylococcus aureus*. *Mol. Microbiol.*, **4**, 961–975.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Schauer,M. and Billich,A. (1992) The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. *Biochem. Biophys. Res. Commun.*, **185**, 874–880.
- Scottoline,B.P., Chow,S., Ellison,V. and Brown,P.O. (1997) Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.*, **11**, 371–382.
- Sherman,P.A. and Fyfe,J.A. (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.
- Shin,C.G., Taddeo,B., Haseltine,W.A. and Farnet,C.M. (1994) Genetic analysis of the human immunodeficiency virus type 1 integrase protein. *J. Virol.*, **68**, 1633–1642.
- Taddeo,B., Haseltine,W.A. and Farnet,C.M. (1994) Integrase mutants of human immunodeficiency virus type 1 with a specific defect in integration. *J. Virol.*, **68**, 8401–8405.
- van Gent,D.C., Elgersma,Y., Bolk,M.W., Vink,C. and Plasterk,R.H.A. (1991) DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Res.*, **19**, 3821–3827.
- van Gent,D.C., Groeneger,A.A.M. and Plasterk,R.H.A. (1992) Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc. Natl Acad. Sci. USA*, **89**, 9598–9602.
- van Gent,D.C., Vink,C., Groeneger,A.A.M. and Plasterk,R.H.A. (1993) Complementation between HIV integrase proteins mutated in different domains. *EMBO J.*, **12**, 3261–3267.
- Vincent,K.A., Ellison,V., Chow,S.A. and Brown,P.O. (1993) Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. *J. Virol.*, **67**, 425–437.
- Vink,C. and Plasterk,R.H.A. (1993) The human immunodeficiency virus integrase protein. *Trends Genet.*, **9**, 433–438.
- Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.*, **21**, 1419–1425.
- Vink,C., Puras Lutzke,R.A. and Plasterk,R.H.A. (1994) Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. *Nucleic Acids Res.*, **22**, 4103.
- Weiss,A., Wiskocil,R.L. and Stobo,J.D. (1984) The role of T3 surface molecules in the activation of human T cells: a two-stimulus requirement for IL 2 production reflects events occurring at a pre-translational level. *J. Immunol.*, **133**, 123–128.
- Wetlauffer,D.B. (1962) Ultraviolet spectra of proteins and amino acids. *Adv. Prot. Chem.*, **17**, 303–390.
- Willis,M.C., Hicke,B.J., Uhlenbeck,O.C., Cech,T.R. and Koch,T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, **262**, 1255–1257.
- Wiskerchen,M. and Muesing,M.A. (1995) Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J. Virol.*, **69**, 376–386.
- Woerner,A.M. and Marcus-Sekura,C.J. (1993) Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.*, **21**, 3507–3511.
- Woerner,A.M., Klutch,M., Levin,J.G. and Marcus-Sekura,C.J. (1992) Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. *AIDS Res. Hum. Retrovir.*, **8**, 297–304.
- Yang,W. and Steitz,T.A. (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, **3**, 131–134.
- Zheng,R., Jenkins,T.M. and Craigie,R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl Acad. Sci. USA*, **93**, 13659–13664.

Received on August 6, 1997; revised on September 8, 1997