Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration

Zhiqi Yin, Mikalai Lapkouski, Wei Yang, and Robert Craigie*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received 7 August 2012; Accepted 19 September 2012

DOI: 10.1002/pro.2166

Published online 25 September 2012 proteinscience.org

Abstract: Integrase is the key enzyme that mediates integration of retroviral DNA into cellular DNA which is essential for viral replication. Inhibitors of HIV-1 that target integrase recognize the nucleoprotein complexes formed by integrase and viral DNA substrate (intasomes) rather than the free enzyme. Atomic resolution structures of HIV-1 intasomes are therefore required to understand the mechanisms of inhibition and drug resistance. To date, prototype foamy virus (PFV) is the only retrovirus for which such structures have been determined. We show that PFV strand transfer complexes (STC) can be assembled on product DNA without going through the normal forward reaction pathway. The finding that a retroviral STC can be assembled in this way may provide a powerful tool to alleviate the obstacles that impede structural studies of nucleoprotein intermediates in HIV-1 DNA integration.

Keywords: integrase; prototype foamy virus; human immunodeficiency virus type 1

Introduction

Integration of viral DNA into cellular DNA is an essential step in the replication of HIV-1 and other retroviruses.^{1,2} Retroviral DNA made by reverse transcription after infection remains tightly associated with viral and cellular proteins in a large nucleoprotein complex called the preintegration complex (PIC).³ The PIC is a poorly defined complex because the low abundance of PICs in extracts of infected cells precludes direct biophysical studies. Purified HIV-1 integrase and DNA substrate mimicking the ends of the viral DNA can associate *in vitro* to form a stable synaptic complex (SSC).⁴ The viral DNA ends within the SSC are bridged by a

tetramer of integrase that protects about 16 bp of terminal sequence in footprinting experiments.⁴ The integrase tetramer and associated viral DNA are also referred to as the intasome. Although only about 16 bp of terminal DNA sequence is protected in the intasome in the case of HIV-1, several hundred base pairs of non-specific flanking DNA are required for efficient intasome assembly; the role of this non-specific internal DNA is not clearly understood. This stable association of integrase with viral DNA ends reproduces the functionality of integrase with viral DNA in the PIC. *In vitro* assembled SSCs efficiently integrate into a target DNA with all the hallmarks of integration *in vivo*. After forming the intasome, integrase cleaves two nucleotides from each 3' end of the viral DNA. The complex then captures a target DNA and a pair of transesterification reactions covalently joins viral to target DNA forming the strand transfer complex (STC) [Fig. 1(A)]. Integration target sites exhibit very limited DNA specificity, so STCs made through the normal reaction pathway are typically heterogeneous with respect to the sites of viral to target DNA joining.

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Intramural Program of NIDDK, NIH, AIDS Targeted Antiviral Program of the Office of the Director of the NIH.

*Correspondence to: Robert Craigie, Building 5, Room 301, 5 Memorial Drive, Bethesda, MD 20892-05670.
E-mail: bobc@helix.nih.gov

Figure 1. Strategy for assembly of the STC from “half-DNA intermediate” DNA substrate. A: In the normal DNA integration reaction, a tetramer of integrase synapses a pair of viral DNA ends to form the SSC or intasome. 3' end processing occurs within the intasome. The SSC captures a target DNA and the 3'OH ends of the viral DNA attack a pair of phosphodiester bonds in the target DNA to covalently join the 3' ends of the viral DNA to target DNA; in the case of PFV the sites of joining are separated by four base pairs on the two target DNA strands. A tetramer of integrase (shaded oval) remains associated with product DNA in the resulting STC. B: The “half-DNA intermediate” consists of a pre-processed viral DNA end segment with the 3' end covalently joined to one strand of a “half-target DNA” segment. Correct juxtaposition of two of these DNAs by integrase would result in the DNA structure of the STC made in the normal DNA integration reaction.

