The EMBO Journal Vol. 20 No. 13 pp. 3565–3576, 2001

Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes

Kui Gao, Scott L.Butler and Frederic Bushman[1]

Infectious Disease Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

[1]Corresponding author
e-mail: bushman@salk.edu

Early steps of retroviral replication involve reverse transcription of the viral RNA genome and integration of the resulting cDNA copy into a chromosome of the host cell. The viral-encoded integrase protein carries out the initial DNA breaking and joining reactions that mediate integration. The organization of the active integrase–DNA complex is unknown, though integrase is known to act as a multimer, and high resolution structures of the isolated integrase domains have been determined. Here we use site-specific cross-linking based on disulfide bond formation to map integrase–DNA contacts in active complexes. We establish that the DNA-binding C-terminal domain of one integrase monomer acts with the central catalytic domain from another monomer at each viral cDNA end. These data allow detailed modeling of an integrase tetramer in which pairs of *trans* interactions link integrase dimers bound to substrate DNA. We also detected a conformational change in integrase–DNA complexes accompanying cleavage of the viral cDNA terminus.

Keywords: AIDS/cross-linking/HIV/integration/recombination

Introduction

Integrase (IN) is the only one of three human immunodeficiency virus (HIV)-encoded enzymes that has not been exploited as a target for anti-viral drugs. Since IN is required for viral replication, and there is no cellular equivalent to IN, it is widely viewed as a promising new drug target (for reviews see Coffin *et al*., 1997; Pommier *et al*., 1997; Hansen *et al*., 1998). Structural information on IN–DNA complexes is potentially valuable for guiding the design of IN inhibitors, particularly since the response of correctly assembled IN complexes to small molecules differs from that of free IN protein (Farnet *et al*., 1996; Hazuda *et al*., 1999, 2000).

The DNA cutting and joining reactions involved in integration have been characterized in detail (for reviews see Coffin *et al*., 1997; Hansen *et al*., 1998). First, IN removes two nucleotides from the 3′-terminus of each strand of the linear viral cDNA. In the next step, the exposed 3′-hydroxyl groups attack the host DNA and become joined by a direct transesterification reaction (Engelman *et al*., 1991). Unfolding of this intermediate yields gaps at each host–virus DNA junction that are probably repaired by host DNA repair enzymes. Purified IN can carry out both the terminal cleavage and strand transfer reactions *in vitro* (Katzman *et al*., 1989; Bushman *et al*., 1990; Craigie *et al*., 1990; Katz *et al*., 1990; Sherman and Fyfe, 1990), and purified cellular gap repair enzymes can complete the gap repair step (Yoder and Bushman, 2000).

IN protein is comprised of three domains (Coffin *et al*., 1997; Hansen *et al*., 1998). The structure of each domain alone has been determined by X-ray crystallography or NMR spectroscopy, and structures of two-domain fragments have also been elucidated. The N-terminal domain contains a conserved motif, $HX_{3–7}HX_{23–32}CX_2C$ that binds zinc (Cai *et al*., 1997; Eijkelenboom *et al*., 1997). The catalytic domain contains a highly conserved $D,DX_{35}E$ motif embedded in a protein fold that is highly conserved among polynucleotide phosphotransferase enzymes (Dyda *et al*., 1994; Bujacz *et al*., 1995; Goldgur *et al*., 1998; Maignan *et al*., 1998; Greenwald *et al*., 1999). The three acidic amino acids comprise a highly conserved metal-binding motif that is required for cleavage and joining by IN and other enzymes. The C-terminal domain contains a less conserved sequence, though the overall structure resembles that of the Src-homology 3 (SH3) domain (Eijkelenboom *et al*., 1995; Lodi *et al*., 1995). Mutational studies indicated that this domain contributes to DNA binding. All three domains alone were proposed to form dimers from structural data (Dyda *et al*., 1994; Bujacz *et al*., 1995; Lodi *et al*., 1995; Cai *et al*., 1997; Eijkelenboom *et al*., 1997; Maignan *et al*., 1998; Greenwald *et al*., 1999), and the two-domain structures, containing the catalytic and C-terminal domains, also formed dimers (J.C.-H.Chen *et al*., 2000; Z.Chen *et al*., 2000; Yang *et al*., 2000).

IN *in vitro* displays similar affinities for DNAs matching the correct viral cDNA substrate and non-specific DNAs, complicating studies of IN–DNA complexes. Nevertheless, functional assays have shown that IN can distinguish between viral cDNA ends and non-specific oligonucleotides. Mutational studies *in vivo* and *in vitro* of the viral cDNA ends have established that the CA/TG dinucleotide pair (residues 3 and 4) at each end is required for efficient catalysis, and the terminal 6–12 bp also contribute (LaFemina *et al*., 1991; Esposito and Craigie, 1998). It remains unclear how IN recognizes the viral cDNA termini during replication.

Studies using photo-cross-linkable groups in synthetic DNA substrates have provided some information on segments of IN close to parts of the substrate, providing constraints useful in modeling IN–DNA contacts (Heuer and Brown, 1997, 1998; Jenkins *et al*., 1997; Esposito and Craigie, 1998). However, these data have not allowed discrimination of the specific contacts made by different

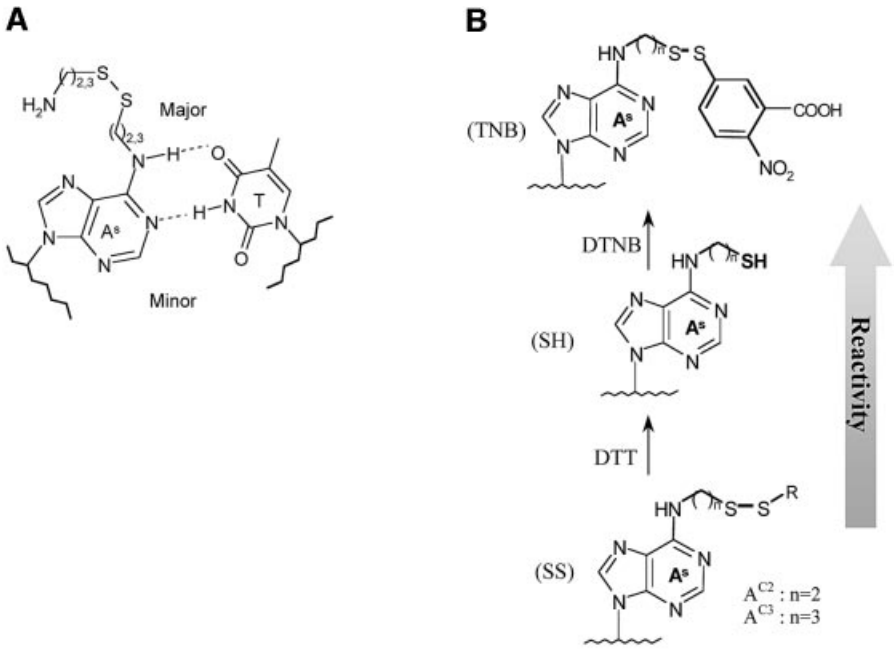

Fig. 1. Oligodeoxynucleotide modifications used for site-specific cross-linking. (**A**) Structure of the modified adenine residue in an A–T base pair. (**B**) Structures of the different modified thio groups used for cross-linking. The relative reactivity increases from the bottom to the top of the figure.

monomers in multimeric complexes. Nevertheless, a cross-linking study from Heuer and Brown (1998) was able to show that IN monomers that do not participate in catalysis may contribute much of the DNA-binding activity, and proposals for the parts of IN involved could be made, supporting initial efforts at modeling IN–DNA complexes.

Here, we have investigated the organization of IN–DNA complexes in detail using a site-specific cross-linking method based on disulfide bond formation. Thio groups were introduced in selected locations in the DNA and new cysteine residues were introduced into the protein so that cross-linking joins the modified DNA to the new cysteine residues. Accumulation of the cross-linked products establishes that a cysteine residue is bound near the modified position on the DNA in the complex. This method thus results in formation of a covalent bond between predetermined groups in the protein and DNA, in contrast to published photo-cross-linking methods that link multiple groups in a protein to a cross-linkable group tethered to DNA. Moreover, the disulfide-mediated method requires closer approach of the two interacting groups than typical photo-cross-linking methods, and the specificity of disulfide cross-linking is more readily controlled (see below). Previously Verdine, Harrison and co-workers have used this method to cross-link HIV-1 reverse transcriptase to DNA and then crystallized the resulting complexes (Huang *et al*., 1998, 2000).

