Review TRENDS in Genetics Vol.22 No.7 July 2006

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Retroviral DNA integration: HIV and the role of LEDGF/p75

Angela Ciuffi and Frederic D. Bushman

University of Pennsylvania School of Medicine, Department of Microbiology, 3610 Hamilton Walk, Philadelphia, PA, 19104-6076, USA

To replicate, a retrovirus must integrate a DNA copy of its RNA genome into a chromosome of the host cell. Integration is not random in the host genome but favors particular regions, and preferences differ among retroviruses. Several mechanisms might play a part in this favored integration targeting: (i) open chromatin might be preferentially accessible for viral DNA integration; (ii) DNA replication during cell division might facilitate access of integration complexes to favored sites; and (iii) cellular proteins bound to the host chromosome might tether integration complexes to favored regions. This review summarizes recent advances in understanding the mechanisms of retroviral integration, focusing on LEDGF/p75 – the first cellular protein shown to have a role in directing HIV DNA integration. Studies on LEDGF/p75 indicate that it directs HIV integration site selection by a tethering interaction, whereas the chromatin accessibility or cell cycle models are less well supported. Understanding viral integration will help improve the safety of retrovirus-based vectors used in gene therapy.

Introduction – retroviral integration

The location of viral genome integration has important consequences for both the virus and the host. Viral gene activity varies in different chromosomal locations [1–3]. Efficient transcription of viral genes, following integration in active transcription units, can facilitate productive host infection. By contrast, repression of transcription can occur when the virus integrates into centromeric heterochromatin, resulting in impaired viral replication and providing a potential mechanism for transcriptional latency [1–3].

Retroviral integration can also affect gene expression in the host cell owing to the insertion of viral promoters or enhancers near cellular genes. One consequence that has been thoroughly studied in animal models is insertional activation of oncogenes, which results in tumorigenesis. Gene inactivation by retroviral integration is also well documented in animal models [4,5].

Unfortunately, insertional mutagenesis leading to oncogene activation has recently also been seen in human gene therapy using retroviral vectors. Retroviral vectors are one of the most popular vehicles for clinical gene delivery because they can stably and precisely insert new DNA sequences into a wide variety of cell types. Retrovirus-based gene therapy has been successful in treating X-linked severe combined immunodeficiency, but in three out of 11 patients treated, the cellular *LMO-2* oncogene was activated by insertion of the murine leukemia virus-based vector used, contributing to the development of leukemia in these patients [6,7]. These setbacks have resulted in intense interest in integration target site selection by the integrating vectors used for gene therapy.

The retroviral life cycle

After binding and recognition of cellular receptors by the viral envelope, fusion between the viral and the cellular membranes occurs, either at the plasma membrane or in a vesicle following endocytosis, leading to release of the viral core into the cytoplasm. The viral RNA genome is then reverse transcribed to yield a linear double-stranded DNA copy. The DNA remains associated with viral and host cellular proteins in a nucleoprotein complex known as the preintegration complex (PIC). This complex is subsequently translocated to the nucleus where the viral DNA becomes integrated into the host chromosome. Once integrated, the viral DNA uses cellular machinery to transcribe viral genomes and express its proteins, which assemble at a cellular membrane. After the budding of new viral particles, the viral protease processes the viral polyproteins into mature proteins, yielding infectious particles (Figure 1) [4].

The integration mechanism

The initial steps of DNA integration are catalyzed by the viral integrase enzyme (IN) (Figure 2) [4,5]. First, IN cleaves two nucleotides from the 3′ end of each viral DNA strand (terminal cleavage reaction), exposing a 3′-OH group. Second, IN simultaneously breaks the host DNA and joins it to the viral 3′ end (strand transfer) by a single step transesterification reaction [8]. The third step of integration is thought to involve host DNA repair proteins to remove the two nucleotide overhang and to fill in the DNA gaps [9].

IN alone is sufficient to perform the first two steps of the integration reaction *in vitro*. Nevertheless, *in vitro* integration assays showed that IN activity can be increased by the presence of some viral and cellular proteins [10–19],

Corresponding authors: Ciuffi, A. (aciuffi@mail.med.upenn.edu), Bushman, F.D. (bushman@mail.med.upenn.edu).

Available online 26 May 2006

 0168-9525/$ - see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2006.05.006