As with DNA transposons, the integrase–DNA complexes progressing along the reaction pathway become increasingly stable, and the STC is the most stable. It must be disassembled before cellular enzymes can complete the integration process. The mechanism of disassembly remains to be determined. In the case of bacterial Mu transposition, an energy-dependent process is required for cellular enzymes to disassemble the STC⁵ before transposition is completed and other transposons and retroviruses may require similar active disassembly of the product nucleoprotein complex. Completion of integration requires dissociation of the STC, removal of the two unpaired bases at each 5' end of the viral DNA, filling in of the single strand gaps, and ligation. Cellular enzymes carry out these latter steps. The increased stability of STC gives the integration/transposition processes a non-reversible character. To date, there is no report of assembling homogeneous retroviral STC without actual chemical reactions.

The current FDA approved integrase inhibitor raltegravir,⁶ and others in clinical trials, recognizes the intasome rather than free integrase protein for which they have only low affinity. Without atomic resolution structures of HIV-1 intasomes, the molecular mechanism of its inhibition and how mutations confer resistance can only be inferred. The prototype foamy virus (PFV) integrase produces the only retroviral intasome (SSC) and STC for which structures have been determined.^{7,8} The PFV intasome is therefore the best model for studying the action of integrase inhibitors, but differences between the HIV-1 and PFV integrases do not allow reliable modeling

except in the immediate vicinity of the active site. Structures of the HIV-1 intasome are therefore required.

The HIV-1 intasome has so far proved resistant to high-resolution structural studies. As mentioned before, HIV-1 requires viral DNA of several hundred base pairs long to form intasomes and the intasomes are unstable when this flanking DNA is removed. HIV-1 STCs are stable even when the flanking DNA is removed by restriction enzyme digestion (unpublished data), but the heterogeneous sites of integration into the target DNA make these complexes unsuitable for crystallization. In contrast, PFV has several distinct advantages over the HIV-1 enzyme for structural studies. It is more soluble and, most importantly, forms intasomes with short oligonucleotide substrates that mimic the ends of PFV DNA. As a tool for structural studies, we therefore explored the feasibility of assembling PFV STCs by mixing integrase and product DNA without going through the reaction pathway [Fig. 1(B)] as a model system to alleviate some of the problems that impede progress of structural studies of nucleoprotein intermediates in HIV-1 DNA integration.

Results

Complexes of integrase and “half integration intermediate” DNA detected by EMSA

As the DNA substrate for complex assembly, we used the “half integration intermediate” DNA shown in Figure 2(A). If integrase can correctly juxtapose a pair of “half integration intermediate” DNAs, the

Figure 2. Nucleoprotein complex formed by integrase with “half-DNA intermediate” DNA substrate. A: The “half-DNA intermediate.” B: The half-DNA intermediate was incubated with integrase in buffer containing 150 mM NaCl and electrophoresed in a polyacrylamide gel either directly or after adjusting the mixture to 400 mM NaCl. The complex was formed with similar efficiency in the absence of divalent metal ion or in the presence of Mg²⁺ or Mn²⁺. C: Size exclusion chromatography of the complex on a Superdex 200 column (GE Healthcare) in the presence of 500 mM NaCl. D: Crosslinking of free integrase in solution (left panel) or integrase within the complex purified by size exclusion chromatography (right panel). Free integrase is mostly crosslinked monomer and dimer whereas the integrase in the complex predominantly crosslinks as a tetramer. The asterisk represents the fluorophore 5' 6-FAM which was used to visualize the DNA.

product would be an STC. After mixing DNA and integrase under optimal conditions for concerted integration (125–150 mM NaCl in the presence of divalent metal ion), only a trace of complex was detected by electrophoretic mobility shift assay (EMSA) when the mixture was directly applied to a 6% polyacrylamide gel electrophoresis (PAGE) gel [Fig. 2(B)]. However a band corresponding to the protein–DNA complexes was readily observed when the ionic strength of the assembly mixture was raised to 400 mM NaCl prior to loading the gel [Fig. 2(B)]. When loaded at low ionic strength, the complex aggregated and failed to enter the gel. Assembly is independent of the presence of divalent metal ion or ethylenediaminetetraacetic acid (EDTA) [Fig. 2(B)]. The insensitivity to EDTA is at first sight surprising because the N-terminal domain contains a bound zinc ion. We speculate that zinc remains bound even in the presence of EDTA.