We engineered new cysteine residues in HIV-1 IN and tested their reactivity with thio-modified substrate DNAs, allowing the identification of groups that could form disulfide bonds efficiently. Subsequent studies using these IN–DNA covalent complexes revealed that the C-terminal domain acts *in trans* to the catalytic domain at each cDNA end and specified aspects of the IN–DNA contacts involved. Similar catalysis *in trans* has been reported for the related transposase proteins Tn*5* and MuA (Aldaz *et al*., 1996; Savilahti and Mizuuchi, 1996; Davies *et al*., 2000; Naumann and Reznikoff, 2000). These data allowed us to propose a structural model incorporating the *trans* interaction in which IN formed an interlocked tetramer. As with the prokaryotic models, the requirement for *trans* interactions may serve to couple completion of assembly to activation of catalysis. In addition, we used site-specific cross-linking to detect a conformational change in the IN–DNA complex accompanying the terminal cleavage reaction. These data provide new information on IN complexes and document the utility of disulfide-mediated cross-linking in exploring the architecture of multimeric protein–DNA complexes.

Results

Design of the DNA substrates

In order to introduce the cross-linkable thio groups into DNA substrates, we first introduced a nucleotide analog by DNA synthesis, then attached the desired thio-tether arm post-synthetically (described in Huang *et al*., 1998, 2000 and references therein). DNA strands were modified to contain mixed disulfides linked by either ethyl or propyl tether arms (called ‘C2’ or ‘C3’ arms hereafter).

The reactivity of the thio groups on the tether arms could be controlled by further post-synthetic modification (Figure 1). The initial synthesis yielded the mixed disulfide form, which could become cross-linked to cysteine groups in IN by disulfide exchange. The reactivity of the thio group in disulfide exchange reactions could be increased by treatment with 5-thio-2-nitrobenzoic acid (TNB) to yield the TNB-modified form. A species with intermediate reactivity could be formed by treatment with dithiothreitol (DTT), yielding the free sulfhydryl, which reacts with cysteine residues after air oxidation. Previous

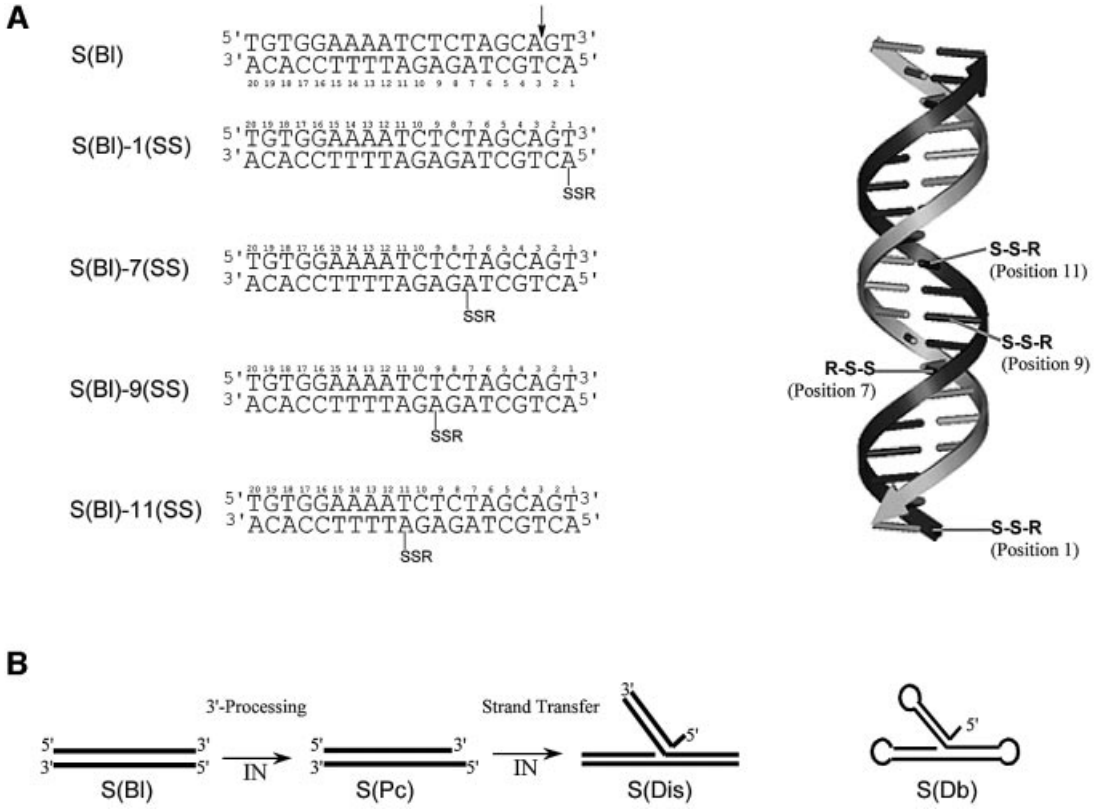

Fig. 2. Derivatives of the viral cDNA end used in this study. (**A**) The blunt-ended substrate S(Bl) mimics the wild-type viral cDNA at the U5 end. The site of cleavage by integrase is denoted by an arrow. Sites of modification are as illustrated. (**B**) The types of integrase substrates used in this study.

studies have described cross-linking of protein to DNA with the mixed disulfide (Huang *et al*., 1998, 2000) and cross-linking of DNA bases with TNB-modified molecules (Wolfe *et al*., 1995). Cross-linking using the sulfhydryl form has not been reported before to our knowledge. As is discussed below, comparison of results with the three thio modifications allowed optimization of reactivity and specificity in cross-linking reactions.

Figure 2A illustrates the DNA substrates studied. The thio-alkyl tether arms were introduced into adenine residues on the bottom strand of the substrates at positions 1, 7, 9 and 11 (numbered from the end of viral cDNA). Only pre-existing adenine groups were modified in order to preserve the wild-type DNA sequence. Addition of the tether arm did not disrupt the capacity of the adenine group for base pairing. The attachment of the tether arm to the N^6 position placed the thio-alkyl group in the DNA major groove, with the four modified positions sampling much of the region of the major groove from 1 to 11.

Four types of DNA substrates were used (Figure 2B). The blunt-ended form S(Bl) mimicked the immediate product of reverse transcription. Substrate S(Pc) mimicked the viral cDNA end after cleavage by IN. A third mimicked the products of integration, named S(Dis) for the disintegration reaction carried out by IN on this substrate (see below). The fourth is the same as the disintegration substrate, but with the DNA strands connected by hairpins to yield a dumbbell form [S(Db)]. The modified DNAs are specified below by (i) a number indicting the position of the modification, (ii) the length of the tether arm and (iii) the type of modification.

Cysteine substitutions in HIV-1 IN

We introduced new cysteine residues into the IN protein to allow site-specific cross-linking with the modified cDNA substrates. X-ray structures of one- and two-domain derivatives of HIV-1 (Dyda *et al*., 1994; Lodi *et al*., 1995; Greenwald *et al*., 1999; J.C.-H.Chen *et al*., 2000) allowed identification of surface-exposed residues for cross-linking tests. However, structures of IN–DNA complexes are not available, necessitating an initial scanning step to identify positions of efficient cross-linking. In our principal series of experiments, we focused on cross-linking to residues in the C-terminal region. Altering this part of IN was also chosen to minimize the possible disruptions of the newly introduced cysteine residues, since changes to the catalytic domain have more chance of reducing enzymatic activity, and cysteine substitutions in the N-terminal domain could disrupt zinc binding.

Changes (I191C, S230C, E246C and R262C) were introduced in and around basic regions in the two-domain structure that are candidate DNA-binding surfaces. I191 is located in a loop region between the catalytic domain and an α-helix extending to the C-terminus. E246, R262 and S230 are on the surface of the C-terminal domain (Figure 3A). Cysteine substitutions were generated in two backgrounds. In the first, four cysteine residues had been removed from the wild-type protein (C56S, C65S, C130S and C280S) in an effort to maximize the specificity of the cross-linking reaction. This left only two cysteine residues in the N-terminal domain, C40 and C43, which are involved in zinc binding. In the second set of mutants,

