Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor

Fabrice Michel¹, Corinne Crucifix¹, Florence Granger¹, Sylvia Eiler¹, Jean-François Mouscadet², Sergei Korolev³, Julia Agapkina³, Rustam Ziganshin³, Marina Gottikh³, Alexis Nazabal⁴, Stéphane Emiliani^{5,6}, Richard Benarous⁷, Dino Moras¹, Patrick Schultz^{1,*} and Marc Ruff^{1,*}

¹IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), Département de Biologie et de Génétique Structurales, UDS, CNRS, INSERM, Illkirch, France, ²Laboratoire de Biotechnologie et Pharmacologie Génétique Appliquée, CNRS, ENS-Cachan, Institut d'Alembert, Cachan, France, ³Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia, ⁴Department of Chemistry and Applied Biosciences ETH Zurich and CovalX, Technoparkstrasse, Zürich, Switzerland, ⁵Institut Cochin, Département des Maladies Infectieuses, Université Paris Descartes, CNRS, Paris, France, ⁶Inserm, Paris, France and ⁷CellVir, Evry, France

Integration of the human immunodeficiency virus (HIV-1) cDNA into the human genome is catalysed by integrase. Several studies have shown the importance of the interaction of cellular cofactors with integrase for viral integration and infectivity. In this study, we produced a stable and functional complex between the wild-type full-length integrase (IN) and the cellular cofactor LEDGF/p75 that shows enhanced *in vitro* integration activity compared with the integrase alone. Mass spectrometry analysis and the fitting of known atomic structures in cryo negatively stain electron microscopy (EM) maps revealed that the functional unit comprises two asymmetric integrase dimers and two LEDGF/p75 molecules. In the presence of DNA, EM revealed the DNA-binding sites and indicated that, in each asymmetric dimer, one integrase molecule performs the catalytic reaction, whereas the other one positions the viral DNA in the active site of the opposite dimer. The positions of the target and viral DNAs for the 3' processing and integration reaction shed light on the integration mechanism, a process with wide implications for the understanding of viral-induced pathologies.

The EMBO Journal (2009) 28, 980–991. doi:10.1038/emboj.2009.41; Published online 19 February 2009

Subject Categories: microbiology & pathogens
structural biology

Keywords: cryo-electron microscopy; DNA integration; HIV-1; integrase; LEDGF–P75

Introduction

The human immunodeficiency virus (HIV-1) integrase (IN) catalyses the stable integration of the viral cDNA into the host genome, an essential step in the establishment of HIV infection. IN is a DNA recombinase that catalyses two endonucleolytic reactions. In the 3'-processing reaction, IN cleaves a dinucleotide from the U5 and U3 ends of viral cDNA, thereby exposing a 3'-OH group at each end of the viral DNA. In the strand-transfer reaction, IN generates a double-strand break in the host DNA and joins the newly formed ends to the viral 3' ends by transesterification (Engelman *et al*, 1991). Host DNA repair proteins remove the two nucleotide overhang and fill in the DNA gaps to complete the integration reaction (Yoder and Bushman, 2000). Both reactions can be carried out *in vitro* with purified IN and DNA substrates that mimic the viral DNA ends, in the presence of cofactors such as Mg²⁺ or Mn²⁺ (Zheng *et al*, 1996). The enzyme consists of three structural and functional domains. The N-terminal zinc-binding domain (residues 1–50), which is required for 3' processing and strand transfer *in vitro*, binds viral DNA sequences and promotes IN multimerisation (Engelman *et al*, 1993). The central catalytic core domain (CCD; residues 50–212) binds specifically to viral DNA. It contains the D, D-35E triad that coordinates divalent ions and is conserved in other retroviral integrases as well as in retrotransposon proteins such as the Tn5 transposase (Haren *et al*, 1999). The C-terminal domain (residues 213–288) of IN has an SH3-like fold (Eijkelenboom *et al*, 1999) that interacts with reverse transcriptase (Hehl *et al*, 2004) and binds DNA nonspecifically (Engelman *et al*, 1994). IN has been the focus of a large number of structural studies, and the design of a soluble IN mutant (F185K) (Jenkins *et al*, 1995) led to several structures of the CCD alone (Dyda *et al*, 1994; Maignan *et al*, 1998; Chiu and Davies, 2004), or with either the N- or the C-terminal domains (Chen *et al*, 2000; Wang *et al*, 2001). These structures initiated computer-aided modelling as well as negatively stained electron microscopy (EM) studies (Ren *et al*, 2007) that led to several models for the integration mechanism (Heuer and Brown, 1998; Gao *et al*, 2001; Karki *et al*, 2004; Wielens *et al*, 2005; Liao *et al*, 2007). *In vitro* integration assays showed that IN activity is increased in the presence of cellular proteins (Turlure *et al*, 2004). In particular, it has been shown that the transcriptional coactivator LEDGF/p75 interacts with HIV-1 IN (Cherepanov *et al*, 2003) as well as with other lentiviral integrases (Cherepanov, 2007) and that its expression is required for proviral integration and subsequent production of HIV-1 virions (Emiliani *et al*, 2005). Interaction occurs through the CCD of IN and the C-terminal integrase-binding domain (IBD) of LEDGF (Busschots *et al*, 2007). The function of LEDGF is to target IN to chromosomes of infected cells (Maertens *et al*, 2003). A structure of an IN(F185K) CCD–LEDGF IBD complex has been solved

*Corresponding authors. P Schultz or M Ruff, Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 rue Laurent Fries, Illkirch 67404, France.
Tel.: +33 0388 6557 50; Fax: +33 0388 6532 76 or
Tel.: +33 0388 6533 50; Fax: +33 0388 6532 76;
E-mails: patrick.schultz@igbmc.fr or marc.ruff@igbmc.fr

Received: 17 December 2008; accepted: 26 January 2009; published online: 19 February 2009

(Cherepanov *et al*, 2005). Unfortunately, in the absence of a structural model for the complete wild-type enzyme, none of the domain structures could give insights into the integration mechanism. To date, the poor solubility of the native full-length IN protein *in vitro* has hindered attempts to determine the organisation of the fully active unit. To overcome these problems, we purified a stable complex between the wild-type full-length HIV-1 IN and the human LEDGF proteins with a stoichiometry of 4 IN and 2 LEDGF as shown by mass spectrometry analysis. A cryo negative stain EM envelope was determined, and its volume is coherent with the stoichiometry of 4 IN and 2 LEDGF molecules. Known atomic structures were fitted and refined using normal mode flexible fitting. The structure in the presence of DNA permits to assess the positions of viral and cellular DNA. The proposed model is validated by the analysis of the DNA–protein interactions found by mutational and cross-linking studies found in the literature and carried out in this study. Taken together, the data reveal the quaternary organisation of the functional integration unit highlighting the role of each of its components and lead to a model for integration of the viral DNA into the host genome.