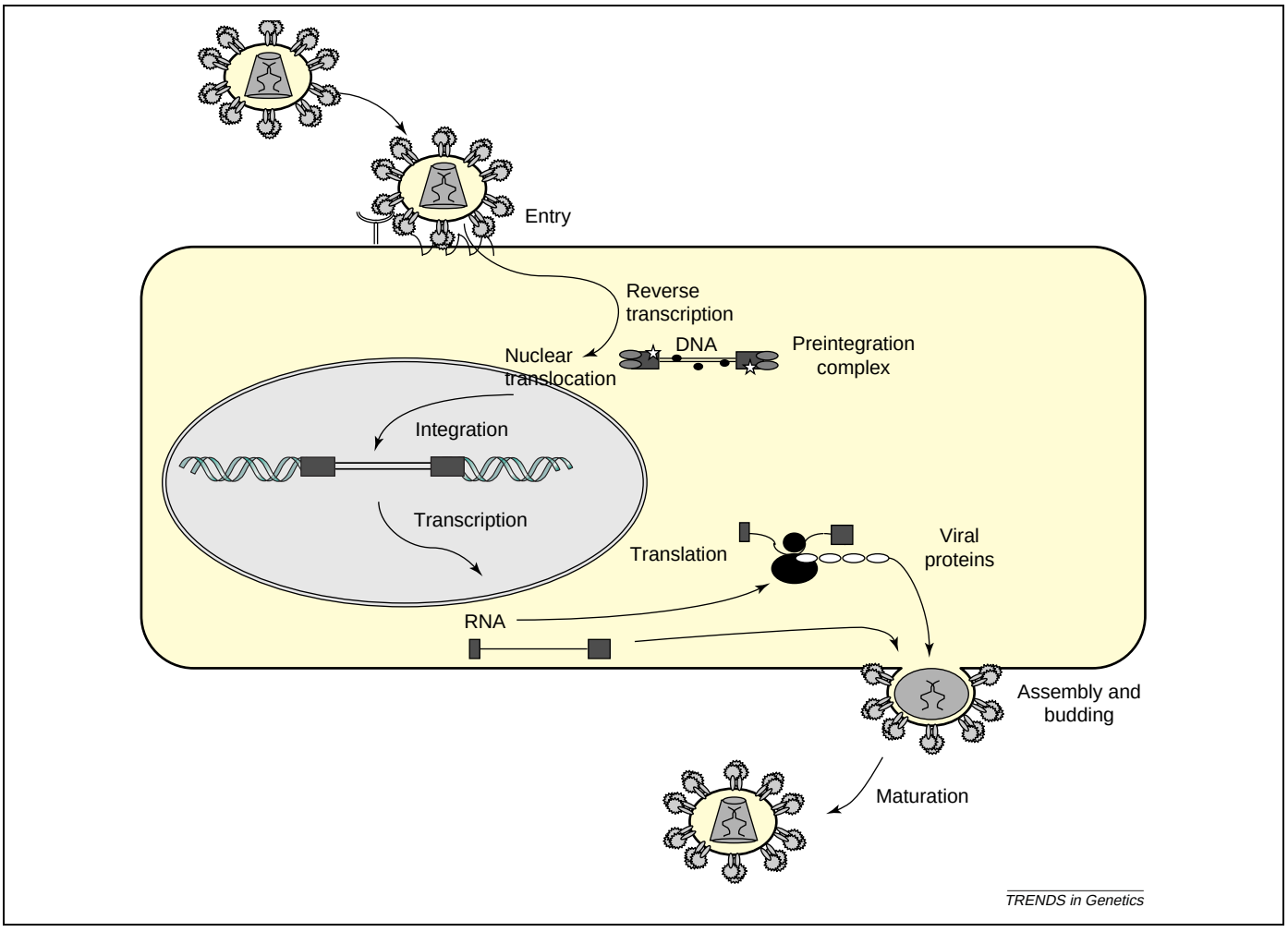

Figure 1. An overview of the retroviral life cycle. The virion attaches to a target cell through an interaction between the viral envelope and a specific receptor (in HIV, depicted here, the viral envelope interacts sequentially with two receptors: the CD4 molecule and co-receptor molecule CCR5 or CXCR4), followed by the fusion of the viral and cellular membranes and the release of the viral core into the cytoplasm of the host cell. Subsequently, the viral genomic RNA is reverse transcribed into a linear double stranded DNA by the virally encoded reverse transcriptase. The viral DNA, still associated with incoming viral proteins and cellular proteins, in the PIC, migrates to the nucleus, where the viral integrase promotes the integration of the viral DNA into the host genome. The cellular RNA polymerase II synthesizes the viral RNAs (genomic RNA and mRNA transcripts) from the proviral DNA. Following protein synthesis, new virions assemble and bud from a cellular membrane (internal or plasma membrane). The processing of virion proteins by the viral protease leads to mature and infectious viral particles.

and several of these might also be involved in target site selection (discussed in the next section).

Integration target site selection differs among retroviruses

With the completion of the human genome sequence, it has become possible to investigate the characteristics of retroviral integration *in vivo* using genomic methods. To perform genome-wide studies of integration, Schroder, A.R. *et al.* infected cultured cells with viral particles and amplified DNA fragments spanning host-virus DNA junctions by ligation-mediated PCR [20]. This method involves (i) digestion of the genomic DNA extracted from the infected cells with restriction enzymes; (ii) ligation to a compatible linker; (iii) PCR amplification of retroviral integration sites using one primer specific to the viral long terminal repeat (LTR) and another primer specific to the linker; (iv) cloning and sequencing of the integration sites. The sequences are then trimmed to remove the viral primer and the viral LTR end, and aligned with the human genome sequence. Chromosomal features at or near the integration sites, such as transcription units (TUs), repeated sequences and G/C content, can then be identified and tabulated. Transcriptional profiling of infected cells can also be performed to assess the transcriptional activity of genes hosting integration events. To determine whether integration is favored in a specific genomic feature, results are compared with a random set of genomic sites generated *in silico*. As another control, integration can be carried out *in vitro* using naked human DNA as an integration target and the integration sites can be sequenced. Under these conditions, *in vitro* integration seemed to be almost random [20]. Control and experimental integration sites are then compared statistically to identify chromosomal features that affect integration frequency.