The complexes migrate as a monodisperse peak in gel filtration at 0.5M NaCl and the elution time is consistent with a tetramer of integrase associated with a pair of viral DNA ends [Fig. 2(C)]. This conclu-

sion is supported by crosslinking of integrase within complexes purified by gel filtration, which gives a major band corresponding to a tetramer [Fig. 2(D)].

DNA requirements for complex assembly

We first examined the sequence corresponding to the target site duplication that can potentially be base-paired in the STC [Fig. 3(panels A,B)]. The optimum length was four base pairs, which corresponds to the length of the PFV target site duplication.⁹ Reducing the duplication site to two base pairs drastically reduced the efficiency of assembly [Fig. 3(B), lane 12]. However, target sequences that cannot base pair [e.g., Fig. 3(B), lanes 4 and 6] formed the complex with only a slightly reduced efficiency. The efficiency of complex formation is relatively insensitive to the length of viral and target DNA segments [Fig. 3(panels C,D)], with 16 bp of viral DNA and 15 bp of target DNA being sufficient for near optimal assembly efficiency. The preference for four base pairs of sequence corresponding to the target duplication is consistent with the complex being an STC.

Figure 3. DNA substrate requirements for complex assembly. A: The half-DNA intermediate showing the viral DNA length, target DNA length, and duplication site. B: Complexes were assembled with half-DNA intermediate containing the indicated sequences at the duplication site. C: Complexes were assembled with half-DNA intermediate containing the indicated target DNA sequence lengths and the four base pair complementary target site. D: Complexes were assembled with half-DNA intermediate containing the indicated viral DNA sequence lengths and the four base pair complementary target site.

Integrase correctly juxtaposes the two half-site integration intermediates

Native gel electrophoresis shows that the efficiency of complex assembly is similar in the presence of EDTA, Mg^{2+} , or Mn^{2+} [Fig. 2(B)]. However when the complexes are deproteinized prior to electrophoresis a new band corresponding to full-length target

DNA is observed in the presence of Mn^{2+} , but not in the presence of Mg^{2+} (Fig. 4), or EDTA (data not shown). This band is the product of the reversal of the normal integration reaction, termed disintegration.¹⁰⁻¹² This reaction does not occur in the presence of the normal metal ion co-factor Mg^{2+} ; if it occurred in the cell, integration would reverse before the STC is disassembled and the integration intermediate repaired by cellular enzymes. Reversal of integration in the presence of Mn^{2+} demonstrates that the 3'-OH of the cleaved target DNA is positioned within the complex for nucleophilic attack on the phosphodiester bond linking viral and target DNA, as would be expected if the complex is an authentic STC.

Structure of the complex reveals it is an authentic STC

We decided to crystallize the complex to unambiguously determine its structure. For crystallization

Figure 4. The complex positions the 3' ends of the target DNA for nucleophilic attack on the phosphodiester bonds between viral and target DNA. A: Complex was assembled with half-DNA intermediate substrate and integrase and deproteinized after incubation in the presence of Mg^{2+} or Mn^{2+} . B: Analysis of the products by gel electrophoresis. In the presence of Mg^{2+} no chemistry takes place, whereas in the presence of Mn^{2+} a reversal of the normal integration reaction (disintegration) reveals the target DNA.