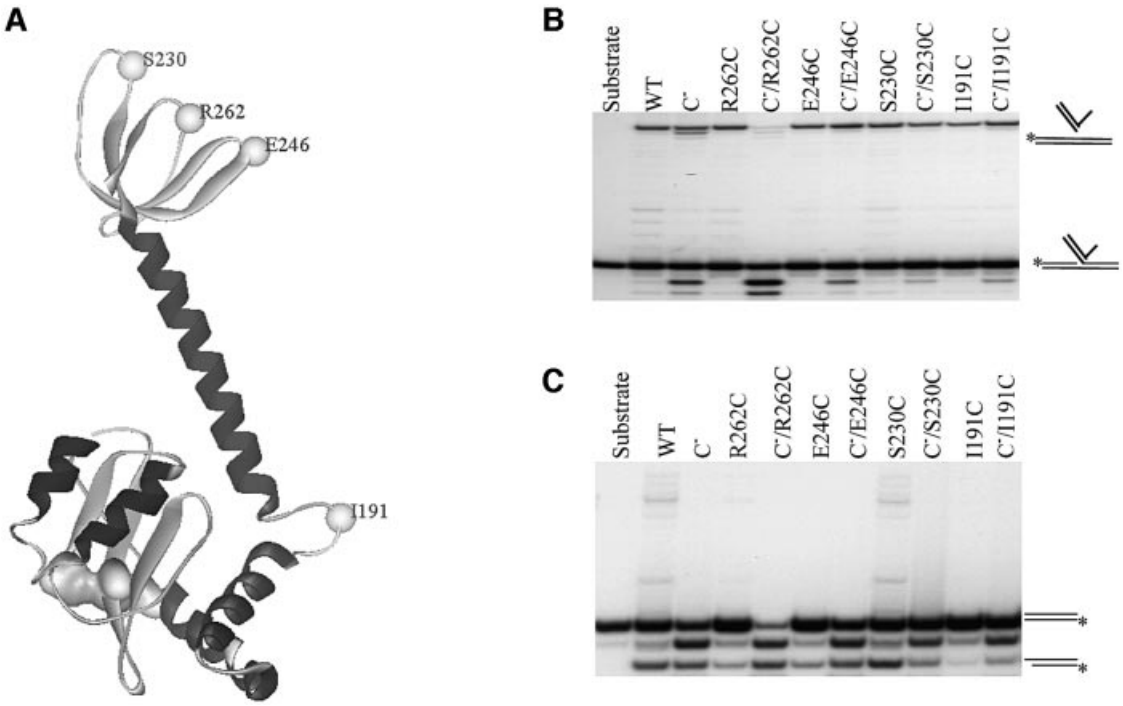

Fig. 3. IN mutants used in cross-linking reactions. (**A**) Illustration of the positions of the mutation sites in the two-domain (50–270) HIV-1 IN structure (J.C.-H.Chen *et al*., 2000) (1EX4). (**B**) Disintegration assay. (**C**) 3′-processing activity assay. The DNA substrates and products are as illustrated beside the gel.

cysteine substitutions were introduced in the wild-type IN background.

The catalytic activities of the modified proteins were tested in assays of disintegration and terminal cleavage. Substrates were 5′-^{32}P labeled then incubated with IN proteins in the presence of Mn^{2+}. The products were analyzed by 20% denaturing PAGE. All proteins except C^{-}/R262C displayed robust disintegration activity (Figure 3B). All proteins also displayed terminal cleavage activity, though several were less active than wild type (I191C, E246C and R262C) and many formed less strand transfer product than the wild type (Figure 3C). Previously, mutants E246A and R262G were tested and found to have reduced terminal cleavage activity (Lutzke and Plasterk, 1998). Assays of some of the Cys-minus IN derivatives yielded unexpected DNA products shorter than the starting substrate, a result of either abnormal IN hydrolysis activities or contaminating nucleases. Proteins in the Cys-minus background were found to aggregate during purification to a greater extent than wild type, so the wild-type background was favored in most experiments.

Disulfide cross-linking of IN with viral cDNA substrates

Figure 4 illustrates cross-linking reactions with wild-type IN and R262C, S230C, E246C and I191C (generated in the background of the wild-type IN sequence). The figure compares the results of cross-linking reactions with the DTNB-activated DNA (Figure 4A), the sulfhydryl form (Figure 4B) and the mixed disulfide form (Figure 4C). For the mixed-disulfide form, the C2 and C3 tether arms were compared. Viral cDNA substrates were end labeled with ^{32}P prior to cross-linking, then products were separated by SDS–PAGE and visualized by autoradiography.

Using the highly reactive TNB-modified substrates S(Bl)-C2(TNB), it was not possible to detect specific cross-linking (Figure 4A). For all substrates, position 1 in the DNA was the most reactive, probably because the terminal base pair can be unpaired transiently, allowing the modified nucleotide more flexibility. Reactions with the wild-type protein yielded abundant cross-linked product, indicating that some of the cysteine residues in the wild-type sequence can be cross-linked to position 1. E246C displayed the highest relative cross-linking efficiency compared with the other IN derivatives.

In order to enhance the specificity of the cross-linking reaction, assays with the intermediate reactivity sulfhydryl-C2 substrates S(Bl)-C2(SH) were compared (Figure 4B). Reactions with the E246C protein yielded the most cross-linked product, with cross-linking to the substrate with the tether arm at position 7 now most efficient, though reactions with substrates modified at positions 1, 9 and 11 also yielded abundant product. For all five proteins studied, the position 1-modified substrate could be cross-linked efficiently. Since substantial product accumulated in reactions with wild-type IN, it was apparent that naturally occurring cysteine residues could be cross-linked to position 1. In other experiments, we found that C65 and C280 each can be cross-linked to position 1 in an otherwise Cys-minus background, documenting two of the residues involved (data not shown).

To increase the specificity of the cross-linking reaction still further, we used the less reactive mixed disulfide C2 and C3 tether arms [S(Bl)-C2(SS) and S(BL)-C3(SS)]. Much less cross-linked product was detected (Figure 4C). Considerable cross-linking occurred between residue E246C and C3 at position 7 (Figure 4C). Little cross-linking was seen with the C3 arm at position 1 or with the

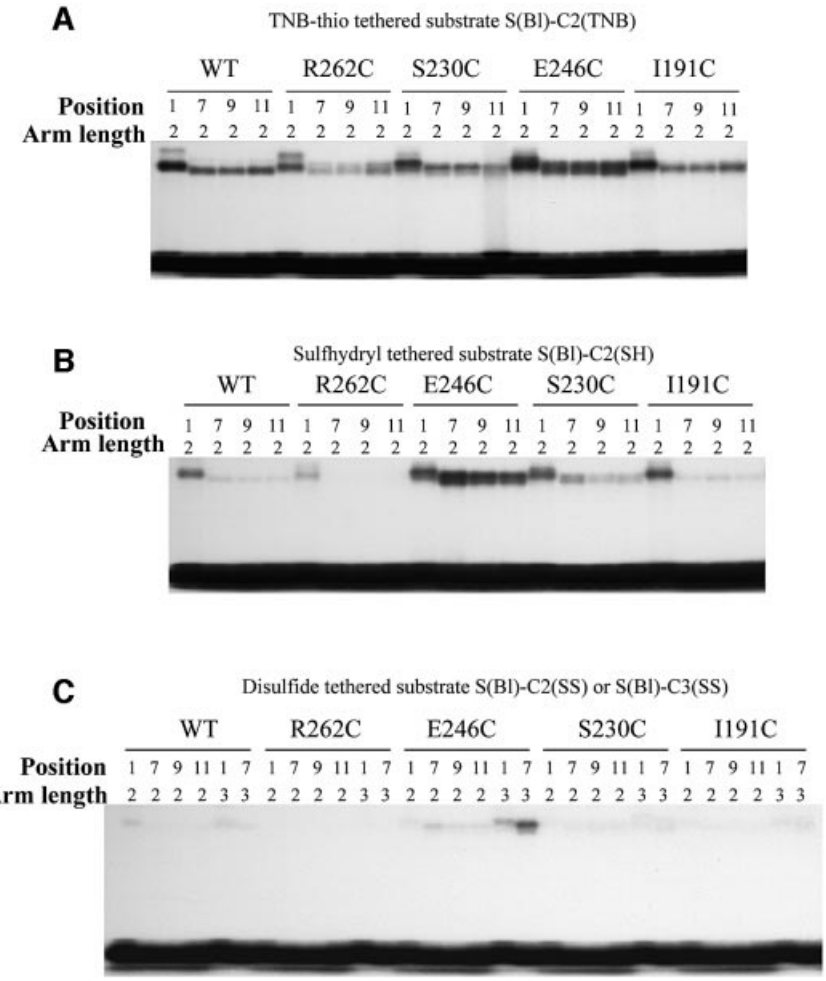

Fig. 4. Cross-linking of IN mutants to the viral cDNA end S(Bl) containing different tether arms. The positions of the tether arm and the arm length are indicated above the gels. (**A**) IN mutants (wild type, R262C, S230C, E246C and I191C) cross-linked to S(Bl) carrying the most reactive DTNB-modified tether arm [S(Bl)-C2(TNB)]. (**B**) IN wild type and mutants cross-link to DNA substrates carrying the intermediate reactivity SH tether arm (S(Bl)-C2(SH)). (**C**) IN mutants cross-linked to DNA substrates carrying the least reactive SS tether arm [S(Bl)-C2(SS) or S(Bl)-C3(SS)].

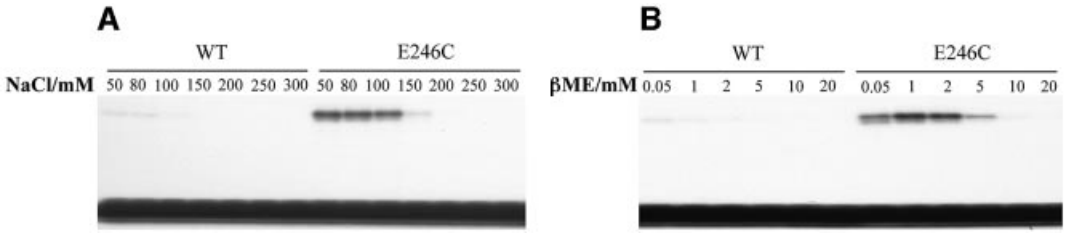

Fig. 5. Effects of salt concentration (**A**) and βME concentration (**B**) on cross-linking of E246C with S(Bl)-7C3(SS). Wild-type IN is also included as a control. Concentrations are indicated above the gel.

other IN derivatives. Of the C2 spacer arms, position 7 was also the optimum for E246, though the amount of product formed was much less than with C3. Thus adjusting the reactivity of the thio group and the length of the tether arm allowed formation of a highly specific covalent complex between Cys246 and position 7. The data in Figure 4A–C illustrate the trade off between reactivity of the tethered thiol and specificity of the cross-linked product.