Such genome-wide studies have shown that retroviral integration into the host genome does not occur randomly but that certain chromosomal features were favored (reviewed in Ref. [21]). With regard to nucleotide sequence at the point of integration, weak preferences have been observed when many *in vivo* sites were aligned, which

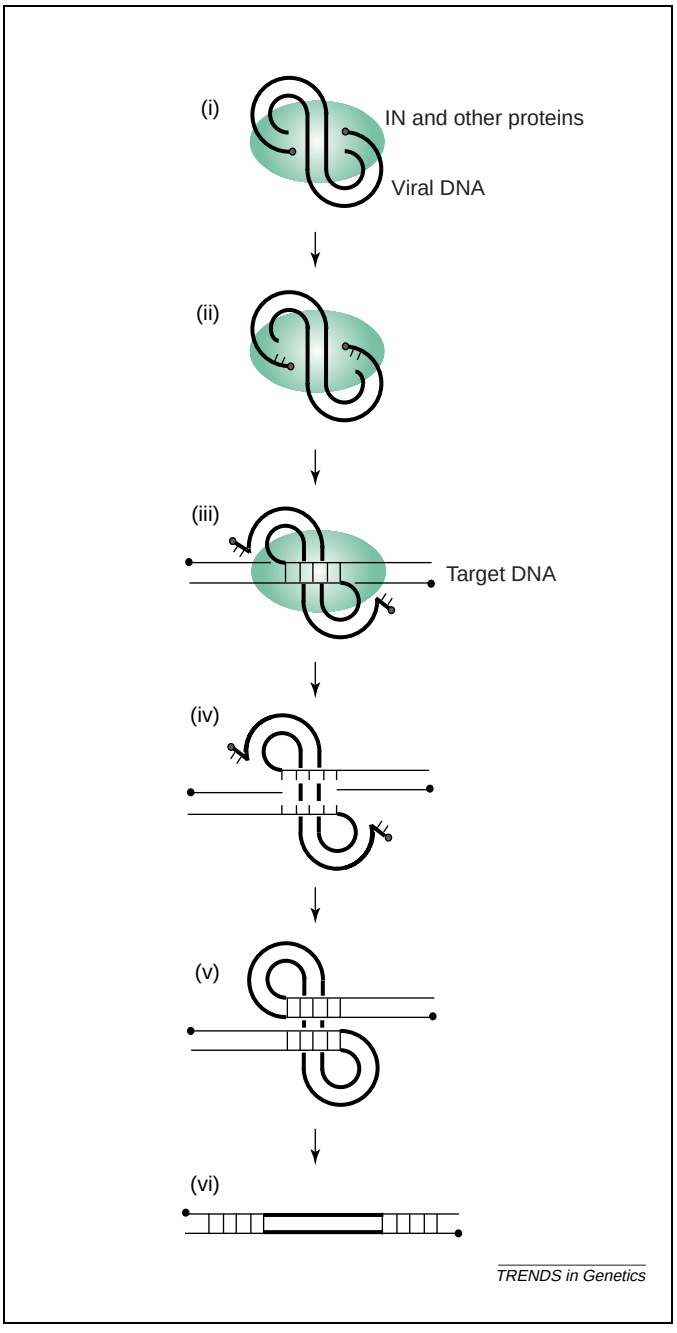

Figure 2. The integration of viral DNA into the host cell. (i) The viral DNA (bold line) is associated with viral and cellular proteins in the PIC (in green). The circles represent the 5′ ends of the DNA. (ii) The viral integrase (a component of the PIC) cleaves two nucleotides off each viral DNA strand, leaving a dinucleotide 5′ overhang. (iii) The viral integrase simultaneously breaks the target DNA (5-bp apart from the HIV IN) and joins it to the 3′ recessed viral DNA end. (iv) The integration intermediate contains unpaired bases at each viral-target DNA junction. (v) The DNA repair machinery of the host is hypothesized to fill in the five nucleotide gaps and cleave out the 5′ protruding viral DNA ends. (vi). The integrated proviral DNA is surrounded by a 5-bp duplication of the host genome (for HIV).

differed among HIV, murine leukemia virus (MLV) and avian sarcoma leukosis virus (ASLV) [22–24]. Synthesis of the expected favored target site from HIV integration and testing *in vitro* revealed that HIV PICs did favor integration at the favored sequence, but other sites were also targeted fairly efficiently [22]. For HIV, integration at the chromosomal level occurs preferentially in transcription units (TUs) that are transcriptionally active; there is no preference for introns or exons (Table 1) [3,20,25–29].

The integration of simian immunodeficiency virus (SIV) in human or simian cells showed a similar pattern of favored integration in active TUs [30–32], indicating that lentiviruses as a group might have similar target site preferences. MLV, by contrast, favors integration at or near gene promoter regions, as defined by location at the 5′ end of TUs and CpG islands [26,30,31,33]. ASLV integration is almost random, showing only a weak preference for transcription units [26,27,34]. Recently, foamy viruses (which are from a distinct retrovirus subfamily) have shown to have yet another integration distribution pattern, with a modest preference for CpG islands, but no preference for TUs [35].

Most studies so far were performed in human cells, revealing a specific integration pattern for each retrovirus. But how much does the species of the target cell affect integration site selection? To investigate this question, Barr *et al.* analyzed integration sites from human and chicken cells infected with HIV or ASLV [27]. They found that HIV favored integration in TUs in chicken cells, as seen for human cells. ASLV integration patterns were also similar in chicken and human cells. Hematti *et al.* analyzed MLV integration sites from simian hematopoietic stem cell infection, and showed that they were located predominantly in transcription start sites and CpG islands [30], consistent with a previous study of MLV integration in human cells [33]. Thus, so far, integration site preferences have not varied significantly in cells from different vertebrate species, suggesting that cellular machinery directing integration site selection is conserved.