Results

Purification and characterisation of the IN/LEDGF complex

GST-IN and His₆-LEDGF were purified separately and solubilised in 1 M NaCl in the presence of 7 mM CHAPS. The complex, which is formed upon removal of the GST tag by specific proteases and of the solubilisation agents by dialysis, was purified to homogeneity by affinity chromatography and gel filtration (Supplementary Figure 1). The stoichiometry of the partners was determined by high-mass MALDI ToF mass spectrometry analysis (Nazabal *et al*, 2006). Control experiments identified the mass of the two components: IN ($MH^+ = 32.6$ kDa) and His₆-LEDGF ($MH^+ = 63.4$ kDa) (Figure 1A). In a second step, the purified complex was chemically cross-linked before mass spectrometry, which detected several protein complexes: [IN•LEDGF] ($MH^+ = 99.8$ kDa), [LEDGF•LEDGF] ($MH^+ = 132.6$ kDa,) [IN•2LEDGF] ($MH^+ = 165.8$ kDa) but the mass of the major species corresponded to [4IN•2LEDGF] ($MH^+ = 267.4$ kDa). No protein complex was detected in the higher molecular weight range between 500 and 1000 kDa.

Functional characterisation

To investigate the 3' processing and strand-transfer activities of the IN/LEDGF complex, we used 21- and 40-mer DNA duplexes that mimic the HIV-1 U5 viral DNA end (Figure 1B). In the presence of the 40-mer DNA, both the initial rate and the efficiency of the 3'-processing reaction are found to be higher for the IN/LEDGF complex compared with the recombinant IN alone, indicating that LEDGF stimulates the 3'-processing activity of IN. Interestingly, this increased activity is not observed when the 21-mer DNA is used, suggesting that a minimal DNA length is required to stimulate the 3'-processing reaction. Moreover, in the presence of LEDGF, the strand-transfer efficiency was strongly enhanced for both the 21- and the 40-mer DNA. Taken together, these results indicate that

IN increases its biological activities upon interaction with LEDGF.

Electron microscopy of the cryo negatively stained IN/LEDGF complex

The IN/LEDGF complex was first observed in transmission EM after negative staining, which showed that more than 85% of the particles were homogeneous in size. The preliminary analysis of an image data set of 3732 isolated particles revealed molecular views that showed a clear two-fold symmetry either in-plane for the side view or normal to the plane for the top view. This symmetry operator was also detected in the eigenvectors before any rotational alignment, indicating that the complex contains two identical units (Supplementary Figure 2). To improve the structural preservation of the sample, a cryo negative stain method was used to observe hydrated IN–LEDGF complexes with high contrast. The particles were sandwiched between two carbon films in the presence of uranyl acetate and quick frozen before complete dehydration (Golas *et al*, 2005). The analysis of 11 441 molecular images again showed a two-fold symmetry and a 3-D model was calculated from 350 different views (Supplementary Figure 3A). The re-projections of the 3-D model closely matched the original views (Figure 1C). The envelope shows a triangular shape when viewed from the side (14 by 15 nm) and a highly elongated shape (10 by 15 nm) when viewed down the two-fold axis (Figure 1D). The volume enclosed by the envelope is consistent with the stoichiometry of 4 IN and 2 LEDGF molecules (~260 kDa) determined by mass spectrometry and the resolution is of 14 Å according to the Rosenthal and Henderson criteria (Rosenthal and Henderson, 2003) (Supplementary Figure 4A). To determine the position of the IN subunit, a complex was assembled in which the IN carried a GST fusion at its N-terminus. The analysis of negatively stained particles revealed a large additional density located in the upper region of the 3-D model (Supplementary Figure 5). This experiment places IN in the upper domain of the model, a position in which the EM envelope revealed a Y-shaped clamp of 8.5 by 7 nm in size emphasised by red dots in Figure 2A. To fit the known atomic structures into the EM map, we used a composite model of the full-length IN in complex to the IBD of LEDGF (residues 347–442) obtained by the superposition of the catalytic domains, which are present in all structures (Figure 2A). This composite model could be readily placed into the Y-shaped clamp of the EM envelope. Normal modes flexible fitting in combination with rigid body refinement and structure idealisation were used to fit the integrase model into the EM map (Supplementary Figure 6, 7 and Movie 1). The fitting parameters were chosen in a way not to modify the fold of the protein domains (Supplementary Protocol 4). The C-terminal IN extensions fit into each of the branches of the 'Y', whereas the dimeric CCD is accurately positioned into the central part of the clamp. In contrast, the N-terminal domains, which fall outside of the envelope, were readily fitted into an empty region of the EM map. The final model of the complex shows no steric clashes between the different domains, and the top part of the envelope is fully occupied by the known atomic structures (Figure 2B). The remaining density forms a W-shaped structure at the bottom of the complex and is consistent with the size of an LEDGF dimer (Figure 2C).