A control experiment was carried out to investigate the requirements for the cross-linking reaction (Figure 5A). Cross-linking of IN E246C to position 7 [disulfide-C3 form, S(Bl)-7C3(SS)] was tested in the presence of increasing concentrations of NaCl. Too high a concentration of NaCl is expected to disrupt the interactions required for complex formation. The cross-linking reaction was largely inhibited at 150 mM NaCl and completely inhibited at 200 mM NaCl, conditions that inhibit the reaction *in vitro*, suggesting that correct complex formation is required for the cross-linking reaction.

To test the expectation that cross-linking required disulfide bond formation, the effect of β-mercaptomethanol (βME) concentration in the reaction was tested

(Figure 5B). High concentrations of βME (>5 mM) were found to block the formation of cross-linking products. The activity of IN was not decreased in the presence of 5 mM βME (data not shown), suggesting that the inhibition of cross-linking was not due to a lack of complex formation. The presence of a low concentration of βME is expected to increase the specificity of the reaction, since the formation of the disulfide is reversible under these conditions, requiring that the reacting thiols be apposed by complex formation for product to accumulate. Thus, a low concentration of βME (2 mM) was used in cross-linking experiments with the mixed disulfide and sulfhydryl-modified DNAs (for the case of DTNB derivatives, βME was not added because it would react with the modified DNA itself).

The autoradiograms shown probably underestimate the extent of cross-linking of E246C to the C3 modification at position 7 (typically ~20–50% by PhosphorImager analysis). βME was presented in the sample loading buffer, and gel samples were heated prior to electrophoresis. This was required to allow the samples to enter the gel but severed some of the disulfide bonds joining the cross-linked species. Evidence presented below indicates that most of the DNA is cross-linked to E246C prior to heating in the presence of βME.

Several cysteine substitutions in the catalytic domain were also tested (data not shown). These studies were carried out using IN 50–288 derivatives in which all cysteine residues were modified to serine, then new cysteine modifications introduced. Position 1 could be cross-linked to several positions (65, 160, 170 or 280) in this background using DTNB-modified substrates. These findings are consistent with the expectation that these residues are exposed on the protein surface and that the catalytic domain can bind near the viral cDNA terminus.

E246C-cross-linked C-domain binds in trans to the catalytic domain at the viral cDNA end

Modeling of IN–DNA complexes has been complicated by a lack of information on the relative positions of different IN monomers bound to DNA. The identification of a site-specific cross-link between E246C and position 7 in the viral cDNA allowed the relative locations of the catalytic and C-terminal domains to be examined. In one model for the complex, a monomer contributing the catalytic domain acting on one viral cDNA end also contributes the C-terminal domain that becomes cross-linked to position 7 (*cis* action). Alternatively, different monomers could contribute the catalytic domain and the C-terminal domain contacts (*trans* action). We investigated this issue by testing the effects of an active site mutation, D64R, in different combinations with monomers containing the cross-linkable E246C modification. A related strategy was used previously by Heuer and Brown (1998) (see Discussion).

Our experimental strategy is illustrated in Figure 6A. For this, we used the dumbbell disintegration substrate S(Db)-7C3(SS), containing the disulfide-C3 at position 7, labeled on the 5′ end with ^{32}P. The dumbbell substrate was used because the complex of IN cross-linked to product has a different mobility on SDS–PAGE than the IN-cross-linked substrate. Two IN proteins, D64R/E246C and wild-type IN, were mixed prior to addition of the substrate. The DNA was then added and incubated with the IN proteins in the absence of Mn^{2+}, so cross-linking could take place but not catalysis. After cross-linking, Mn^{2+} was added to start the reaction. The reaction products were separated by SDS–PAGE and analyzed by autoradiography. IN protein cross-linked to the disintegration substrate migrates more slowly than the form cross-linked to product because of the difference in DNA mass. If the catalytic domain binds *in cis* to the C-terminal domain, then no disintegration product should be found in the cross-linked complex. However, if different monomers can contribute the two domains, then IN cross-linked to disintegration product will accumulate.

Reactions with E246C bearing an intact active site yielded cross-linked forms containing mostly disintegration products (Figure 6B, lane 1). Cross-linking of the catalytically defective mutant E246C/D64R yielded only the covalent complex containing the substrate DNA (Figure 6B, lane 2). We next investigated whether wild-type IN added *in trans* could rescue catalysis in a complex of E246C/D64R cross-linked to the DNA. Indeed, the disintegration substrate linked to E246C/D64R was found to be converted efficiently to disintegration product (Figure 6B, lane 4). Addition of IN D64R instead of wild type did not support disintegration product formation (Figure 6B, lane 3). This supports the idea that the catalytic domain can act *in trans* to another monomer contributing the C-terminal domain contact to position 7, information which is useful in modeling multimeric IN–DNA complexes (see Discussion).

An important question is whether DNAs in the IN–DNA complexes can be released and re-bound to other IN monomers during the *in vitro* assays. It would confound the interpretation if DNAs could encounter the wild-type IN, be converted to product and then bind to E246C/D64R and become cross-linked. To investigate the rate of exchange, we determined whether product formation required pre-mixing of IN E246C/D64R and wild-type protein. In fact, addition of wild-type IN after cross-linking the E246C/D64R monomer failed to yield disintegration product (either cross-linked, Figure 6B, lane 5, or free, data not shown), indicating that added IN monomers could not enter the cross-linked IN–DNA complex under these conditions. Repeating the experiment by adding the cross-linkable E246C form after cross-linking DNA to E246C/D64R also failed to yield product (Figure 6B, lane 6). These data argue against the idea that the DNA substrate was converted to product, released from the enzyme, then rebound and cross-linked. The formation of the stable complex did require cross-linking, however. Figure 6B, lanes 7–10 indicates that the cross-linking of disintegration products to IN D64R/E246C could take place efficiently if the DNAs were pre-incubated with non-cross-linkable wild-type IN forms, indicating exchange. The above data taken together suggest that all of the DNA substrate is sequestered by cross-linking to the catalytically defective monomer, because any free substrate would be expected to become converted to product by the added wild-type IN (Figure 6, lanes 5 and 6).

However, the results in Figure 6B may be subject to another interpretation. In the reaction shown in Figure 6B,