We next consider three models (which are not mutually exclusive) to explain these integration target site preferences: (i) accessibility of chromatin: (ii) cell cycle effects; and (iii) binding to tethering protein(s) [21].

The chromatin accessibility model

One hypothesis to explain integration target site selection is that 'open chromatin' is preferentially accessible to the PIC [36]. Thus, transcribed regions become exposed, thereby facilitating access of retroviral integration complexes. Therefore, one might expect that if chromatin accessibility has a major role in integration site selection, all retroviruses would preferentially integrate at the same accessible hotspots. Retrovirus integration sites, however, differ among viruses, at odds with the simplest version of the model. *In vitro*, nucleosome-associated and highly distorted DNA is actually favored as a target over naked DNA, inconsistent with the idea that packing DNA in nucleosomes obstructs integration [37–40].

Another line of evidence comes from analysis of DNase I hypersensitive (HS) sites, which are thought to be markers for accessible chromatin. Recently, Crawford *et al.* carried out a high-throughput analysis of DNase I HS sites on the human genome in quiescent $CD4^+$ T cells [41,42]. These studies showed that DNase I HS sites were enriched near genomic features such as CpG islands, regions upstream of genes and active transcription units. DNase I HS sites correlated with MLV favored locations for integration, but not with favored sites for HIV integration (M.K. Lewinski *et al.*, unpublished). This is

Table 1. Integration frequency in transcription units and repeated sequences in human cells

Chromosomal feature	HIV[a,b]	MLV[a,c]	ASLV[a,d]	Matched random[e]
Transcription unit catalog				
RefSeq	68.9%[f]	42.3%[f]	43.5%[f]	33.8%
Repeats				
SINEs Alu	13.0%[f]	7.8%[f]	7.1%[f]	9.5%
SINEs MIR	3.1%	7.9%[f]	4.9%[f]	3.1%
DNA elements	3.5%	4.4%[f]	4.1%	3.2%
LTR elements	4.5%[f]	6.4%[f]	8.3%	8.7%
LINEs	17.4%[f]	12.7%[f]	20.7%	22.7%

[a]There are slight differences in the frequencies of integration compared with those given in earlier publications, which reflect more stringent analysis criteria.
[b]Pooled HIV integration sites from infected 293T, IMR90, Jurkat, PBMC, SupT1, HeLa and H9 cells [3,20,22,25,28,33].
[c]MLV integration sites from infected HeLa cells [33].
[d]Pooled ASLV integration sites from infected 293T-tva and HeLa cells [26,35].
[e]Matched random integration sites are computer-generated random sites, corrected for experimental biases as a result of restriction site placement.
[f]Significantly different from random, $P<0.05$.

inconsistent with the idea that chromatin accessibility, as reported by DNase I hypersensitivity, is responsible for integration site selection because the effect is retrovirus specific. Together these data suggest that favorable effects of open chromatin are unlikely to fully account for the mechanism of integration targeting, although chromatin structure might still affect integration site selection to some extent.

Cell cycle effects

Another hypothesis is that the cell cycle phase at the time of integration might affect site selection. Retroviruses differ in their ability to infect dividing and non-dividing cells. HIV can infect dividing and non-dividing cells, whereas MLV PICs require cell division to enter the nucleus and integrate [33,43]. It is therefore possible that the cell-cycle phase at the time of infection might influence the placement of integration sites of the incoming virus, accounting for the differences in integration targeting between HIV and MLV.

To begin to investigate this hypothesis, HIV integration site selection was compared in dividing and non-dividing cells. We compared non-dividing primary lung fibroblast IMR-90 cells, which were arrested by serum starvation and contact inhibition, with dividing IMR-90 cells [29]. Our analysis of HIV integration targeting revealed no reduction of integration in TUs in non-dividing IMR-90 compared with that in dividing IMR-90 cells [29], suggesting that cell division was not important for this bias. Analysis of HIV integration sites in macrophages, another non-dividing cell type that is a clinically relevant HIV target, showed similar results [44]. The effect of arrest in other phases of the cell cycle on the distribution of integration sites by HIV and other retroviruses has yet to be determined. However, the data available so far have failed to make a compelling case that the cell cycle phase at the time of integration is a major determinant.

Integration targeting by cellular tethering protein(s)

The tethering model proposes that cellular proteins serve as a 'bridge' between the PIC and specific sites in cellular DNA. Integration targeting by tethering is implicated in site selection by the Ty retrotransposons of *Saccharomyces cerevisiae* [45–48]. Ty elements lack an envelope gene and thus an extracellular step in their life cycle, but otherwise replicate like retroviruses. Ty elements integrate into the genome of the host cell through integrase enzymes related to the retroviral integrases. All the steps of retrotransposition occur within the same cell, thus the choice of integration site is important, because any disruption of the host genome can be lethal for both the host and the transposon. The favored target sites for the Ty5 element are heterochromatin at telomeres and the silent mating loci of the yeast genome, which do not disrupt gene activity. These sites are bound by the Sir4p protein, which specifically recruits the Ty5 IN, promoting integration at nearby sites [47]. Similarly, Ty3 IN requires RNA polymerase III transcription factors to direct highly selective integration into Pol III transcript start sites [45,46,49], a target that does not impair Pol III transcription.