Figure 1 Functional characterisation and structure determination of the IN/LEDGF complexes. (A) IN/LEDGF complex analysis by high-mass MALDI mass spectrometry in the absence and presence of cross-linking agent showing a major peak of (IN)₄-(LEDGF)₂. (B) Kinetic study of the 3' processing and strand-transfer reactions catalysed by HIV-1 integrase or by the IN-LEDGF complex using a 21- or 40-mer DNA substrate. Bands correspond to substrate DNA (S), 3'-processed product (P) and strand-transfer product (T). Incubation times are in min. The bands intensities are reported as a function of time, showing the increase of the 3' processing kinetic when using the IN-LEDGF complex with a 40-mer DNA substrate. (C) Views of the cryo negatively stained IN/LEDGF complex (upper row) and the corresponding reprojections of the 3-D model (lower row). (D) The 3-D model of the cryo negatively stained IN/LEDGF complex. (E) The 3-D model of the cryo negatively stained IN/LEDGF complex in the presence of the 21-bp DNA substrate.

Figure 2 Docking of atomic structures into the cryoEM maps of IN/LEDGF and IN/LEDGF/DNA. **(A)** Composite atomic model constructed from available X-ray structures. **(B)** Superposition of the fitted 3-D model and the IN/LEDGF EM envelope. The Y-shaped structure that accommodates the atomic structure of the catalytic and C-terminal IN domains is contoured by red dots. **(C)** Surface representation of the IN/LEDGF structure: the IN tetramer is in gold and the LEDGF dimer visualised by the difference map between the atomic model and the EM map is in grey. **(D)** Superposition of the IN/LEDGF (gold) and IN/LEDGF/DNA (blue) envelopes. The Y-shape in the top of the envelope is highlighted with red dots. The protruding densities corresponding to DNA are circled in orange. Densities (1) and (3) are assigned to viral DNA, whereas density (2) is assigned to target DNA. **(E)** Atomic model fitted into the IN/LEDGF/DNA envelope. **(F)** Difference map between the EM map and the fitted model. Fitted DNA is shown in blue. 1 and 3 represent the viral DNA, respectively, in the integration and 3'-processing step, and 2 represents the target DNA. **(G)** Atomic models of IN and DNA fitted into the EM map. DNA molecules assigned to viral DNA are shown in yellow and DNA assigned to target DNA is shown in red. **(H)** Crystal structure of the Tn5-DNA complex superposed to the IN-CCD identifying the viral DNA in the 3'-processing position. Two viral DNA positions in the integration intermediate and 3'-processing state are represented by red and green arrows, respectively. IN is indicated in gold, the viral DNA in orange, the target DNA in red and the envelope for the LEDGF dimer in grey.

Architecture of the functional IN/LEDGF/DNA complex

To determine the DNA-binding sites, the IN/LEDGF complex was incubated with a 10-fold molar excess of the 21-mer U5-substrate. A data set of 11 706 images of cryo negatively stained complexes was recorded and analysed independently of the previous model. The resulting 3-D envelope was slightly more elongated than the native complex (Figure 1E). The envelopes of the IN/LEDGF and the IN/LEDGF/DNA complexes were superimposed to reveal the structural changes induced by the addition of DNA

(Figure 2D). The upper domain appears larger, suggesting that additional mass is bound to the IN molecules. Comparison of the two maps revealed a displacement of the top branch of the Y domain (C-terminal domain of integrase), which results in a closure of the clamp upon DNA binding (Figure 2D). The crystal structures were adjusted and refined as described earlier (Figure 2E; Supplementary Figures 6, 8 and Movie 2). In each IN dimer, three-stain exclusion spikes are detected at the surface of the model (labelled 1, 2 and 3 in Figure 2D). These spikes are likely to reflect protruding DNA

molecules that exclude the stain as shown by (Schultz *et al*, 1996; Wagenknecht *et al*, 1988) (Supplementary Figure 3B). A difference map between the EM envelope and the fitted model reveals 20 Å wide rod-like structures that are consistent with the size and shape of DNA molecules. The extremity of these rods corresponds to the above-mentioned spikes (Figure 2F; Supplementary Movie 4). Rod 1 is connected to the catalytic domains of IN, indicating that this density corresponds to oligonucleotides mimicking the viral DNA. Rod 2 lies between the two arms of the clamp and forms a continuous density with the two-fold related rod 2' of the second clamp, thus, producing a bent filament connecting the two IN dimers. As the ends of this filament do not point towards the active sites, we interpreted this density as the target DNA. A DNA fragment from the nucleosome structure (Richmond and Davey, 2003) could be fitted into this additional volume, which is consistent with the observation that nucleosomes promote the HIV integration reaction (Pruss *et al*, 1994a). Moreover, the positions of the target DNA that are closest to the 3' ends of the two viral DNA are separated by a distance of 15 Å or five base pairs, which corresponds to the documented distance of viral DNA integration into the target DNA. The structure formed by the IN/LEDGF complex and these four DNA fragments (labelled 1,1',2,2' in Figure 2D) could, thus, correspond to an integration intermediate between the 3' processing and the integration step. Rod 3 and the two-fold related rod 3' is consistent with a straight B DNA and lies at the interface with the LEDGF moiety (Figure 2D and F). To gain some insight into the role of this unexpected DNA fragment, we superimposed the atomic structure of the IN catalytic site with that of the homologous Tn5 transposase in complex with DNA (Davies *et al*, 2000). The path of the Tn5 DNA fits properly into rod 3, thus, indicating that the corresponding DNA fragment is in the position of the viral DNA during the 3'-processing step. The excess of 21-mer DNA can explain the concomitant occupancy of the 3' processing and integration site of IN. Using this information, a structural model for the 3' processing intermediate could be constructed (Figure 2H).

Model validation by DNA-protein cross-linking

Our proposed model fully agrees with previously published biochemical and genetic data. To further validate our structural model and to identify the protein domains in contact with the viral DNA, cross-linking experiments were carried out to identify proximal DNA bases and protein amino acids. The found cross-linked peptide had the RKAKIIRDYGK sequence corresponding to amino acids 263–273 in the C-terminal domain. This result indicates that the sixth nucleotide from the nonprocessed strand end is located near Lys264 or Lys266. Our structures are in excellent agreement with these findings, as the lysine 264 is in the vicinity of the phosphate between nucleotides 6 and 7 of the unprocessed DNA strand (Supplementary Figure 9). This confirms the importance of the C-terminal domain for the mechanism of integration and viral DNA binding.