Table I. *Data Collection and Refinement Statistics*

Data collection	
Space group	$P4_12_12$
Cell dimensions	
a, b, c (Å)	160.61, 160.61, 125.09
Resolution (Å)	30–3.26 (3.32–3.26)
No. of reflections	25,740 (1277)
Multiplicity	6.3 (6.3)
Completeness (%)	98.8 (100)
Mean $I/\sigma I$	14.6 (2.0)
R_{sym}	0.109 (0.62)
Refinement	
Resolution (Å)	30–3.26
No. of reflections	24,101
$R_{\text{work}}/R_{\text{free}}$	18.5/22.6
No. of atoms	
Protein/DNA	4378/1305
Zn/Mg	1/2
B -factors	
Protein/DNA	55/100
Zn/Mg	56/40
R.m.s. deviations	
Bond distances (Å)	0.009
Bond angles (°)	1.3

Highest resolution shell is shown in parenthesis.

X-ray diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-BM at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. Data were collected at 100 K and $\lambda = 1$ Å with the crystal-to-detector distance was set to 250 mm and oscillation range per frame was 0.5°. One hundred fifty-five frames were processed using HKL2000.¹⁷ The crystal structure of the prototype foamy virus (PFV) intasome (PDB ID 3L2R) was used for rigid body refinement against the collected data. The resulting model was then refined with Phenix,¹⁹ including individual ADP and combination of rigid body and individual coordinates refinement, and manually corrected using Coot.¹⁸ The data collection and refinement statistics are shown in Table I. The refined structure includes PFV IN residues 8–375 in chain A and 116–299 in chain B. The complete non-transferred strand of the vDNA is resolved (19 bp in chain C). Density for the continuous strand of the vDNA integration product is traceable up to 33 bp (out of 38, chain D). The model has 95% of the amino acid residues in most preferred and 0.2% in disallowed regions.

trials, complexes were assembled with viral DNA (V) ranging from 16 bp to 36 bp and “half-target” DNA (T, not counting the central 4 nts) ranging from 13 bp to 30 bp. The protein–DNA complexes were purified by size-exclusion chromatography and concentrated to 4–6 mg/mL before crystallization screening. Only one crystal form, obtained with the DNA combinations V19T14, V19T15, V19T16, V19T17, and V19T18, diffracted beyond 4 Å. Although the sample preparation and crystallization condition are different, our crystals belong to the space group $P4_12_12$ and are isomorphous to those of the PFV STC complexes made through the normal forward integration reaction.⁸ The main difference between the two complexes is the length of target DNA, ours is 7–11 bp longer than the previously reported.⁸ The viral DNA

ends, which form crystal lattice contacts, are the same.

Diffraction data were collected to 3.26 Å on BM-22 at the APS synchrotron radiation source. The structure was solved by molecular replacement using the PFV SSC structure (PDB ID 3L2R) as the search model (Table I), and the protein structures are virtually identical [Fig. 5(A,B) Supporting Information Fig. 1(A)]. The coordinates and structure factors have been deposited in the Protein Data Bank with the accession code 4BAC. The integrase (IN) tetramer is a dimer of dimers, and each dimer contains an inner and outer subunit. The inner subunits are completely traceable and each is composed of the NTD, the core, the CTD domain, and the N-terminal extension (NED) unique to PFV [Fig. 5(C)]. The N-terminal and core domain are located at the ends of the elongated structure and the C-terminal domain is in the middle. The two inner subunits of the intasome are packed head-to-tail with one core domain stabilizing the N-terminal domain of the other subunit [Fig. 5(C)]. Only the core domain of each outer subunit is traceable and forms the conserved dimer interface with the inner subunit, but the N-terminal and C-terminal domains are disordered as reported.^{7,8} The viral and target DNAs in the STC are like four arms (two viral and two target) arranged in a tetrahedral configuration and converge at the integration sites [Fig. 6(A)]. The last three nucleotides on the 5' end of the viral DNA are flipped out and sandwiched between the core and CTD domains of the inner subunit. All the observed DNA contacts are with the inner two subunits [Fig. 5(C)]. P214, which stacks with the viral DNA 3' base, and the following α -helix with the catalytic carboxylate E221 have very different structures in the outer subunits, which are not involved in DNA binding [Fig. 5(D)]. In the core domain structure alone (PDB: 3DLR, Ref. 13), P214 and the surrounding loop is disordered and the helix with E221 is more similar to that in the outer subunit.