In principle, any component of the retroviral PIC, viral or cellular, could serve as a docking point for a cellular tethering factor that targets retroviral integration. Among the candidates potentially affecting HIV integration site selection are the viral PIC proteins, MA, Vpr, IN and cellular factors associated with PIC, such as barrier to autointegration factor (BAF), High Mobility Group A1 (HMGA1), EED, p300, Integrase Interactor 1 (Ini-1) and LEDGF/p75 [3,21,48,50–53,73]. Recent data indicate that LEDGF/p75 in particular influences HIV target site selection [28].

The LEDGF/p75 protein

LEDGF/p75 was first identified by several laboratories because of its ability to bind tightly to HIV IN [18,52]. LEDGF/p75 binding is specific for lentiviral IN proteins (e.g. those of HIV, SIV and FIV) (Box 1) [54–56]. LEDGF/p75 binding also protects IN against proteolysis *in vivo* [18,57].

Box 1. The integrase enzyme

Viral integrase proteins range from 288 to 450 amino acids in length. Integrase proteins are composed of three domains: the N-terminal domain contains a Zinc-finger-like motif (HHCC motif), the catalytic core domain containing the essential acidic residues DD35E and the C-terminal domain, which is the least conserved. The lentiviral integrases are more closely related to each other than viral integrases. The α-retrovirus group, including ASLV, forms another clade. The γ-retroviruses including MLV form another distinct clade [70–72]. Only the lentiviruses have been found to bind to LEDGF/p75, potentially providing an explanation for some of the differences in integration site distribution.

LEDGF/p75 was originally isolated and characterized as a general transcriptional co-activator. Ge *et al.* identified two transcriptional cofactors using functional assays *in vitro*, p75 and p52, that are encoded by differentially spliced transcripts derived from *PSIP1* [58]. Both proteins share the same first 328 amino acids and differ only in their C-terminal portion. They can both interact with transcriptional activators, although different ones, and also interact with components of the basal transcription machinery [58]. In an early study, Singh *et al.* screened a human lens epithelial cell library to identify proteins involved in lens epithelial cell growth and survival, yielding a protein named lens epithelium-derived growth factor (LEDGF) [59,60], which proved to be identical to the transcriptional co-activator p75 (hence the name LEDGF/p75). LEDGF/p75 has been suggested to increase cell survival following stress-induced apoptosis, in part through the induction of stress-related gene expression [61,62]. LEDGF/p75 is ubiquitously expressed and contains a nuclear localization signal (NLS), consistent with a role in transcription in the nucleus. Although particular DNA sequences have been reported to be bound by LEDGF/p75 *in vitro*, LEDGF/p75 has been shown recently to bind to many unrelated sequences, suggesting that its DNA-binding ability is not sequence specific [63].

The NLS of LEDGF/p75 drives the nuclear localization of HIV IN when both proteins are produced at high levels [54,64,65]. However, during infection, HIV IN nuclear localization is not dependent on LEDGF/p75 [54,65], suggesting that multiple mechanisms exist to mediate the nuclear translocation of the HIV PIC. In addition to the NLS, LEDGF/p75 contains a PWWP domain (i.e. Pro-Trp-Trp-Pro, a domain implicated in protein–protein

