

**The Role of LEDGF/p75 in HIV
Integration**

Keshet Ronen

May 24, 2007

A. Specific Aims

Integration of viral cDNA into the host cell genome is an essential step in the life-cycle of retroviruses such as HIV, as well as a process that is exploited in the development of retroviral gene therapy vectors. While the virally encoded enzyme integrase is sufficient for the integration reaction *in vitro*, cellular cofactors are thought to play a role in the process *in vivo*. Elucidation of the involvement of host factors in retroviral integration can impact both the development of antiviral agents, and the refinement of gene therapy strategies.

LEDGF/p75 is a cellular transcription coactivator that interacts with the integrase of HIV and other lentiviruses, protecting it from proteolysis, enhancing its activity *in vitro* and causing its accumulation on chromosomes, suggestive of a role in tethering to chromatin. In support of this, in cellular HIV infection, LEDGF has been shown to play a role in determining the sites of the host genome at which HIV integrates. Recent data show it to be essential for efficient infection. Whether this block to infection in the absence of LEDGF can be ascribed to its function as a tether for integration or to another function is unclear. Furthermore, if LEDGF does function as an essential tether, an important question is whether it is the sole determinant of integration site selection, or whether there remains an independent role for other cellular factors, such as chromatin accessibility or epigenetic marks.

Two roles can be hypothesized for LEDGF in HIV integration: (i) mediation of integration site selection, and (ii) enhancement of stability or activity of pre-integration complexes (PICs). In this proposal we will address these roles through the following Specific Aims:

Specific Aim 1: To determine the contribution of LEDGF/p75 to integration site selection

1a As tools to test LEDGF's hypothesized role as an essential tether for HIV integration, we will establish a genome-wide profile of LEDGF binding sites and LEDGF-responsive genes. These sites will be used to analyze genomic sites of HIV integration in wild-type and newly developed intensified knockdown cells, with the expectation that genomic sites of LEDGF occupancy and the location of LEDGF-responsive genes should strongly predict the sites of HIV integration.

1b To assess the role of LEDGF as a necessary integration tether, and determine the extent of its cooperation with epigenetic marks shown to correlate with HIV integration, we will analyze integration site targeting in intensified LEDGF-knockdown cells, asking whether previously reported correlations with various genomic features persist in the absence of LEDGF.

1c To test the ability of LEDGF to act as an integration tether, we will engineer artificial tethers based on LEDGF fused to alternative DNA-binding domains and test integration targeting in cells expressing these constructs for redistribution to sites targeted by the new tethers.

Specific Aim 2: To determine the effect of LEDGF/p75 on pre-integration complexes

2a To investigate the effect of LEDGF on the efficiency of the integration reaction we will compare the *in vitro* integration activity of PICs from wild-type and LEDGF-deficient cells.

2b To investigate the effect of LEDGF on the stability of PICs we will compare the resistance of PICs from wild-type and LEDGF-deficient cells to high salt and temperature *in vitro*.

2c To identify the domains of LEDGF that mediate any effects found in Aims 2a and 2b we will test for rescue with truncations of LEDGF.

B. Background and Significance

Human Immunodeficiency Virus (HIV), the causative agent of AIDS, is a lentivirus, in the retrovirus family. Retroviruses are RNA viruses that, upon entry into the host cell, reverse transcribe their genome into double-stranded cDNA, and integrate this into the host genome. Integration is an essential step in the retroviral lifecycle, allowing the provirus to be transmitted as a stable element of the host genome and be used as a template for subsequent expression of viral genes^[1]. Advancement of our understanding of the process of retroviral integration and its consequences impacts a number of scientific areas of interest. Pharmacological inhibitors of the integration step of HIV's life-cycle are currently in clinical trials, and a greater understanding of the process may generate additional therapeutic candidates. The process of retroviral integration has also been harnessed by the gene therapy field, where retroviral vectors are now widely used for gene delivery^[2, 3]. However, some concerns exist regarding the safety of such vectors^[4] and specificity of integration targeting, so the study and manipulation of integration may offer potential improvements in this field.

The integration reaction is composed of three main steps: processing of viral DNA ends, joining of viral to host DNA, and repair of gaps. The viral enzyme integrase (IN) mediates the first two steps. IN removes two nucleotides from the 3' termini of the viral DNA, leaving recessed 3' hydroxyl groups, and then catalyzes the attack by these hydroxyl groups on phosphodiester bonds in the target DNA. This leaves single-stranded gaps between the points of

joining on the two strands, which are probably repaired by host DNA repair enzymes. *In vitro*, purified HIV IN is sufficient for end processing and joining.

In vivo, integration is carried out by a nucleoprotein pre-integration complex (PIC). PICs can be isolated from infected cells and carry out coordinate integration of both ends of endogenous cDNA into target DNA supplied *in vitro* ^[5, 6]. HIV PICs have been shown to contain, as well as IN, the viral proteins MA, RT, NC and Vpr ^[7]. In addition, a number of host proteins have been found to be associated with PICs, including the non-histone chromatin protein HMGA1 ^[8]; a SWI/SNF chromatin remodeling protein, Ini1 ^[9]; a chromatin-associated protein, BAF ^[10] and the transcriptional mediator LEDGF/p75 ^[11]. Studies of the role of HMGA1 *in vivo* suggested it was not important for infection ^[12], the role of Ini1 in infection is thought to be in virion assembly ^[13], and the importance of BAF in HIV infection remains contentious ^[14, 15]. However, evidence is accumulating for an important role of LEDGF/p75 in HIV integration (see below).