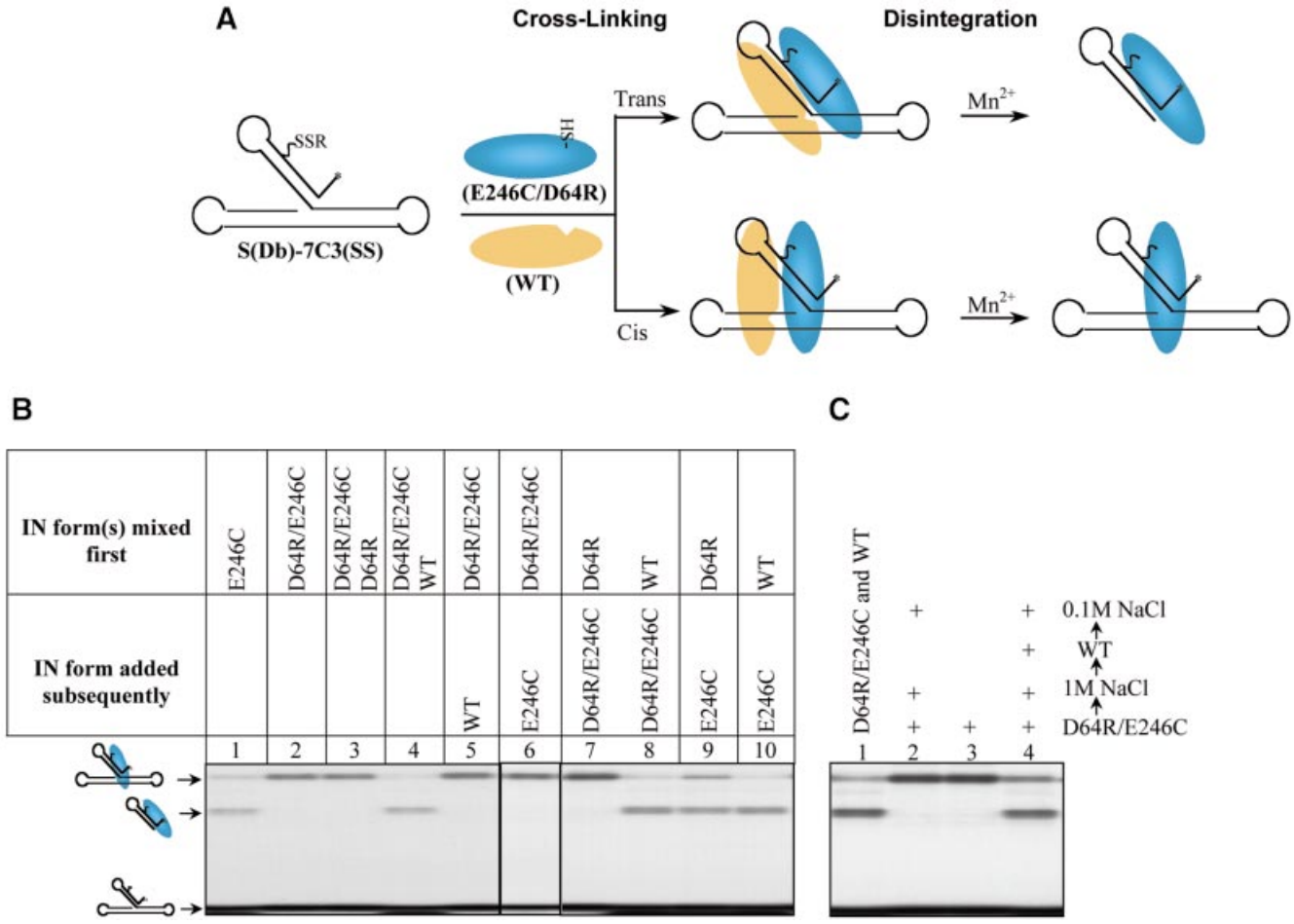

Fig. 6. *Cis–trans* test of integrase function. (**A**) Schematic illustration of the *cis–trans* analysis. The dumbbell disintegration substrate S(Db)-7C3(SS) was pre-incubated with the indicated integrase derivatives to allow cross-linking, then the disintegration was initiated by addition of Mn^{2+}. The predictions for *cis* or *trans* action with IN E246C/D64R and wild-type (WT) proteins are shown. See the text for details. (**B**) Autoradiogram illustrating the results of the complementation assay. Whether proteins were added sequentially or together in a pre-incubation step is specified above the gel. The structure of the products formed is illustrated beside the gel. (**C**) Staging of the cross-linking to D64R/E246C and addition of wild-type IN. The S(Db)-7C3(SS) substrate was cross-linked to D64R/E246C, and then subjected to further treatments: either no treatment (lane 3), exposure to 1 M NaCl, then reduction of the salt to 0.1 M NaCl (lane 2), or exposure to 1 M NaCl, addition of wild-type IN, then reduction of the salt to 0.1 M NaCl (lane 4). For comparison, a reaction in which D64R/E246C was pre-mixed with wild-type IN is shown in lane 1. Incubation with 1 M NaCl was for 30 min, then the salt concentration was reduced by dilution in buffer, the solution was adjusted to 5 mM $MnCl_2$ and the samples incubated an additional 20 min at 37°C. Samples were concentrated using a microcon YM-3 unit prior to electrophoresis.

lane 4, could the DNA sequestered in a mixed complex containing D64R/E246C and wild-type IN have become converted to product by wild-type IN, then switched binding partners and become cross-linked to D64R/E246C? Such an exchange, according to this idea, would take place within the IN–DNA complex, misleadingly indicating *trans* cleavage.

To address this issue, we staged the reaction so that the DNA substrate was first cross-linked to D64R/E246C, then wild-type IN introduced into the complex in a subsequent step (Figure 6C). Figure 6C, lane 1 shows the mobility of D64R/E246C cross-linked to product, generated by pre-mixing D64R/E246C and wild-type IN. Figure 6C, lane 3 shows the mobility of D64R/E246C cross-linked to the unreacted substrate. To introduce active IN into the complex, a reaction in which D64R/E246C was first cross-linked to substrate was treated with 1 M NaCl, and wild-type IN was added subsequently (Figure 6C, lane 4). Upon lowering the salt concentration to 0.1 M NaCl and addition of 5 mM Mn^{2+}, much of the cross-linked substrate was converted to product, indicating that the high salt treatment allowed the wild-type IN access to the cross-linked DNA. Without the high salt treatment, the wild-type IN cannot enter the cross-linked complex (Figure 6B, lanes 5 and 6). A reaction in which the salt concentration was raised and lowered, but no wild-type IN was added, yielded only the substrate cross-linked to D64R/E246C, as expected (Figure 6C, lane 2). The simplest interpretation of this experiment is that the added wild-type IN was able to enter the cross-linked D64R/E246C–DNA complex after disruption with 1 M NaCl and convert cross-linked substrate to product. These data further support the idea that the catalytic domain of one monomer acts *in trans* to a different monomer donating the C-terminal E246C–position 7 contact.

We also attempted to study domain interactions with another active site mutant, IN D116N. However, disintegration reactions containing mixtures of the IN D116N

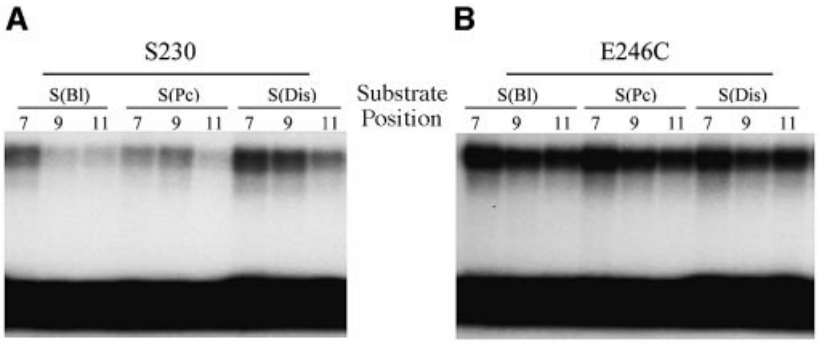

Fig. 7. Biochemical assay of conformation change in different steps of integration. (**A**) Cross-linking assay of IN S230C with the indicated substrates. (**B**) Cross-linking assay of IN E246C with the indicated substrates. In all cases, the C2(SH) tether arm was used.

mutant and wild-type proteins yielded little product, indicating that the IN D116N mutant form acted as a dominant-negative inhibitor *in vitro* (data not shown). Thus we could not use the IN D116N mutant in complementation tests as described above. Why the two active site mutants behaved differently is unknown.

A conformational change in the IN–DNA complex

Previous studies have suggested that IN–DNA complexes undergo conformational changes during the integration reaction (Ellison and Brown, 1994; Asante-Appiah *et al*., 1998; Greenwald *et al*., 1999; Espeseth *et al*., 2000). The site-specific cross-linking method is well suited to studying such changes, since disulfide formation is highly sensitive to local changes in the positions of the reacting thiol groups. We investigated possible conformational changes by testing cross-linking to substrates mimicking different reaction steps. S(Bl) mimicked the blunt-ended product of reverse transcription, S(Pc) the post-cleavage cDNA end, and S(Dis) the product of integration. Thioethyl tether arms at positions 7, 9 and 11 were assayed. The sulfhydryl form of the C2-tether arm was used because the less reactive disulfide yielded little product, while the highly reactive DTNB-modified form showed little specificity. The shorter C2 tether arm was used to reduce flexibility in order to maximize the sensitivity to small conformational changes. Each substrate was tested in reactions with the four cysteine-substituted IN derivatives described above (Figure 7 and data not shown).

Assays with S230C revealed a prominent difference in cross-linking efficiency with the different substrates. In substrate S(Bl), position 7 was the most efficient site for cross-linking, showing an ~3:1 preference over positions 9 and 11. However, in S(Pc), position 9 formed cross-linked product most efficiently, while in S(Dis) positions 7 and 9 were about equal in efficiency. Thus the complexes with substrates mimicking different reaction steps displayed different conformations as revealed by the cross-linking assay. Reactions with E246C, in contrast, were not sensitive to the differences among S(Bl), S(Pc) and S(Dis), with each displaying a preference for position 7. These results begin to provide a specific molecular picture of IN conformational changes during the reaction.

Discussion

Here we present a cross-linking study that revealed specific contacts between IN and the viral cDNA despite the non-specific binding seen with IN in most assays. We initially tested cross-linking with different combinations of thio-modified DNAs and cysteine residues in or near the IN C-terminal domain. Many combinations yielded cross-linked product with the DTNB-activated disulfide and sulfhydryl-containing DNAs, necessitating use of more discriminating disulfide exchange reactions to reveal specific cross-linking. Reactions with IN E246C and DNA containing the mixed-disulfide at position 7 [S(Bl)-7C3(SS)] yielded abundant product under conditions in which other combinations of modified INs and DNAs yielded little or no product. The DNA substrate with the C3 thio-propyl tether arm formed cross-linked product to IN E246C at a higher rate than the C2-modified substrate, indicating the importance of optimizing the tether arm length as well. Formation of the cross-linked product was disrupted by addition of high concentrations of NaCl, consistent with the expectation that correct IN–DNA complex formation is required for specific cross-linking. These findings indicate that use of the relatively vigorous TNB-modified reagent permitted cross-linking of physically distant or transient species, while use of the less reactive reagents permitted cross-linking of only stably bound and precisely positioned thiol groups. Inclusion of βME in reactions also allowed disulfide bonds to be formed under equilibrium conditions, reducing kinetic trapping of species not dependent on complex formation (Huang *et al*., 2000).