Discussion

Architecture of the IN-LEDGF assembly

Mass spectrometry and cryoEM showed that the complex formed between the HIV-1 integrase and its cellular cofactor

LEDGF contains 4 IN and 2 LEDGF molecules. This finding is consistent with biochemical data, indicating that an IN tetramer is the minimal complex required for the concerted DNA integration reaction (Li *et al*, 2006). The *in vitro* catalytic activity and stability of this assembly are superior to that of IN molecules alone, indicating that LEDGF acts as a chaperone by preventing IN from aggregation or higher multimerisation. The existing atomic structures could be readily fitted into the EM envelope, thus, positioning the functional domains of the enzyme. Two main differences were detected between our model and the atomic structures. First, the crystal structure of the IN CCD:LEDGF IBD complex contains two copies of each protein, whereas our structure shows a 2:1 stoichiometry where each LEDGF interacts with two IN molecules. Second, according to the available X-ray crystal and NMR structures, the IN N-terminal domains form dimers with different interfaces, suggesting that this interaction is not specific (Li *et al*, 2006). In our model, this domain does not dimerise but instead contacts the CTD of its own IN molecule (NtA-CtA, NtB-CtB) (Figure 3A). Very recently, the structure of the two domains HIV-2 integrase (CCD and NTD) in complex with the LEDGF IBD has been solved and is in agreement with our model (Hare *et al*, 2009). In this structure, the IN-NTD does not dimerise and is positioned as in the EM model. Moreover, a single LEDGF molecule is associated to each IN dimer. The contact between the residues E10 from IN and R405 from LEDGF shown in the crystallographic structure is in agreement with our model, as these residues are within reach in our complex (Supplementary Figure 10). The domains arrangement has important functional consequences as the two IN molecules form a dimer, within which they adopt an asymmetric organisation and fulfil distinct functions. Thus, the 4 IN molecules (A1, A2, B1 and B2) are organised into two asymmetric IN dimers (1 and 2) related by a two-fold axis. Such an asymmetry has been inferred previously from chemical cross-linking experiments, which showed that intra-dimer and inter-dimer contacts were not equivalent, thus, excluding the possibility that the IN tetramer exhibits a four-fold symmetry (Faure *et al*, 2005).

Architecture of the IN-LEDGF-DNA complex

The 3-D model of the functional IN/LEDGF/DNA assembly revealed the interaction sites and the paths of the host and viral DNA for the 3'-processing mode and for an integration intermediate. The global domain organisation of the integrase tetramer in complex with DNA is similar to the architecture of the protein assembly except for conformational changes of the N- and C-terminal parts of IN. The target DNA binds within the IN clamp. Upon DNA binding, the C-terminal domain of monomer B and the N-terminal domain of monomer A move to clamp the target DNA (Figure 3B). Our data reveal two viral DNA sites that may correspond to two distinct functional states. In the 3'-processing mode, the viral DNA ends bind integrases from two different dimers, as the DNA is positioned in the 3'-processing active site of monomer B through an interaction with monomer A from the opposite dimer. The C-terminal end of monomer A interacts strongly with both host and viral DNA, close to the integration site of the target DNA and to the 3'OH end of the viral DNA. In the integration intermediate, the viral DNA and the C-terminus of monomer A are displaced to position the 3'OH of the viral DNA towards the target integration sites separated

Figure 3 Integrase domain organisation. (A) Two perpendicular views of the $(IN)_4$ – $(LEDGF)_2$ complex. Each IN/LEDGF complex contains 4 IN molecules (A1, A2, B1 and B2) organised in two IN dimers (1 and 2). The different IN monomers are represented with different colours: monomer A1 in light blue, A2 in dark blue, B1 in light yellow and B2 in yellow. Monomers 1 and 2 are related by a two-fold axis. The density of LEDGF (grey) is obtained by subtracting the model from the EM map. (B) Same views for the complex with DNA in the 3'-processing state. The viral DNA is shown in orange and the target DNA in red.

by the canonical five base-pair stagger (15 Å) (Pruss *et al*, 1994b). These findings agree with previous studies that established the important role of the C-terminal domain for both the concerted integration and the 3' processing (Li and Craigie, 2005). Our results highlight the role of LEDGF as a molecular chaperone that stabilises IN as a tetramer, which was shown to be the catalytically active form of the enzyme (McKee *et al*, 2008).

Analysis of protein–DNA interactions from cross-linking and mutational experiments

Table I summarises results from the cross-linking and mutational experiments. Numerous experiments have been carried out by Heuer and Brown (1997, 1998) and all are in excellent agreement with our model. The N-terminal IN fragment (1–12) has been shown to interact with target DNA and our model reveals that monomer A1 is involved in this contact, as it is in interaction with the nucleotides 10 to 14 (Figure 4A). The C-terminal domains (fragments 213–246 and 247–270) have been shown to interact with both target and viral DNA. Our model confirms these cross-linking data, as the C-terminal domain of monomer A is placed between the target and

viral DNA (Figure 4B). Mutational analysis showed that residues L234, R262 interact with the viral DNA (Lutzke and Plasterk, 1998). Arginine 262 is in proximity of the nucleotides 10–12 of the processed strand, whereas L234 interacts with the first nucleotides of the unprocessed strand (Figure 4C). Cross-linking experiments shown in this report further indicate that K264 interacts with nucleotide 6 from the unprocessed strand, which is fully consistent with our model (Figure 4D). Our structure confirms the findings that two fragments of the CCD (49–69, 139–152) interact with the viral DNA (Figure 4G), fragment 49–69 with nucleotides 1–4 of the processed strand and fragment 139–152 with nucleotides 1–4 of the unprocessed strand. Interestingly, a flexible loop between residues 141 and 145, unresolved in the X-ray structures, is in a key position between the 3'-processing active site and the scissile phosphate in the target DNA. A photo cross-linking experiment showed that K156 and K159 from the core domain interact with the viral DNA (Jenkins *et al*, 1997). In our structure, these lysines interact with nucleotides 4–5 of the processed viral DNA strand (Figure 4H). Close to these residues, glutamine 148, which has been shown to make a specific interaction with nucleo-