As reported previously, the target DNA is severely bent and unwound at the 4 bp staggered integration site, which would be duplicated upon completion of integration. The central AT dinucleotides are only partially paired and separated by over 5 Å due to the unwinding and severe bending [Fig. 6(B)]. The distortion supports the observation that mismatches at the central 4 bps are well tolerated in the intasome assemblies (Fig. 3). The longer target DNA in our structure allows definite tracing of an additional 5 bps and reveals the trajectory of the rest of the extension (Fig. 6). The longest half target that can be accommodated in the crystal lattice is 18 bps, where the end of target DNA reaches a neighboring molecule [Fig. 6(C)]. The electron density near the target end has an ill-defined shape. Additional unaccounted densities are observed between

Figure 5. Structure of the PFV integrase-product complex. A, B: Orthogonal views of the tetrameric PFV integrase in complex with two viral DNA ends inserted into a target DNA. The two inner subunits are shown in orange and cyan, and their dimeric partners (outer subunits) are shown in yellow and green, respectively. The two DNA molecules are shown in magenta and blue, each consisting of one viral DNA end (vDNA) and a half of the target DNA (tDNA). C: Ribbon diagrams of the two inner PFV integrase subunits in the same orientation as in (A). There is a large hole in the middle of the PFV tetramer. D: Structural differences between the inner and outer subunits. A close-up view of the viral and target DNA junction. The 3'-end base of the target DNA is shown in sticks. P214 of the inner subunit stack with the base. Residues from 214 to 242 (marked by the blue arrowheads) are significantly different in inner (orange) and outer (cyan) subunits while the rest of the core domains are superimposable (shown in pale yellow and light cyan). One of the three active site carboxylates, E221, is located in the middle of the helix that differs in the two subunits.

the extended target DNAs, which may correspond to the disordered NTD and CTD of the outer subunit. However, due to the missing link, these densities can be connected to more than one IN tetramer. It is interesting to note that in the structure of the phage Mu STC, domain IIIa of the Mu transposase occupies a similar position between the arms of the target DNA (S.P. Montano and P.A. Rice, personal communication). Regardless, the outer subunits have to be different from the inner subunits in order to fit into the crystal [Supporting Information Fig. 1(B)], and their NTD and CTD likely contribute to target DNA binding *in vivo* and aggregation in the intasome assemblies *in vitro*.

Discussion

We demonstrate that PFV integrase can assemble the STC directly on a DNA substrate that mimics the integration intermediate without going through the normal forward reaction pathway. The crystal structure of this complex reveals that it is indistinguishable from the STC made by the forward reaction. Assembly of the STC without going through

the forward reaction pathway also enabled us to unambiguously determine that the reverse “disintegration” does not occur in the presence of the physiologically relevant Mg^{2+} ion, although it occurs quite efficiently in the presence of Mn^{2+} . The population of STCs appears to be quite homogeneous. This contrasts with assembly of Mu transposase on DNA substrate mimicking the transposition product, where multiple species were inferred.¹⁴

The structure of the PFV intasome and STC remains the only high-resolution structures of a retroviral integrase in complex with DNA substrate. Although the structures of the analogous HIV-1 complexes are likely to be similar and can be modeled on the PFV structure,¹⁵ especially in the immediate vicinity of the active site, the structure of the HIV-1 complexes are required to fully understand the mechanism of integrase inhibitors and the mutations that confer drug resistance. The role of the domains that are disordered in the PFV structures may also be elucidated when structures of other retroviral integrases are obtained. The finding that the PFV STC can be assembled on product DNA provides a