We found that the IN–DNA complex at each long terminal repeat end involved two monomers, one contributing the C-terminal domain and the other contributing the catalytic domain. This was deduced from reactions in which various mixtures of active and catalytically inactive IN mutants were cross-linked to a DNA substrate, then the monomers bound to DNA products identified. Related results were reported by Heuer and Brown (1998) based on photo-cross-linking studies, which indicated that a catalytically inactive monomer could nevertheless contribute much of the DNA-binding activity required for catalysis. Their data also allowed proposals to be made for possible contacts between DNA and the bound IN domains. However, the high reactivity of the photo-cross-linking

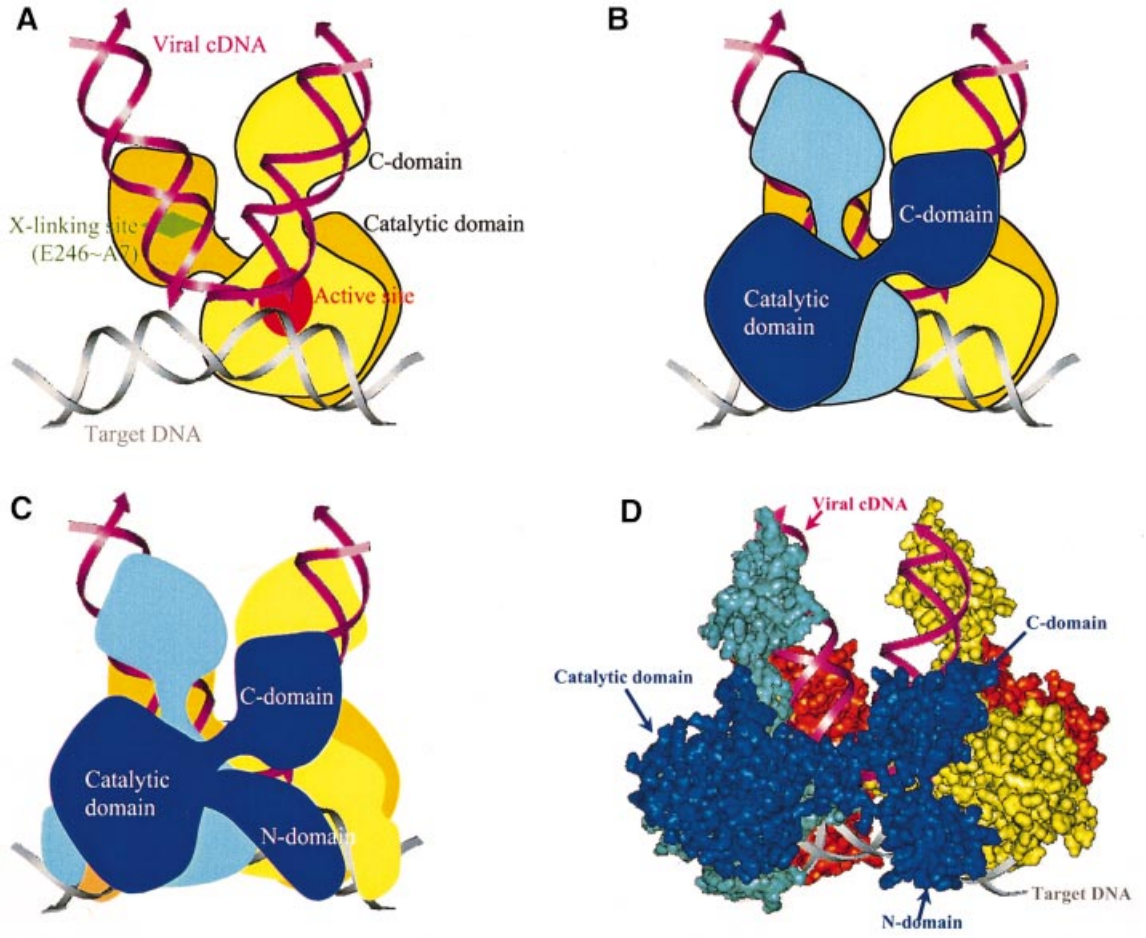

Fig. 8. A candidate model for the IN–DNA complex based on cross-linking and structural studies. (**A**) Illustration of the modeled interaction between one IN two-domain dimer and DNAs (target DNA, gray; viral cDNA ends, pink). The position of the active site in the catalytic domain is shown in red. The green diamond indicates the position of the E246–A7 cross-link. (**B**) Illustration of a complex containing two dimers of the two-domain IN bound to the DNA substrates. (**C**) Organization of an IN tetramer bound to DNA, illustrating a possible location for the N-terminal domain. (**D**) Overview of the complex model with integrase shown as the calculated Van der Waal's surface (shaded light blue and blue, yellow and brown for the four monomers). Viral cDNA ends are indicated as pink ribbons, target DNA (bent) is indicated by the gray ribbons. Coordinates for the model are available upon request.

reagent used resulted in formation of heterogeneous cross-linked species in the products of each reaction, complicating straightforward interpretation. An advantage of disulfide-mediated cross-linking is that under favorable circumstances it can yield a homogeneous cross-linked product, though the method does require a relatively laborious initial scanning step to find sites in the protein and DNA that can be cross-linked specifically.

Studies of related DNA transposition complexes have also revealed an interwoven arrangement of enzyme subunits, an architecture that has been proposed to couple completion of assembly with activation of catalysis. The X-ray structure of Tn*5* transposase bound to DNA revealed such a *trans* complex, and biochemical studies have documented *trans* catalysis for Tn*5* and phage Mu transposases (Aldaz *et al*., 1996; Savilahti and Mizuuchi, 1996; Davies *et al*., 2000; Naumann and Reznikoff, 2000). Such a requirement suppresses partial reactions that potentially could yield abnormal DNA products deleterious to the element or host. For HIV IN, under relatively physiological conditions *in vitro* (as in Fitzgerald *et al*., 1992; Engelman and Craigie, 1995; Miller *et al*., 1995), requirements for catalysis generally match requirements for coupled joining of the two cDNA ends, suggesting that correct assembly may also regulate catalysis by HIV IN.

We have generated an initial model of the IN–DNA complex using constraints from disulfide-mediated cross-linking, photo-cross-linking and published structural data. For the DNA, the points of joining of the HIV cDNA ends are known to be separated by 5 bp across a DNA major groove. The integration target DNA is modeled as bent, since DNA bending is known to favor integration (Pryciak and Varmus, 1992; Muller and Varmus, 1994; Bor *et al*., 1995). Multimerization by IN has been detected in biochemical and biophysical studies, and multimerization during catalysis is implied by studies indicating complementation among integrase mutants (Engelman *et al*., 1993; van Gent *et al*., 1993; Heuer and Brown, 1998), so IN is modeled as a multimer.

The recently solved structure of an IN dimer containing the catalytic and C-terminal domains (J.C.-H.Chen *et al*., 2000) could be arranged on the DNA substrates to satisfy the observed *trans* interaction (illustrated in Figure 8A). The catalytic site from one monomer contacts the right

viral DNA end as drawn, while the other monomer donates E246 to make the position 7 contact to the left cDNA end, thus making the *trans* contact. Figure 8B shows that a second dimer can be added so as to make two-fold symmetric DNA contacts, yielding a well packed tetramer. In the model, the α-helices linking the catalytic and C-terminal domains extend across the complex, linking IN complexes at the two viral cDNA ends. In Figure 8C, the N-terminal domains (Cai *et al*., 1997) have been positioned in the space between two IN dimers based on the location of the N-terminus of the catalytic domain in the model and photo-cross-linking data (Heuer and Brown, 1997). Figure 8D presents an atomic model for the complex with IN protein shown as a Van der Waal's surface.

The proposed complex provides a relatively simple means of satisfying the biochemical and structural constraints, but more complicated arrangements cannot be ruled out. For example, our studies do not exclude higher-order IN complexes such as the octamer proposed by Heuer and Brown (1998). In the model in Figure 8, some of the proposed interactions do not yield a precise local fit between the DNAs and IN, suggesting that the protein and DNA may adjust their positions in these regions upon complex formation.