Table 1 DNA protein interactions

Domain	Amino acid	Target DNA	Viral DNA		References
			Unprocessed strand	Processed strand	
Nter	<i>1–12</i>	10–14		1–4	Heuer and Brown, 1997
	<i>49–69</i>				Heuer and Brown, 1997
Catalytic core	<i>139–152</i>		1–4		Heuer and Brown, 1997
	<i>E148</i>		2		Johnson <i>et al</i> , 2006
	<i>K156</i>			4–5	Jenkins <i>et al</i> , 1997
	<i>K159</i>			4–5	Jenkins <i>et al</i> , 1997
	CC			3–6	Agapkina <i>et al</i> , 2006
Cter	C-ter		<i>4–9</i>		Agapkina <i>et al</i> , 2006
	<i>213–246</i>	6–10	5–6		Heuer and Brown, 1997
	<i>L234</i>		1		Lutzke and Plasterk, 1998
	<i>247–270</i>	0–4	6–8	10–12	Heuer and Brown, 1997
	<i>R262</i>			10–12	Lutzke and Plasterk, 1998
	<i>K264</i>		6		This work

Summary of interactions seen in cross-linking and mutational experiments. The nucleotides numbering is as described in Figure 4F. In italic are the published residues found to interact with DNA and in bold are the corresponding interacting nucleotides that are found in our structure (for reference (Agapkina *et al*, 2006), in italic are the published nucleotides and in bold the interacting IN domain).

tide 2 of the unprocessed strand (Johnson *et al*, 2006), is placed at a suitable position in our structure. Finally, the probing of IN–DNA interactions with viral DNA analogue highlighted the nucleotides of the processed (Pr) and unprocessed (Un) strands of viral DNA in interaction with IN (Agapkina *et al*, 2006). These interactions are fully consistent with our model, as nucleotides 3–6 from the unprocessed strand interact with the catalytic core domain of monomer B2, whereas nucleotides 4–9 from the unprocessed strand interact with the C-terminal end of monomer A1 (Figure 4I). Taken together, the full agreement of these independent biochemical and genetic experiments validates our model.

Integration mechanism

Physiological integration of the viral cDNA into the host genome requires to join both ends of the viral cDNA on opposite strands of the target host DNA. The enzymatic reaction can be divided into two steps; an early 3' processing of the viral DNA ends and a later concerted integration step occurring after target DNA binding. Our data revealed two viral DNA positions that reflect these two steps (Figure 5A). In the first step, the viral DNA ends interact with the IN tetramer allowing its precise positioning into the catalytic site for 3' processing. In the second state, after target DNA binding, two main conformational changes have to occur. First, the N- and C-terminal domains move to clamp the host DNA and second, the C-terminal of monomer A is rotated (140°) to direct the 3'OH of the viral DNA, which is rotated by about 90° towards the scissile phosphate of the target DNA (Figure 5B; Supplementary Figure 8 and Movie 4). A higher resolution structure of biochemically stabilised intermediates will be needed to precisely describe these conformational changes. This proposed mechanism for HIV-1 integration into the host genome clarifies the positions and the roles of the integrase domains and of the DNA molecules involved in the reaction. The need of an asymmetric dimer in which each monomer performs a distinct function explains that the use of the F185K mutation (strongly implicated in the dimerisation interface) that leads to a symmetric dimer structure did not result in a functional structural model. The importance of the interactions between the C-terminal domain with viral and host DNA for the integration mechanism make it a target of

choice for structure-based drug design experiments. For example, our structure predicts that mutations that affect HIV-1 integrase sensitivity to the new generation of integrase inhibitors (e.g., Raltegravir) (Malet *et al*, 2008) are mainly implicated in the interactions with viral and host DNA and would have no effect on the IN/LEDGF interactions.

The essential role of cellular cofactors in the assembly and folding of individual molecules into large functional units is further illustrated and will prove to have major implications for other retrovirus families. The understanding of the integration process and of the role of cellular cofactors has wide implications for basal mechanisms of numerous viral-induced pathologies such as retroviral-induced cancers. Viral integration near cellular proto-oncogenes or tumour suppressors can affect the expression of these proteins by proviral insertional mutagenesis and can lead to tumorigenesis (Lewinski and Bushman, 2005). The mechanism of DNA recombination as catalysed by the IN/LEDGF complex may represent a model system to understand retroviral-induced pathologies that provokes genome instability.

Materials and methods

Production and purification of IN and LEDGF

The GST-IN and His₆-LEDGF were produced in *Escherichia coli* BL21(DE3) host strain (Novagen) transformed by pRARE isolated from Rosetta DE3 strain (Novagen). Cells were grown at 37°C in 1 L of LB containing 10% sucrose, 100 µg/ml ampicillin and 17 µg/ml chloramphenicol up to an OD₆₀₀ of 0.6. Protein expression was then induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside and cells were further grown overnight at 18°C before harvesting. Cells were resuspended in 50 ml of buffer A [50 mM HEPES pH 7.0, 1 M NaCl, 5 mM MgCl₂, 7 mM CHAPS, 2 mM β-mercaptoethanol, 5% glycerol, 1 mM PMSF] per litre of culture and lysed by a microfluidizer processor (Microfluidics). GST-IN was enriched from soluble fraction by chromatography on a 5-ml GSTrap4B column (GE Healthcare) and eluted with buffer A containing 20 mM reduced glutathione. The His₆-tagged LEDGF was processed in the same way except that the lysate was passed through a 5-ml Hitrap-Ni column (GE Healthcare), and the protein was eluted by a linear gradient of imidazole (from 7.5 mM to 250 mM) in buffer A. GST-IN and His₆-LEDGF were mixed at a 1:3 molar ratio, respectively, and dialysed against buffer B [50 mM HEPES pH7.0, 150 mM NaCl, 5 mM MgCl₂, 2 mM β-mercaptoethanol]. The complex was adsorbed onto glutathione-Sepharose (GE Healthcare) and eluted with 20 mM reduced glutathione in buffer B. Fractions were pooled and the GST-tag was removed by 2 days digestion at 4°C with 50U of P3C