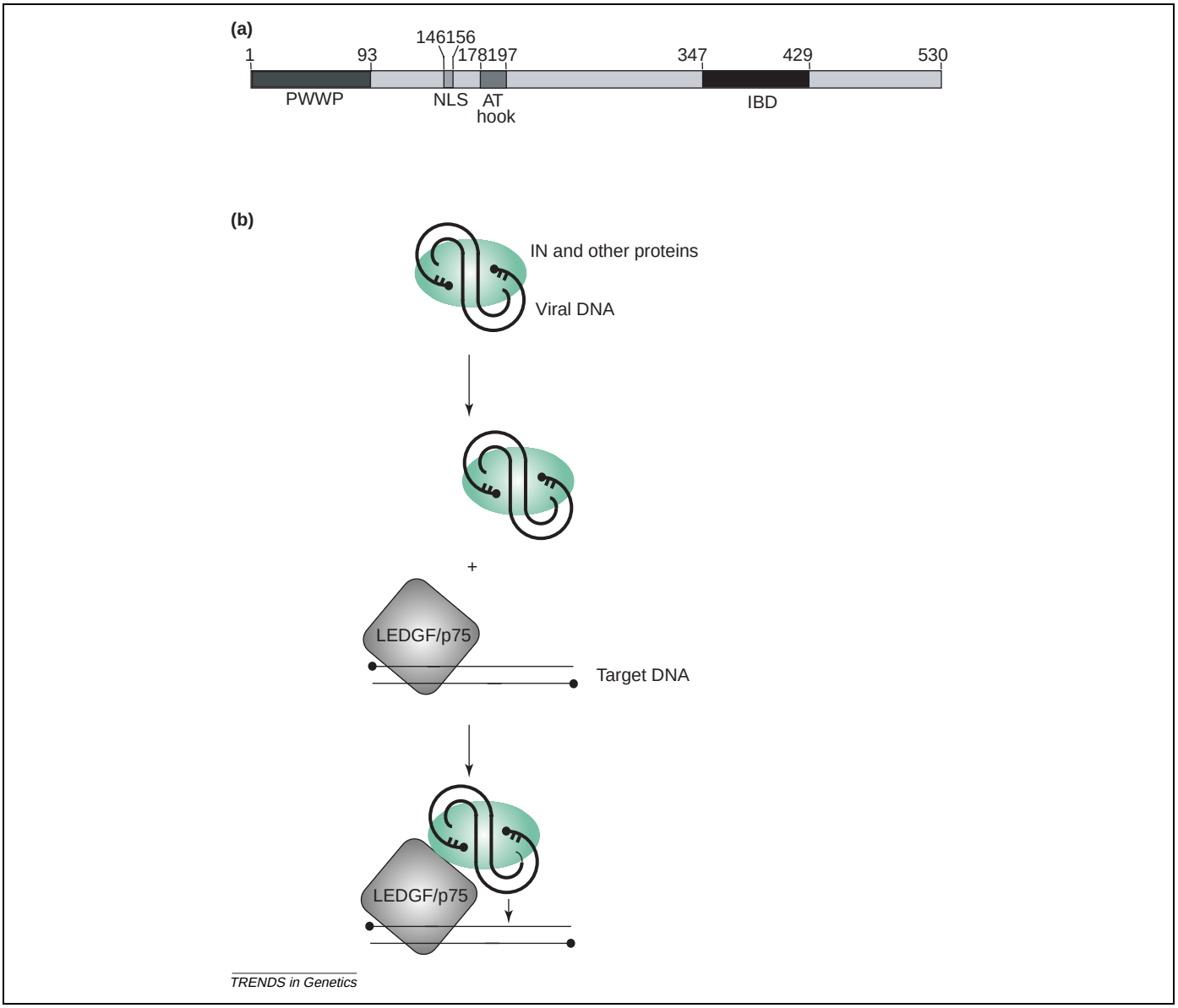

Figure 3. LEDGF/p75 recruits the HIV PIC and promotes HIV integration at a nearby site. **(a)** The structure of LEDGF/p75, showing the location of its domains (PWWP, NLS, AT hook and IBD). **(b)** LEDGF/p75 directs integration via a tethering mechanism. The PIC (containing IN and other proteins), shown in green, surrounds the viral DNA. The viral IN interacts with the LEDGF/p75 protein (shown in grey). Whether LEDGF/p75 is initially bound to the host genomic DNA, or first binds the PIC in solution, has yet to be determined. The LEDGF/p75–IN association promotes HIV DNA integration near the chromosomal sites of LEDGF/p75 binding, via a tethering mechanism.

interactions), an AT-hook motif and an integrase-binding domain (IBD) (Figure 3a). Both the NLS and AT-hook motifs have been shown to be involved in DNA-binding and chromosome association [19,65].

The LEDGF/p75-HIV IN interaction is mediated by the IBD, consisting of LEDGF/p75 residues 341–429 [54,65–68], and the core domain of HIV IN [55]. The crystal structure of the catalytic domain of HIV IN (residues 50–212) in a complex with IBD has been solved; it revealed a pair of LEDGF-IBD molecules bound at symmetry-related positions at the interface of the IN dimer [67,68]. Interhelical loop residues (I^{365}, D^{366} and F^{406}) of the IBD become buried and interact with exposed residues at the dimer interface, specifically V^{165} to K^{173}, and with a hydrophobic patch composed of residues A^{129} to W^{132}.

LEDGF/p75 is conserved among vertebrate species [66]. Human LEDGF/p75 shares 92% identity with murine LEDGF, 69% with chicken LEDGF and 56% with frog LEDGF. Interestingly, the LEDGF-IBDs are >83% identical, providing a potential explanation for the maintenance of HIV integration site selection in chicken cells [27].

Nevertheless, LEDGF/p75 seems dispensable for HIV replication in cell culture because HIV infectivity is unaffected when cells are knocked down for LEDGF/p75 by RNA interference [19,28,54]. However, one caveat is that residual expression of LEDGF/p75 protein could suffice for function, thus potentially explaining the absence of phenotype on HIV replication, although knock-downs with <10% residual expression have been characterized.

LEDGF/p75 in targeting HIV integration

Recently, using cells knocked-down for LEDGF/p75, we showed that LEDGF/p75 can affect HIV integration site distribution [28]. We used RNA interference to knock-down the expression of LEDGF/p75 to undetectable levels (as assayed by western blot) in 293T and Jurkat-derived cells, and to a lesser extent in HOS cells. After infection, HIV integration was favored in TUs in the control cell lines, as in previous studies, but all three cell lines depleted for LEDGF/p75 showed reduction in the frequency of integration into TUs. However, the preference for HIV to integrate into TUs was not completely abolished in the cell lines that were depleted for LEDGF/p75, suggesting either that residual expression of LEDGF/p75 in the knock-down was sufficient for some function, or that other factors are involved.

Furthermore, analysis of the GC content of HIV integration sites indicated that relatively AT-rich regions were more favored for integration than GC-rich regions in cells when LEDGF/p75 was present. This is consistent with the presence of an AT-hook motif in LEDGF/p75, which has been shown in heterologous proteins to mediate favored DNA-binding at AT-rich regions – thus LEDGF/p75 binding to AT-rich sites apparently boosted integration into these sequences.