Figure 6. Structure of the DNA products. A: DNA is shown as tube-and-ladders and colored in blue and magenta as in Figure 5(A). The last three nucleotides at the viral 5' end are unpaired and flipped out. The top image is in the same orientation as in Figure 5(A), and the bottom is an orthogonal view (from the angle indicated by the light brown arrowhead). For comparison, the shorter tDNAs reported by Maertens *et al.*⁸ are shown in silver and gold in the bottom panel. B: A close up view of the central 4 bp of the target DNA, which would be duplicated after repair and completion of integration. The palindromic GTAC are hydrogen bonded between base pairs as indicated by yellow dashed lines, but are highly distorted. The central dinucleotide TA are bent and unstacked (separated by over 5 Å instead of the usual 3.4 Å) and unwound. These four base pairs fit into the central hole as seen in Figure 5(C) and have little contact with the integrase except for limited van der Waals interactions with P132 and S332 through phosphosugar backbone. The contacting protein sidechains are shown as sticks and colored in cyan and orange according to the subunits, which they belonged to. C: The placement of the extended target DNA and unaccounted electron density in the structure. The STC complex is placed in the center. The inner subunits are visible and most of the outer subunits are clipped for clarity. The end of the target DNA (tDNA) at 18 bp makes lattice contact. The $2F_o - F_c$ map is calculated without the last 8 bp of the tDNA and contoured at 0.7 σ . The vDNAs lie outside the density box. The unaccounted densities between two tDNAs are indicated by the question marks.

new strategy toward this elusive goal. Unlike PFV, HIV-1 integrase does not assemble intasomes on short oligonucleotide substrate. Several hundred base pairs of non-specific flanking DNA are required for efficient assembly and stability.⁴ Furthermore, intasomes assembled with the long DNA fall apart when the flanking DNA is cut off after assembly.¹⁶ The strategy that the Cherapanov group successfully applied to PFV therefore cannot be used with the HIV-1 system. However, HIV-1 STCs made through the normal forward reaction pathway are stable when the flanking DNA is removed (data not shown). Unfortunately, these complexes are unsuitable for crystallization because the sites of integration into the target DNA are heterogeneous. The proof of concept that PFV STCs can be assembled on product DNA suggests that the same strategy may help overcome the obstacles to structural studies of

nucleoprotein intermediates in HIV-1 DNA integration.

Materials and Methods

PFV IN expression and purification

The sequence of PFV IN was obtained from the human spumaretrovirus complete genome (GenBank: U21247.1). Full-length PFV IN (IN 1-392, POL 752-1143) was synthesized by GenScript and cloned into pET-15b (Novagen). Plasmids were transformed into *Escherichia coli* BL21 (DE3). Cells were grown in LB at 37°C to OD₆₀₀ 0.8–1.0, induced by addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to 0.4 mM, grown for an additional 3 h at 37°C and harvested. The cell paste was suspended in lysis buffer (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 20 mM

imidazole, 0.5M NaCl, 2 mM 2-mercaptoethanol, 0.4 mg/mL lysozyme). Lysed cells were sonicated and then centrifuged at 35,000g for 45 min. The supernatant was filtered through a 0.2 μ m filter and applied to a HisTrap HP (GE healthcare) Ni-affinity column. After extensive washing with elution buffer (20 mM HEPES, pH 7.5, 0.5M NaCl, 2 mM 2-mercaptoethanol, 10% (wt/vol) glycerol) containing 20 mM and 60 mM imidazole, protein was eluted with a 10-column-volume linear gradient of 60 mM to 1M imidazole in elution buffer. The histidine tag was removed using thrombin (Sigma) at a concentration of 6 NIH U/mg of protein. Thrombin was removed by adsorption to benzamidine-Sepharose 6B (GE Healthcare). The protein was dialyzed against 20 mM HEPES, pH 7.5, 0.2M NaCl, 2 mM dithiothreitol (DTT), 1 mM EDTA, 10% glycerol, loaded onto a Mono S 10/10 column (GE Healthcare) and eluted with a linear gradient of 0.2–0.6M NaCl in 20 mM HEPES, pH 7.5, 2 mM DTT, 1 mM EDTA, 10% glycerol. Peak fractions were pooled and dialyzed overnight against 20 mM HEPES pH 7.5, 0.4M NaCl, 100 μ M ZnCl₂, 5 mM DTT, 10% glycerol. The purified protein was concentrated to 6–10 mg/mL using a Centrprep YM-10 (Millipore).