Figure 4 Consistency of the proposed model with published experimental protein-DNA interactions data. The nucleotides or amino acids that have been shown to interact by biochemical or mutational experiments are shown in green. (A) N-terminal domain of monomer A1 interacting with target DNA (red). (B-D) The C-terminal domain of monomer A1 is interacting with the target DNA (red) and viral DNA (orange). (E) Full complex: the squares represent the position of the magnified views. (F) Nucleotide numbering used in this publication. *Un* represents the unprocessed and *Pr* represents the processed strand of the viral DNA. The target DNA numbering starts at 0, which is placed at the pseudo two-fold axis relating the two IN monomers. (G-I) Interactions of the viral DNA with the catalytic core of IN.

protease per mg of complex during dialysis against buffer B. The cleaved IN-(His₆-LEDGF) complex was again passed through glutathione-Sepharose to remove noncleaved protein, concentrated using a Centrprep YM-50 device (Amicon), and loaded at 1 ml/mn onto a Superdex 200HR gel filtration column (GE Healthcare) pre-equilibrated in buffer B. Peak fractions were used directly for electronic microscopy. Protein concentration was determined using the Bradford colorimetric assay (Bio-Rad).

High-mass MALDI ToF mass spectrometry analysis

High-mass MALDI mass spectra were obtained using an MALDI-TOF Reflex IV (Bruker Daltonics, Bremen, Germany) equipped with a HM1 high-mass retrofit system (CovalX AG, Zürich, Switzerland). The instrument was operated in linear mode by applying an accelerating voltage of 25 kV. The HM1 system was operated in the high-mass mode with acceleration voltage set to 20 kV and gain voltage set to 2.7 kV. Mass spectra were acquired by averaging 300 shots (three different positions in each spot and 100 shots per position). All subsequent mass spectra acquisitions were carried out

by applying the same laser fluency. The cross-linking reactions were realised using a solution containing different cross-linkers specific for amino and sulphhydryl groups. The cross-linking reactions were carried out using a reagent composed of iodoacetic acid *N*-hydroxysuccinimide ester, Octanedioic acid di-*N*-hydroxysuccinimide ester and ethylene glycol bis-succinimidylsuccinate (K200 MALDI MS analysis Kit, CovalX AG, Zürich, Switzerland). Cross-linking reagents were added to 10 µl at 0.6–1.2 µM of the complex and incubated for 2 h to achieve a complete reaction. The sample was then mixed with a matrix solution (1:2 v/v) of sinapic acid (10 mg/ml) containing 70% acetonitrile (v/v) diluted in deionised water with 0.1% TFA. After mixing, 0.5 µl of the mixture was deposited on the MALDI target using the dried-droplet method.

IN and IN-LEDGF complex activity assay

Integrase activity was studied by mixing 100 nM of integrase (free or in complex with 50 nM LEDGF/p75) and 2.5 nM of ³²P-labelled U5-duplex (21- or 40-mer) in 20 µl of a buffer containing 20 mM Hepes

Figure 5 Model for DNA integration. **(A)** The distance of 15 Å (5 bp) corresponds to the integration distance of the viral DNA (yellow) in the target DNA (red). The red arrows represent the movement of the viral DNA towards the integration intermediate structure. **(B)** Sketch showing the relative positions of the DNA molecules for the 3' processing and integration reactions (the target DNA is in black and the viral DNA in orange and red). The integration reaction, which takes place through a nucleophilic attack by the 3'OH end on a 5' phosphate of the target DNA, requires the displacement of the integrase C-terminus together with the viral DNA.

(pH 7.5), 10 mM dithiothreitol, 7.5 mM $MgCl_2$ at 37 °C. Free recombinant integrase protein was produced in *E. coli* and purified as described previously (Leh *et al*, 2000). Various incubation times were used. The reaction was stopped with 80 µl of a buffer containing 9 mM Tris/HCl (pH 7.5), 6 mM EDTA, 0.125 mg ml⁻¹ glycogen, 400 mM NaOAc. DNA fragments were precipitated with ethanol, then suspended in loading dye (80% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on a 20% polyacrylamide denaturing gel. Gels were analysed on a STORM 840 PhosphorImager (Molecular Dynamics). The amount of reaction products was measured and quantified using IMAGE QUANT 4.1 software.

Negative stain cryo-electron microscopy and image processing

The purified complexes were diluted to a concentration of 20 µg/ml in buffer B containing 20% glycerol. Ten microliters of this preparation were placed on a 10-nm thick carbon film treated by a glow discharge in air. After 2 min of adsorption, the specimen was

negatively stained with a 2% (w/v) uranyl acetate solution, sandwiched between a second carbon film and allowed to partially dry for 2 min before being plunged into liquid nitrogen. Images were recorded under low-dose conditions with an FEI Tecnai 20 TEM operating at 200 kV and at a temperature of -172 °C. The image data set was acquired on photographic plates (SO163, Eastman-Kodak, Rochester, NY) at 62,000 × magnification and at defocus values ranging from 1.2 to 2.1 µm. The micrographs were digitised using a drum scanner (Primescan D7100, Heidelberg) to obtain a final pixel spacing of 0.164 nm. Image analysis was carried out using the IMAGIC software package (Van Heel *et al*, 1996) (Image Science Software, Berlin, Germany), as described earlier (Jawhari *et al*, 2006). The negatively stain cryoEM images were CTF corrected by phase inversion using the EMAN software (Ludtke *et al*, 1999), a semiautomated software for high-resolution single-particle reconstructions (Supplementary Protocol 1). The analysis of negatively stained single IN/LEDGF complexes revealed the presence of a two-fold axis that was used to reconstruct a first 3-D model (Supplementary Figure 2). The negatively stain cryoEM data

set was analysed independently to obtain a stable clustering of the images into 1000 classes whose projection directions were determined using the initial model to produce an anchor set for angular reconstruction. The 3-D model was further refined by multiple rounds of alignment using reprojection of the model as references followed by standard clustering and 3-D reconstruction steps. Resolution was assessed using the Fourier shell correlation (FSC) function between two 3-D reconstructions obtained by splitting the data set in two and by using the 0.145 FSC criterion (Rosenthal and Henderson, 2003). The IN/LEDGF/DNA complex was processed identically and prepared as described in Supplementary Protocol 2. An independent model was constructed making use of the two-fold symmetry to avoid any bias from the DNA-free model. The final model was calculated after having sorted out 25% of the images, which correlated better with the DNA-free model and were likely to have not bound DNA.