Studies using disulfide-mediated cross-linking also indicated that the C-terminal domain has different positions on DNAs mimicking different stages of the integration reaction. Several previous studies have indicated that the different steps of integration are probably accompanied by conformational changes (Ellison and Brown, 1994; Asante-Appiah *et al*., 1998; Espeseth *et al*., 2000; Hazuda *et al*., 2000). Following terminal cleavage, complexes are more stable than those present at earlier reaction steps, probably helping to promote correct assembly and suppress reversal of the reaction. S230C cross-linked most readily to A7 on the uncleaved substrate, but most readily to A9 after cleavage, specifying one part of the complex that participates in the conformational change. In the model shown in Figure 8, this could be accommodated by a number of possible motions, such as flexing of the α-helix connecting the catalytic and C-terminal domains, rotation of the dimer as a unit or displacement of the cDNA ends.

In summary, analysis of IN–DNA cross-linking by disulfide formation supports the following conclusions. (i) The IN residue E246 binds near position 7 in the viral cDNA. (ii) The C-terminal domain from one IN monomer binds with the catalytic domain of another monomer at each cDNA end. This finding allowed modeling of an IN tetramer linked by *trans* interactions between IN and the cDNA. (iii) A conformational change accompanying terminal cleavage involves repositioning the C-terminal domain in the vicinity of residue S230. These findings may also aid efforts to crystallize IN bound to DNA. The disulfide cross-linking strategy has been used previously to prepare complexes of HIV-1 reverse transcriptase and its substrates, allowing crystallization and structure determination by X-ray crystallography (Huang *et al*., 1998, 2000). The disulfide-mediated cross-linking reactions reported here provide the methods for preparing homogenous complexes of IN bound to DNA for crystallization trials.

Materials and methods

Oligodeoxynucleotides

Modified nucleotides with thio-tether arms (Figure 1A) were introduced into oligodeoxynucleotides using post-synthetic modification (Ferentz and Verdine, 1991). Oligonucleotides containing the convertible nucleoside O^6-phenyl-2′-deoxyinosine (hereafter called Φ) were synthesized on an Expedite™ Nucleic Acid Synthesis System (PerSeptive Biosystem, Inc.) using phenoxyacetyl-protected phosphoramidites by incorporating O^6-phenyl-dI phosphoramidite (GlenResearch). After synthesis, the resin-bound synthesis product was treated with concentrated ammonia at room temperature for 6 h to release and deprotect oligonucleotides. The released oligonucleotides were desalted and aminolysized by treatment with bis(3-aminoethyl) disulfide (n = 2) or bis(3-aminopropyl) disulfide (n = 3) at 65°C for 16 h. Bis(2-aminoethyl) disulfide hydrochloride was purchased from Fluka, and converted to free bis(2-aminoethyl) disulfide by treatment with sodium hydroxide. Bis(3-aminopropyl) disulfide (Evans *et al*., 1990) was synthesized from 3-bromopropylamine hydrobromide through a Bunte salt (Westlake and Dougherty, 1942). This converted Φ to disulfide-protected N^6-thioethyl-dA containing two methylene units or N^6-thiopropyl-dA containing three methylene units (Figure 1A).

The oligonucleotides containing Φ were as follows: HU520B1Φ (5′-ΦCTGCTAGAGATTTTCCACA), HU520B7Φ (5′-ACTGCTΦGAGATTTTCCACA), HU520B9Φ (5′-ACTGCTAGΦGATTTTCCACA) and HU520B11Φ (5′-ACTGCTAGAGΦTTTTCCACA).

The following modified oligodeoxynucleotides containing a thioalkyl tether arm were prepared from HU520B1Φ, HU520B7Φ and HU520B9Φ: (A^{C2} denotes N^6-thioethyl-dA protected as a disulfide; A^{C3} denotes N^6-thiopropyl-dA protected as a disulfide): HU520B1C2 (5′-A^{C2}CTGCTAGAGATTTTCCACA), HU520B7C2 (5′-ACTGCTA^{2}GAGATTTTCCACA), HU520B9C2 (5′-ACTGCTAGA^{C2}GATTTTCCACA) and HU520B11C2 (5′-ACTGCTAGAGA^{C2}TTTTCCACA); HU520B1C3 (5′-A^{C3}CTGCTAGAGATTTTCCACA) and HU520B7C3 (5′-ACTGCTA^{C3}GAGATTTTCCACA).

The following oligodeoxynucleotides (unmodified) were purchased from Integrated DNA Technologies, Inc.: HU520B (5′-ACTGCTAGAGATTTTCCACA), HU520T (5′-TGTGGAAAATCTCTAGCAGT), HU518T (5′-TGTGGAAAATCTCTAGCA), DIS1 (5′-ACTGCTAGAGATTTTCCACA), DIS2 (5′-TGTGGAAAATCTCTAGCAGGGGCTATGGCGTCC), DIS3 (5′-GAAAGCGACCGCGCC), DIS4 (5′-GGACGCCATAGCCCCGGCGCGGTCGCTTTC) and HPL (5′-CCATCGTTCGATGGTGTGGAAAATCTCTAGCACGCTCCGGTTCCGGAGCGCCGAGGCCTTGGCCTCGG).

The modified oligonucleotides were 5′-end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP (Maniatis *et al*., 1982). HIV-1 viral cDNA end duplex mimics [S(Bl)] were formed by annealing each of the above modified oligonucleotides with HU520T. Post-processing substrates [S(Pc)], in which two nucleotides had been removed, were formed by annealing the modified oligonucleotides with HU518T. Disintegration substrates [S(Dis)] were formed by annealing with DIS2, DIS3 and DIS4 and the appropriate modified oligonucleotide.

The hairpin three-way junction substrate was prepared by the ligation of HU520B7C3 and HPL using T4 ligase. The ligation products were gel purified to obtain S(Db)-7C3(SS): 5′-ACTGCTA^{C3}GAGATTTTCCACACCATCGTTCGATGGTGTGGAAAATCTCTAGCACGCTCCGGTTCCGGAGCGCCGAGGCCTTGGCCTCGG. Before it was used as substrate for the *cis–trans* test, S(Db)-7C3(SS) was 5′-^{32}P-end-labeled and annealed to form the hairpin structure.

The unmodified HIV-1 viral cDNA end duplex mimic S(Bl) was formed by annealing the mixture of 5′-^{32}P-labeled HU520T and HU520B in the same molarity. Disintegration substrate S(Dis) was formed by annealing the mixture of DIS1, DIS2, 5′-^{32}P-labeled DIS3 and DIS4 in the same molarity.

IN mutants

A His_6 tag followed by a thrombin cleavage site was added at the N-terminus of IN to facilitate purification. All mutations were introduced into a synthetic full-length HIV-1 IN sequence (Gerton *et al*., 1998) by cleavage of the integrase gene (containing the F185K substitution) and ligation with synthetic double-stranded oligonucleotides encoding the desired sequences. A T7 promoter was used for inducible expression of the IN protein. All mutations were confirmed by DNA sequencing. HIV-1 IN derivatives were purified using nickel-chelating agarose essentially as described in Miller *et al*. (1995).

Activity tests

Reaction mixtures contained 20 mM Tris–HCl pH 7.4, 80 mM NaCl, 10 mM $MnCl_2$, 2 mM βME, 1 mM CHAPS, 1 nM ^{32}P-labeled DNA substrate and 0.7 μM IN in a 20 μl volume. The reaction mixtures were incubated at 37°C for 30 min. A 2 μl aliquot of reaction solution was mixed with 40 μl of sequencing gel stop buffer (Maniatis *et al*., 1982). The sample was denatured by heating to 95°C for 5 min, chilled on ice, analyzed by 20% denaturing PAGE and visualized by autoradiography.

Cross-linking

Activated cross-linkable substrates were prepared freshly before being used. The sulfhydryl-tethered adenine IN substrates (SH) were obtained by reduction of disulfide-protected substrates (SS) with DTT. Disulfide-protected substrates (SS) were incubated with 0.1 M Tris–HCl pH 7.9, 75 mM DTT, 1 mM EDTA under argon at 55°C for 2 h, and cooled slowly to room temperature. The sulfhydryl-modified substrate was used after desalting by passage over a QuickSpin G-25 column (Boehringer Mannheim).

To prepare the TNB-thio-activated species (TNB), the sulfhydryl-tethered substrates (SH) prepared as described above were incubated with 1 mM DTNB (Sigma) (from fresh 10 mM solution of DTNB in 0.1 M phosphate buffer pH 8) at room temperature for 40 min. TNB-thio-activated substrates were desalted by QuickSpin G-25 columns and immediately used in cross-linking reactions.

The cross-linking reaction solutions contained 20 mM Tris–HCl pH 7.4, 80 mM NaCl, ~20 nM substrates, 1.4 μM proteins, 1 mM CHAPS and 2 mM βME (no βME in the case of TNB-thio-activated substrates) in a 10 μl volume. Cross-linking reactions were carried out at room temperature for 30 min. The reactions were stopped by adding 38 μl of stopping buffer, which contains 60 mM Tris–HCl pH 6.8, 2.5% SDS, 0.01% xylene cyanol, 60 mM EDTA, 500 mM NaCl and 5.2 M urea. After being denatured at 95°C for 4 min, the reaction products were electrophoresed on a 10% SDS–polyacrylamide gel without a stacking gel.