Complex assembly

DNA substrates were prepared by annealing three HPLC-purified synthetic oligonucleotides (IDT). For EMSA assays, 50 μ L assembly mixtures contained 1 μ M DNA and 2 μ M PFV IN in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM DTT, and 20% (vol/vol) dimethyl sulfoxide (DMSO) in the presence of 0.1 mM EDTA, or 5 mM MgCl₂, or 5 mM MnCl₂. After incubation at 37°C for 60 min, a 4 μ L aliquot was subjected to electrophoresis in Novex 6% TBE gel (Invitrogen). To prepare IN–DNA complexes for crystallization, 40 μ M DNA substrate and 100 μ M PFV IN were mixed in 20 mM HEPES, pH 7.5, 400 mM NaCl, 20% (wt/vol) glycerol, 5 mM DTT. The mixture was dialyzed against low salt buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 20% (wt/vol) glycerol, 5 mM DTT) at 37°C for 4 h. At the end of this first dialysis, precipitate appeared. Then the mixture was dialyzed against high salt buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 20% (wt/vol) glycerol, 5 mM DTT) at room temperature for 2 h. At the end of this second dialysis, the reaction mixture became clear. The IN–DNA complexes were purified by size-exclusion chromatography on a Superdex 200 10/300 column (GE Healthcare) equilibrated in 20 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM MgCl₂, 8% (wt/vol) glycerol, 5 mM DTT.

Chemical crosslinking

Chemical crosslinking of PFV IN or IN–DNA complexes was performed using ethylene glycol-bis

(EGS). EGS was freshly dissolved at 100 mM in DMSO. The reaction was initiated by the addition of 0.2–2.5 mM (final concentration) to IN alone or IN–DNA complex isolated through size-exclusion chromatography. Incubation was at room temperature for 15 min. Reactions were stopped by the addition of 100 mM Tris, and after incubation at room temperature for 5 min subjected to electrophoresis in SDS-PAGE.

Crystallization and structure determination

Complexes assembled from DNA substrates with different combinations of viral DNA lengths (from 16 bp to 36 bp) and half target DNA lengths (from 13 bp to 30 bp, corresponding to full target DNA from 30 bp to 64 bp), were concentrated to 4–6 mg/mL using a Vivaspin 2 centrifugal concentrator (Sartorius) and screened in crystallization trials. Several crystal forms were identified. Only one crystal form, obtained from the DNA constructs (V19T14, V19T15, V19T16, V19T17, and V19T18) diffracted beyond 4 Å. For diffraction data collection, crystals were grown through reverse vapor diffusion in hanging drops at 23°C by mixing 1 μ L IN–DNA complex solution (20 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM MgCl₂, 2 mM TCEP, and 8% (wt/vol) glycerol) with 1 μ L reservoir solution (50 mM Na cacodylate, pH 6.5, 100 mM MgCl₂, 1 mM CoCl₃, and 6% (vol/vol) ethanol). Crystals appeared within 1–2 days and reached a size of 200–300 μ m within 1–2 weeks. Crystals were cryoprotected in 18% (vol/vol) MPD, 40 mM MgCl₂, 100 mM NaCl, 2% (wt/vol) glycerol, 6% (vol/vol) ethanol, and 20 mM Na cacodylate, pH 6.5, and frozen by rapid immersion in liquid nitrogen. Diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-BM at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. Data were collected at 100 K and $\lambda = 1$ Å with the crystal-to-detector distance set to 250 mm and the oscillation range per frame was 0.5°. One hundred fifty-five frames were processed using HKL2000.¹⁷ The crystal structure of the PFV intasome (PDB ID 3L2R) without target DNA was used for rigid body refinement against the collected data. The map revealed clear density for the target DNA, which was manually built using COOT.¹⁸ The resulting model was then refined using Phenix,¹⁹ including individual ADP and combination of rigid body and individual coordinates refinement. The model was manually corrected in COOT.¹⁸ The data collection and refinement statistics are shown in Table I. The refined structure includes PFV IN residues 8–375 in chain A and 116–299 in chain B. The complete non-transferred strand of the viral DNA is resolved (19 bp in chain C and D). Density for the continuous strand of the DNA integration product is traceable up to 33 bp (out of 38, chain D and E). The model has 95% of