Another approach to investigate the role of LEDGF/p75 queried the frequency of integration in LEDGF/p75-modulated genes. LEDGF/p75 modulates gene expression by binding to DNA, so LEDGF/p75-regulated genes should have more bound LEDGF/p75 than randomly selected genes, and thus should be preferential integration targets. To investigate this possibility, we performed transcriptional profiling of an LEDGF/p75 knocked-down cell line and matched control using microarrays to identify differentially expressed genes [28]. A comparison of differentially expressed genes with those hosting HIV integration events indicated that LEDGF/p75-modulated genes were favored targets in the control cell line, but this preference was abolished in LEDGF/p75 depleted cells, consistent with the idea that LEDGF/p75 acts as a tethering protein. These data suggest that LEDGF/p75 is involved in promoting HIV DNA integration in TUs and in AT-rich regions, thus providing the first example of an apparent tethering protein for retroviral integration (Figure 3b).

This work was extended using studies of HIV integration *in vitro*. In these studies, the LEDGF/p75 IBD was artificially tethered to a specific site on the integration target DNA *in vitro* using synthetic fusion proteins. Integration by HIV was found to favor integration at DNA sites near the bound fusions. Thus a protein–protein interaction between the LEDGF/p75 IBD and HIV IN can cause favored local integration *in vitro* (A. Ciuffi *et al.*, unpublished), modeling the proposed tethering interaction *in vivo*.

Conclusions and perspectives

HIV integration is not random but occurs preferentially in active TUs, probably as a result of a strategy to promote efficient viral gene expression [1–3,20,69]. LEDGF/p75, a transcriptional coactivator, is the first cellular protein shown to affect HIV integration site distribution *in vivo* [28] and provides evidence that a tethering mechanism can mediate retroviral integration site selection. The recently reported crystal structure of the LEDGF/p75 IBD bound to the HIV IN catalytic domain provides a structural picture of a key interaction [67]. The observation that two IBD monomers bound simultaneously to the catalytic domain dimer suggests that polyvalent binding of LEDGF/p75 might increase the avidity of the interaction *in vivo* and so have a cooperative effect. However, LEDGF/p75 depletion reduces HIV integration frequency in TUs but does not abolish it, so it seems likely that there are other factors contributing to HIV integration site selection. Currently, there are no convincing data to support the chromatin accessibility or cell cycle models (but centromeres seem to be disfavored at the chromosomal level, [22]), although it is certainly possible that future experiments might also unveil a role for these mechanisms. These recent studies help unravel the mechanisms involved in retrovirus integration site selection and might eventually allow the modification of retrovirus-based vectors to improve the safety of gene therapy.

Acknowledgements

We thank members of the Bushman laboratory for helpful discussions and for manuscript improvements. This work was supported from grants by US National Institutes of Health AI52845 and AI34786. A.C. was supported in part by a fellowship by the Swiss National Science Foundation.

References

1 Jordan, A. *et al*. (2001) The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* 20, 1726–1738

2 Jordan, A. *et al*. (2003) HIV reproducibly establishes a latent infection after acute infection of T cells *in vitro*. *EMBO J.* 22, 1868–1877

3 Lewinski, M.K. *et al*. (2005) Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* 79, 6610–6619

4 Coffin, J.M. *et al*. (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press

5 Bushman, F.D. (2001) *Lateral DNA Transfer: Mechanisms and Consequences*, Cold Spring Harbor Laboratory Press

6 Hacein-Bey-Abina, S. *et al*. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348, 255–256

7 Hacein-Bey-Abina, S. *et al*. (2003) LMO2-associated clonal T-cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 400–401

8 Engelman, A. *et al*. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67, 1211–1221

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

10 Kalpana, G.V. *et al*. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 266, 2002–2006

11 Farnet, C.M. and Bushman, F.D. (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. *Cell* 88, 483–492

12 Carteau, S. *et al*. (1997) Human immunodeficiency virus Type 1 nucleocapsid protein specifically stimulates MG^{2+}-dependent DNA integration *In Vitro*. *J. Virol.* 71, 6225–6229

13 Chen, H. and Engelman, A. (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15270–15274

14 Li, L. *et al*. (1998) Modulation of activity of Moloney murine leukemia virus preintegration complexes by host factors *in vitro*. *J. Virol.* 72, 2125–2131

15 Carteau, S. *et al*. (1999) Coupled integration of human immunodeficiency virus cDNA ends by purified integrase *in vitro*: stimulation by the viral nucleocapsid protein. *J. Virol.* 73, 6670–6679

16 Harris, D. and Engelman, A. (2000) Both the structure and DNA binding function of the barrier-to-autointegration factor contribute to reconstitution of HIV type 1 integration *in vitro*. *J. Biol. Chem.* 275, 39671–39677

17 Gao, K. *et al*. (2003) Cofactors for human immunodeficiency virus type 1 cDNA integration *in vitro*. *J. Virol.* 77, 1598–1603

18 Cherepanov, P. *et al*. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278, 372–381

19 Vandegraaff, N. *et al*. (2005) Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology* 346, 415–426

20 Schroder, A.R. *et al*. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529

21 Bushman, F. *et al*. (2005) Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* 3, 848–858

22 Carteau, S. *et al*. (1998) Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* 72, 4005–4014

23 Holman, A.G. and Coffin, J.M. (2005) Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. U. S. A.* 102, 6103–6107

24 Wu, X. *et al*. (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* 79, 5211–5214

25 Mitchell, R. *et al*. (2003) Global analysis of cellular transcription following infection with an HIV-based vector. *Mol. Ther.* 8, 674–687

26 Mitchell, R.S. *et al*. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* DOI: 10.1371/journal.pbio.0020234 (http://biology.plosjournals.org/)