The maximal extent of cross-linking between E246C and the adenine 7-modified substrate was ~20% after SDS–PAGE, but this represents a minimum estimate due to loss of product due to heating the sample in the presence of βME prior to electrophoresis. This was necessary since samples entered the gel only poorly in the absence of βME. Efforts to analyze the extent of cross-linking by addition of methyl methanesulfonate, which reacts with free sulfhydryls, indicated at least 50% conversion to cross-linked product, though samples were difficult to analyze by electrophoreses due to aggregation. Almost all of the DNA appears to be cross-linked to protein as assayed by the experiment in Figure 6.

Cis–trans test

Substrate concentration was ~20 nM and the total IN concentration was 1.4 μM (equimolar mixture). The $MnCl_2$ (0.2 M) was added last to a final concentration of 10 mM. The cross-linking mixtures were incubated at room temperature for 30 min, then incubated at 37°C for 20 min after addition of Mn. The reactions were stopped and analyzed by 10% SDS–PAGE as described above. In exchange experiments, the second protein was added after incubation of substrate with the first proteins at room temperature for 30 min. Reaction solutions were then incubated at 37°C for an additional 20 min after addition of Mn and analyzed by SDS–PAGE.

Acknowledgements

We thank members of the Salk Institute Infectious Disease Laboratory and Dr Leslie Orgel for helpful discussions, Dr Patrick Brown for sharing his synthetic HIV-1 IN gene, and Lynn Artale for manuscript preparation. This work was supported by NIH grants GM56553 and AI34786 to F.D.B., the James B.Pendleton Charitable Trust, the Berger Foundation and Cornelia Mackey. F.D.B. is a Scholar and S.B. is a fellow of the Leukemia and Lymphoma Society of America.

References

Aldaz,H., Schuster,E. and Baker,T.A. (1996) The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell*, **85**, 257–269.

Asante-Appiah,E., Seeholzer,S.H. and Skalka,A.M. (1998) Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J. Biol. Chem.*, **273**, 35078–35087.

Bor,Y.-C., Bushman,F. and Orgel,L. (1995) *In vitro* integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl Acad. Sci. USA*, **92**, 10334–10338.

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.

Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science*, **249**, 1555–1558.

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.

Chen,J.C.-H., Krucinski,J., Miercke,L.J.W., Finer-Moore,J.S., Tang, A.H., Leavitt,A.D. and Stroud,R.M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl Acad. Sci. USA*, **97**, 8233–8238.

Chen,Z. *et al*. (2000) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293)—an initial glance of the viral DNA binding platform. *J. Mol. Biol.*, **296**, 521–533.

Coffin,J.M., Hughes,S.H. and Varmus,H.E. (1997) *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Craigie,R., Fujiwara,T. and Bushman,F. (1990) The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell*, **62**, 829–837.

Davies,D.R., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2000) Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. *Science*, **289**, 77–85.

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.

Eijkelenboom,A.P.A.M., van den Ent,F.M.I., Vos,A., Doreleijers,J.F., Hard,K., Tullius,T., Plasterk,R.H.A., Kaptein,R. and Boelens,R. (1997) The solution structure of the amino-terminal HHCC domain of HIV-2 integrase; a three-helix bundle stabilized by zinc. *Curr. Biol.*, **7**, 739–746.

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.

Engelman,A. and Craigie,R. (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J. Virol.*, **69**, 5908–5911.

Engelman,A., Mizuuchi,K. and Craigie,R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211–1221.

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.

Espeseth,A.S. *et al*. (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl Acad. Sci. USA*, **97**, 11244–11249.

Esposito,D. and Craigie,R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein–DNA interaction. *EMBO J.*, **17**, 5832–5843.

Evans,B.J., Doi,J.T. and Musker,W.K. (1990) Kinetics of the aqueous periodate oxidation of aliphatic disulfides and thioethers. *J. Org. Chem.*, **55**, 2580–2586.

Farnet,C., Lipford,R., Wang,B. and Bushman,F.D. (1996) Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl Acad. Sci. USA*, **93**, 9742–9747.

Ferentz,A.E. and Verdine,G.L. (1991) Disulfide cross-linked oligonucleotides. *J. Am. Chem. Soc.*, **113**, 4000–4002.

Fitzgerald,M.L., Vora,A.C., Zeh,W.G. and Grandgenett,D.P. (1992) Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *J. Virol.*, **66**, 6257–6263.

Gerton,J.L., Ohgi,S., Olsen,M., Derisi,J. and Brown,P.O. (1998) Effects of mutations in residues near the active site of human immuno-

deficiency virus type 1 integrase on specific enzyme–substrate interactions. *J. Virol.*, **72**, 5046–5055.

Goldgur,Y., Dyda,F., Hickman,A.B., Jenkins,T.M., Craigie,R. and Davies,D.R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.

Greenwald,J., Le,V., Butler,S.L., Bushman,F.D. and Choe,S. (1999) Mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*, **38**, 8892–8898.

Hansen,M.S.T., Carteau,S., Hoffmann,C., Li,L. and Bushman,F. (1998) Retroviral cDNA integration: mechanism, applications and inhibition. In Setlow,J.K. (ed.), *Genetic Engineering. Principles and Methods*, Vol. 20. Plenum Press, New York and London, pp. 41–62.

Hazuda,D. *et al*. (1999) Isolation and characterization of a novel class of human immunodeficiency virus integrase inhibitors from fungal metabolites. *Antimicrob. Agents Chemother.*, **10**, 63–70.

Hazuda,D.J. *et al*. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*, **287**, 646–650.

Heuer,T.S. and Brown,P.O. (1997) Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme–DNA complex by photo-cross-linking. *Biochemistry*, **36**, 10655–10665.

Heuer,T.S. and Brown,P.O. (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type-1 integrase–DNA complex. *Biochemistry*, **37**, 6667–6678.

Huang,H., Chopra,R., Verdine,G.L. and Harrison,S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669–1675.

Huang,H., Harrison,S.C. and Verdine,G.L. (2000) Trapping of a catalytic HIV reverse transcriptase template:primer complex through a disulfide bond. *Chem. Biol.*, **7**, 355–364.

Jenkins,T.M., Esposito,D., Engelman,A. and Craigie,R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.*, **16**, 6849–6859.

Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell*, **63**, 87–95.

Katzman,M., Katz,R.A., Skalka,A.M. and Leis,J. (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration. *J. Virol.*, **63**, 5319–5327.

LaFemina,R.L., Callahan,P.L. and Cordingly,M.G. (1991) Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.*, **65**, 5624–5630.

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.

Lutzke,R.A. and Plasterk,R.H. (1998) Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.*, **72**, 4841–4848.

Maignan,S., Guilloteau,J.P., Zhou-Liu,Q., Clement-Mella,C. and Mikol,V. (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J. Mol. Biol.*, **282**, 359–368.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Miller,M., Bor,Y.-C. and Bushman,F.D. (1995) Target DNA capture by HIV-1 integration complexes. *Curr. Biol.*, **5**, 1047–1056.

Muller,H.-P. and Varmus,H.E. (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.*, **13**, 4704–4714.

Naumann,T.A. and Reznikoff,W.S. (2000) Trans catalysis in Tn*5* transposition. *Proc. Natl Acad. Sci. USA*, **97**, 8944–8949.

Pommier,Y., Pilon,A.A., Bajaj,K., Mazumder,A. and Neamati,N. (1997) HIV-1 integrase as a target for antiviral drugs. *Antiviral Chem. Chemother.*, **8**, 463–483.

Pryciak,P.M. and Varmus,H.E. (1992) Nucleosomes, DNA-binding proteins and DNA sequence modulate retroviral integration target site selection. *Cell*, **69**, 769–780.

Savilahti,H. and Mizuuchi,K. (1996) Mu transpositional recombination: donor DNA cleavage and strand transfer *in trans* by the Mu transposase. *Cell*, **85**, 271–280.

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.

van Gent,D.C., Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) Complementation between HIV integrase proteins mutated in different domains. *EMBO J.*, **12**, 3261–3267.

Westlake,H.E.J. and Dougherty,G. (1942) The use of bunte salts in synthesis. III. The preparation of aliphatic disulfides. *J. Am. Chem. Soc.*, **64**, 149–150.

Wolfe,S.A., Ferentz,A.E., Grantcharova,V., Churchill,M.E. and Verdine,G.L. (1995) Modifying the helical structure of DNA by design: recruitment of an architecture-specific protein to an enforced DNA bend. *Chem. Biol.*, **2**, 213–221.

Yang,Z.N., Mueser,T.C., Bushman,F.D. and Hyde,C.C. (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.*, **296**, 535–548.

Yoder,K.E. and Bushman,F.D. (2000) Repair of gaps in retroviral DNA integration intermediates. *J. Virol.*, **74**, 11191–11200.

Received January 15, 2001; revised April 18, 2001; accepted May 1, 2001