Protein–DNA cross-linking

Cross-linking experiment. Detailed procedures as well as the sequences and the synthesis of modified oligonucleotides are described in Supplementary Protocol 3. To prepare the substrate analogue with an aldehyde group, solution of the DNA duplex 40-U5ald (15 pmol) containing the oligonucleotide U5A6-ald in 10 µl of 30 mM Na-acetate, pH 4.5 was supplemented by 15 µl of 200 mM fresh solution of NaIO₄ in 30 mM NaOAc and incubated at 25°C for 1 h. Then 2 M LiClO₄ water solution was added; the duplex was precipitated with acetone, dissolved in 600 µl of 20 mM Hepes, pH 7.2, 7.5 mM MgCl₂, 1 mM DTT and incubated with 10 µg of IN-LEDGF complex at 37°C for 1 h. The mixture was supplemented with 5 µl of 4 M fresh solution of NaBH₃CN and incubated at 37°C for 30 min. This reaction was repeated 25 times and 7 µl of 10% SDS solution was added to the total mixture and evaporated to 150 µl. This mixture was then desalted on Microcon YM-3 centrifugal filters (Millipore Corporation, Bedford, MA, USA). In parallel, the same procedure was carried out by using the DNA duplex 40-U5ald (2 pmol) containing ³²P-labelled oligonucleotide U5A6-ald.

Trypsin-mediated hydrolysis of the integrase covalent complex with the substrate analogue. The reaction mixture obtained after the cross-linking was dissolved in 200 µl of 40 mM NH₄HCO₃, 1 mM DTT, 2 mM EDTA and 2.5 µg of trypsin (Sigma-Aldrich Inc., St Louis, MO, USA) and incubated at 37°C for 20 h. Trypsin was extracted with phenol/chlorophorm/iso-amyl alcohol (25:24:1 v/v/v); the cross-linking products containing mixture was purified by gel-filtration on Microcon YM-3 centrifugal filters and loaded onto 20% acrylamide/urea gel. Gel images obtained for ³²P-labelled products were recorded on a STORM 840TM Phosphorimager (Molecular Dynamics, Sunnyvale, USA). The main band corresponding to cross-linked conjugate was cut out. The conjugate was eluted from

gel using 300 mM Na-acetate and then purified using a series of successive procedures, that is, desalting on Microcon YM-3 centrifugal filters, ion-exchange chromatography on Fe³⁺-NTA-agarose prepared as described in Stensballe *et al* (2000) followed by the second desalting on Microcon YM-3 centrifugal filters. A part of the purified samples was analysed by MALDI-TOF mass-spectrometry and another part underwent HF treatment to hydrolyze the oligonucleotide fragment of the conjugates.

HF treatment. The DNA-peptide conjugate was dried in vacuum, supplemented with 40 µl of HF (48%; Fluka, Neu-Ulm, Germany) and incubated at 4°C for 16 h. Then the excess of HF was evaporated and HF traces were completely removed by three successive evaporations with methanol.

Model building and fitting

Flexible fitting of the atomic structures in the NS-cryoEM map was done using normal modes analysis (Tama *et al*, 2004; Suhre *et al*, 2006). The composite model and structure superposition were done using the O program (Jones, 2004). Normal mode flexible and rigid body fitting were done using the procedure described (Supplementary Protocol 4) using NORMA (Suhre *et al*, 2006) and URO (Navaza *et al*, 2002). Regularisation of the structure was done with REFMAC (Winn *et al*, 2001) (Supplementary Figures 6–8, Tables 1–3 and Movies 1–3). Map manipulation and difference maps were carried out using CCP4 (The CCP4 suite, 1994), COOT (Emsley and Cowtan, 2004) and UCSF chimera (Pettersen *et al*, 2004). Figures were produced with UCSF chimera and Pymol (DeLano, 2002).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

This work was supported by grants from the CNRS, the INSERM, the French National Agency for Research against AIDS (ANRS), the RFBR (Russian Foundation for Basic Research, 08-04-01252 and 08-04-01293), the TrioH European Project (FP6 Grant 503480), the SPINE 1 and 2 European Project (FP6 Contract N° QLG2-CT-2002-00988 and 031220) and the ARC. FM was supported by a post-doctoral fellowship from SIDACTION. We thank Julie Thompson and Jean Claude Thierry for useful suggestions about the paper. We also thank the members of the Structural Biology and Genomics platform, CEGBS, Illkirch and the members of the IGBMC's common services.