27 Barr, S.D. *et al*. (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* 79, 12035–12044

28 Ciuffi, A. *et al*. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* 11, 1287–1289

29 Ciuffi, A. *et al*. (2006) Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. *Mol. Ther.* 13, 366–373

30 Hematti, P. *et al*. (2004) Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* DOI: 10.1371/journal.pbio.0020423 (http://biology.plosjournals.org/)

31 Dunbar, C.E. (2005) Stem cell gene transfer: insights into integration and hematopoiesis from primate genetic marking studies. *Ann. N. Y. Acad. Sci.* 1044, 178–182

32 Crise, B. *et al*. (2005) Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* 79, 12199–12204

33 Wu, X. *et al*. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749–1751

34 Narezkina, A. *et al*. (2004) Genome-wide analyses of avain sarcoma virus integration sites. *J. Virol.* 78, 11656–11663

35 Trobridge, G.D. *et al*. (2006) Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1498–1503

36 Panet, A. and Cedar, H. (1977) Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. *Cell* 11, 933–940

37 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

38 Pryciak, P.M. *et al*. (1992) Retroviral integration into minichromosomes *in vitro*. *EMBO J.* 11, 291–303

39 Pruss, D. *et al*. (1994) Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5913–5917

40 Pruss, D. *et al*. (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* 269, 25031–25041

41 Crawford, D.H. *et al*. (2004) Identifying gene regulatory elements by genome-wide recovery of DNAse hypersensitive sites. *Proc. Natl. Acad. Sci. U. S. A.* 101, 992–997

42 Crawford, G.E. *et al*. (2006) Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 16, 123–131

43 Yamashita, M. and Emerman, M. (2004) Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J. Virol.* 78, 5670–5678

44 Barr, S.D. *et al*. (2006) HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. *Molecular Therpy* DOI:10.1016/j.ymthe.2006.03.012 (http://www.sciencedirect.com/science/journal/15250016)

45 Kirchner, J. *et al*. (1995) *In Vitro* position-specific integration of a retrovirus-like element requires Pol III Transcription Factors. *Science* 267, 1488–1491

46 Sandmeyer, S. (2003) Integration by design. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5586–5588

47 Zhu, Y. *et al*. (2003) Controlling integration specificity of yeast retrotransposon. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5891–5895

48 Bushman, F.D. (2003) Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115, 135–138

49 Boeke, J.D. and Devine, S.E. (1998) Yeast retrotransposons: finding a nice quiet neighborhood. *Cell* 93, 1087–1089

50 Bushman, F.D. (2002) Integration site selection by lentiviruses: biology and possible control. *Curr. Top. Microbiol. Immunol.* 261, 165–177

51 Engelman, A. (2003) The roles of cellular factors in retroviral integration. *Curr. Top. Microbiol. Immunol.* 281, 209–238

52 Turlure, F. *et al*. (2004) Human cell proteins and human immunodeficiency virus DNA integration. *Front. Biosci.* 9, 3187–3208

53 Engelman, A. (2005) The ups and downs of gene expression and retroviral DNA integration. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1275–1276
54 Llano, M. *et al*. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* 78, 9524–9537
55 Emiliani, S. *et al*. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J. Biol. Chem.* 280, 25517–25523
56 Busschots, K. *et al*. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* 280, 17841–17847
57 Llano, M. *et al*. (2004) Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *J. Biol. Chem.* 279, 55570–55577
58 Ge, H. *et al*. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J.* 17, 6723–6729
59 Singh, D.P. *et al*. (1999) Lens epithelium-derived growth factor: increased resistance to thermal and oxidative stresses. *Invest. Ophthalmol. Vis. Sci.* 40, 1444–1451
60 Singh, D.P. *et al*. (2000) Lens epithelium-derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. *Gene* 242, 265–273
61 Singh, D.P. *et al*. (2001) LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes. *Biochem. Biophys. Res. Commun.* 283, 943–955
62 Shinohara, T. *et al*. (2002) LEDGF, a survival factor, activates stress-related genes. *Prog. Retin. Eye Res.* 21, 341–358
63 Singh, D.P. *et al*. (2006) DNA Binding Domains and Nuclear Localization Signal of LEDGF: Contribution of two helix-turn-helix (HTH)-like domains and a stretch of 58 amino acids of the n-terminal to the trans-activation potential of LEDGF. *J. Mol. Biol.* 355, 379–394
64 Maertens, G. *et al*. (2004) Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *J. Biol. Chem.* 279, 33421–33429
65 Vanegas, M. *et al*. (2005) Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J. Cell Sci.* 118, 1733–1743
66 Cherepanov, P. *et al*. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/-transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.* 279, 48883–48892
67 Cherepanov, P. *et al*. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17308–17313
68 Cherepanov, P. *et al*. (2005) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat. Struct. Mol. Biol.* 12, 526–532
69 Lewinski, M.K. and Bushman, F.D. (2005) Retroviral DNA integration–mechanism and consequences. *Adv. Genet.* 55, 147–181
70 Yokoyama, S. *et al*. (1988) Molecular evolution of the human immunodeficiency and related viruses. *Mol. Biol. Evol.* 5, 237–251
71 Malik, H.S. and Eickbush, T.H. (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J. Virol.* 73, 5186–5190
72 Llorens, C. and Marin, I. (2001) A mammalian gene evolved from the integrase domain of an LTR retrotransposon. *Mol. Biol. Evol.* 18, 1597–1600
73 Miller, F.D. *et al*. (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* 71, 5382–5390