the amino acid residues in the most preferred and 0.2% in disallowed regions. The protein structure is nearly identical to the PFV intasome structure with C α atoms rmsd of 0.49 Å. A month later after the structure was solved, the coordinates for the PFV STC at 2.81 Å were deposited (PDB: 3OS0), which is almost identical to our structure [Fig. 5(A,B) and Supporting Information Fig. 1(A)]. The atomic coordinates and the structure factors have been deposited in the Protein Data Bank, with the accession code 4BAC.

References

1. Varmus H, Brown PO. Retroviruses. In: Berg DE, Howe MM, Eds. (1989) *Mobile DNA*. American Society of Microbiology, pp 53–108.
2. Brown PO, Integration. In: Coffin JM, Hughes SH, Varmus HE, Eds. (1997) *Retroviruses*. Cold Spring Harbor Laboratory Press, New York, pp 161–203.
3. Bowerman B, Brown PO, Bishop JM, Varmus HE (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* 3:469–478.
4. Li M, Mizuuchi M, Burke TR, Craigie R (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J* 25:1295–1304.
5. Levchenko I, Luo L, Baker TA (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* 9:2399–2408.
6. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, Fiore F, Gardelli C, Paz Odalys G, Hazuda DJ, Jones P, Kinzel O, Laufer R, Monteagudo E, Muraglia E, Nizi E, Orvieto F, Pace P, Pescatore G, Scarpelli R, Stillmock K, Witmer MV, Rowley M (2008) Discovery of Raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem* 51:5843–5855.
7. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464:232–236.
8. Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468:326–329.
9. Enssle J, Moebes A, Heinkelein M, Panhuysen M, Mauer B, Schweizer M, Neumann-Haefelin D, Rethwilm A (1999) An active foamy virus integrase is required for virus replication. *J Gen Virol* 80: 1445–1452.
10. Chow SA, Vincent KA, Ellison V, Brown PO (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 255:723–726.
11. Chow SA, Brown PO (1994) Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. *J Virol* 68:7869–7878.
12. Mazumder A, Engelman A, Craigie R, Fesen M, Pommier Y (1994) Intermolecular disintegration and intramolecular strand transfer activities of wild-type and mutant HIV-1 integrase. *Nucleic Acids Res* 22: 1037–1043.
13. Valkov E, Gupta SS, Hare S, Helander A, Roversi P, McClure M, Cherepanov P (2009) Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res* 37:243–255.
14. Mizuuchi M, Rice PA, Wardle SJ, Haniford DB, Mizuuchi K (2007) Control of transposase activity within a transpososome by the configuration of the flanking DNA segment of the transposon. *Proc Natl Acad Sci USA* 104:14622–14627.
15. Krishnan L, Li XA, Naraharisetty HL, Hare S, Cherepanov P, Engelman A (2010) Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc Natl Acad Sci USA* 107:15910–15915.
16. Li M, Ivanov V, Mizuuchi M, Mizuuchi K, Craigie R (2012) DNA requirements for assembly and stability of HIV-1 intasomes. *Protein Sci* 21:249–257.
17. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
18. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D* 60:2126–2132.
19. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* 66:213–221.