References

- Agapkina J, Smolov M, Barbe S, Zubin E, Zatsen T, Deprez E, Le BM, Mouscadet JF, Gottikh M (2006) Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA. *J Biol Chem* **281**: 11530–11540
- Busschots K, Voet A, De MM, Rain JC, Emiliani S, Benarous R, Desender L, Debyser Z, Christ F (2007) Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J Mol Biol* **365**: 1480–1492
- Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM (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
- Cherepanov P (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity *in vitro*. *Nucleic Acids Res* **35**: 113–124
- Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci USA* **102**: 17308–17313
- Cherepanov P, Maertens G, Proost P, Devreese B, Van BJ, Engelborghs Y, De CE, Debyser Z (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* **278**: 372–381
- Chiu TK, Davies DR (2004) Structure and function of HIV-1 integrase. *Curr Top Med Chem* **4**: 965–977
- Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**: 77–85
- DeLano WL (2002) *The PyMOL Molecular Graphics System*. Delano Scientific: Paolo Alto, CA
- Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**: 1981–1986
- Eijkelenboom AP, Sprangers R, Hard K, Puras Lutzke RA, Plasterk RH, Boelens R, Kaptein R (1999) Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins* **36**: 556–564
- Emiliani S, Mousnier A, Busschots K, Maroun M, Van MB, Tempe D, Vandekerckhove L, Moisan F, Ben-Slama L, Witvrouw M, Christ F, Rain JC, Dargemont C, Debyser Z, Benarous R (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in

- chromosome tethering and HIV-1 replication. *J Biol Chem* **280**: 25517–25523
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132
- Engelman A, Bushman FD, Craigie R (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J* **12**: 3269–3275
- Engelman A, Hickman AB, Craigie R (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J Virol* **68**: 5911–5917
- Engelman A, Mizuuchi K, Craigie R (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**: 1211–1221
- Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V (2005) HIV-1 integrase crosslinked oligomers are active *in vitro*. *Nucleic Acids Res* **33**: 977–986
- Gao K, Butler SL, Bushman F (2001) Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J* **20**: 3565–3576
- Golas MM, Sander B, Will CL, Luhrmann R, Stark H (2005) Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell* **17**: 869–883
- Hare S, Shun MC, Gupta SS, Valkov E, Engelman A, Cherepanov P (2009) A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog* **5**: e1000259
- Haren L, Ton-Hoang B, Chandler M (1999) Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* **53**: 245–281
- Hehl EA, Joshi P, Kalpana GV, Prasad VR (2004) Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins. *J Virol* **78**: 5056–5067
- Heuer TS, Brown PO (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 TS, Brown PO (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
- Jawhari A, Uhring M, De CS, Crucifix C, Tocchini-Valentini G, Moras D, Schultz P, Poterszman A (2006) Structure and oligomeric state of human transcription factor TFIIIE. *EMBO Rep* **7**: 500–505
- Jenkins TM, Esposito D, Engelman A, 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
- Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci USA* **92**: 6057–6061
- Johnson AA, Santos W, Pais GC, Marchand C, Amin R, Burke Jr TR, Verdine G, Pommier Y (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem* **281**: 461–467
- Jones TA (2004) Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr D Biol Crystallogr* **60**: 2115–2125
- Karki RG, Tang Y, Burke Jr TR, Nicklaus MC (2004) Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *J Comput Aided Mol Des* **18**: 739–760
- Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, Brochon JC, LeCam E, Coulaud D, Auclair C, Mouscadet JF (2000) Determinants of Mg²⁺-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* **39**: 9285–9294
- Lewinski MK, Bushman FD (2005) Retroviral DNA integration—mechanism and consequences. *Adv Genet* **55**: 147–181
- Li M, Craigie R (2005) Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem* **280**: 29334–29339
- Li M, Mizuuchi M, Burke Jr TR, Craigie R (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J* **25**: 1295–1304
- Liao C, Karki RG, Marchand C, Pommier Y, Nicklaus MC (2007) Virtual screening application of a model of full-length HIV-1 integrase complexed with viral DNA. *Bioorg Med Chem Lett* **17**: 5361–5365
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**: 82–97
- Lutzke RA, Plasterk RH (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
- Maertens G, Cherepanov P, Pluymers W, Busschots K, De CE, Debyser Z, Engelborghs Y (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* **278**: 33528–33539
- Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, 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
- Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M, Tchertanova L, Peytavin G, Reynes J, Mouscadet JF, Katlama C, Calvez V, Marcelin AG (2008) Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor *in vitro*. *Antimicrob Agents Chemother* **52**: 1351–1358
- McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M (2008) Dynamic modulation of HIV-1 integrase structure and function by cellular LEDGF protein. *J Biol Chem* **283**: 31802–31812
- Navaza J, Lepault J, Rey FA, Alvarez-Rua C, Borge J (2002) On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr* **58**: 1820–1825
- Nazabal A, Wenzel RJ, Zenobi R (2006) Immunoassays with direct mass spectrometric detection. *Anal Chem* **78**: 3562–3570
- Petersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612
- Pruss D, Bushman FD, Wolffe AP (1994a) Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc Natl Acad Sci USA* **91**: 5913–5917
- Pruss D, Reeves R, Bushman FD, Wolffe AP (1994b) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J Biol Chem* **269**: 25031–25041
- Ren G, Gao K, Bushman FD, Yeager M (2007) Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA. *J Mol Biol* **366**: 286–294
- Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* **423**: 145–150
- Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**: 721–745
- Schultz P, Olland S, Oudet P, Hancock R (1996) Structure and conformational changes of DNA topoisomerase II visualized by electron microscopy. *Proc Natl Acad Sci USA* **93**: 5936–5940
- Stensballe A, Jensen ON, Olsen JV, Haselmann KF, Zubarev RA (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom* **14**: 1793–1800
- Suhre K, Navaza J, Sanejouand YH (2006) NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr D Biol Crystallogr* **62**: 1098–1100
- Tama F, Miyashita O, Brooks III CL (2004) Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* **337**: 985–999
- The CCP4 suite (1994) Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760–763
- Turlure F, Devroe E, Silver PA, Engelman A (2004) Human cell proteins and human immunodeficiency virus DNA integration. *Front Biosci* **9**: 3187–3208
- Van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* **116**: 17–24

- Wagenknecht T, Frank J, Boublik M, Nurse K, Ofengand J (1988) Direct localization of the tRNA—anticodon interaction site on the Escherichia coli 30 S ribosomal subunit by electron microscopy and computerized image averaging. *J Mol Biol* **203**: 753–760
- Wang JY, Ling H, Yang W, Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* **20**: 7333–7343
- Wielens J, Crosby IT, Chalmers DK (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J Comput Aided Mol Des* **19**: 301–317
- Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **57**: 122–133
- Yoder KE, Bushman FD (2000) Repair of gaps in retroviral DNA integration intermediates. *J Virol* **74**: 11191–11200
- Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* **93**: 13659–13664