While integration shows little sequence specificity ^[16, 17] it has become clear that retroviruses exhibit preferences for the genomic sites at which they integrate ^[18-21]. Indeed different classes of retroviruses show different integration site preferences. In a number of cell types HIV ^[19, 22, 23] and other lentiviruses ^[24, 25] strongly favor transcription units as integration sites, particularly active genes. Recent data (see Preliminary Data) also show a positive correlation between HIV integration and histone modifications associated with transcription (H3 and H4 acetylation, H3 lysine 4 methylation), and a negative correlation with those associated with transcriptional repression (H3K27 methylation and DNA CpG methylation). In contrast, Murine Leukemia Virus (MLV), in the gammaretrovirus group, prefers CpG islands and gene 5' ends ^[20, 21], while Avian Sarcoma and Leukosis Virus (ASLV), an alpharetrovirus, shows the most random integration pattern, only weakly favoring transcription units ^[20]. Hypotheses that have been proposed to explain integration site targeting by retroviruses center around the ideas of chromatin accessibility or tethering. Since much of the DNA in mammalian cells is tightly wrapped into higher order chromatin structures, and these structures change with transcriptional status and cell cycle phase, it may be that integration can only occur in regions that are in an exposed conformation. The bias of MLV integration toward DNase I hypersensitive sites ^[26] and of HIV away from aliphoid repeats located in pericentric heterochromatin ^[27] lend some support to this idea. However, the distinct patterns of different retroviruses in the same cell types are suggestive of virus-specific tethers, rather than simple accessibility. There is growing support

for the idea that retroviral PICs are targeted to particular regions of the host chromatin through interactions with host proteins. Tethering interactions are well documented in yeast retrotransposons, which are closely related to retroviruses^[28, 29], and artificial fusions of integrase to sequence-specific DNA-binding domains have been shown to direct integration to their recognition sites^[30, 31]. Any component of the PIC described above could act as or make contact with a cellular tethering factor. Indeed, studies with LEDGF/p75 have provided good evidence of its role as a host integration tether.

Interest in LEDGF/p75 in the HIV field began when it was found to interact with overexpressed HIV IN in the nucleus^[11]. The protein was identified as p75, one of two splice variants from the PSIP1/LEDGF gene, reported to be a transcriptional coactivator that co-purified with the general transcription factor PC4^[32], and has been implicated in cellular stress responses^[47, 50]. The two splice variants of LEDGF, p52 and p75 (see Fig. 1 in Appendix), share an N-terminus containing a PWWP domain, a nuclear-localization signal and two AT-hook motifs, which have all been implicated in chromatin binding^[33, 34]. p75's C-terminus (divergent from p52) contains a domain that binds lentiviral integrases^[35, 36]. LEDGF/p75 (referred to as LEDGF in this proposal) enhances IN activity *in vitro*^[11], enhances the solubility of IN^[37], and protects it from proteasomal degradation^[38]. It has also been suggested to be a component of functional PICs^[39] and can partly reconstitute the activity of salt-disrupted PICs^[40]. Furthermore, it has been shown to recruit recombinant IN to chromatin^[37, 39, 41] and its depletion by RNAi^[39] or mutations of IN that abrogate interaction with LEDGF^[42] lead to loss of nuclear localization of IN. These observations led to a model of LEDGF as a molecular tether of the integration machinery to host DNA. Indeed, a role for LEDGF in integration site selection was demonstrated by siRNA knockdown^[43]. This led to significant, but partial, reduction in the frequency of integration in transcription units, specifically in LEDGF-responsive genes (as determined by transcriptional profiling of knockdown cells) and A/T-rich regions of the genome hypothesized to be bound by LEDGF by virtue of its AT-hook motif. However the knockdown failed to show any effect on susceptibility to infection. Recently, this has been argued to be due to residual chromatin-associated LEDGF, and intensified knockdown cells have been engineered by stable expression of two lentiviral vectors containing LEDGF shRNA constructs. In these cells, no residuum was detected and HIV infection was reduced 31-fold^[44]. Rescue of the infection defect required both chromatin- and integrase-binding capabilities of LEDGF. The

block was shown to be after nuclear import: quantification of the 2-LTR circle form of viral reverse transcription products, widely used as a mark of nuclear import since host nuclear enzymes mediate the necessary recombination^[45], showed LEDGF-deficient cells were enriched in this form of viral DNA. An essential role for LEDGF's tethering function in HIV infection is also supported by the observation that expression of the integrase-binding domain alone inhibits infection^[44, 46]. Integration site selection in intensified knockdown cells has not been studied.

In understanding the role of LEDGF in HIV infection, a number of questions remain. Is LEDGF essential in directing HIV integration to transcription units? Is the recently reported essential role for LEDGF in HIV infection simply a function of its role as an integration tether? Are other reported effects of LEDGF on IN activity and stability relevant in the context of cellular infection by PICs, and therefore also responsible for this essential role? Furthermore, if LEDGF is essential as a molecular tether, is it the only factor responsible for targeting HIV integration to actively transcribed genes? In this proposal, we seek to investigate the extent to which these two sides of LEDGF/p75 are involved in HIV infection: the role in integration site selection and the role in the stability of PICs and enhancement of the integration reaction.

C. Preliminary Data

An association between certain epigenetic marks and HIV integration was revealed in a recent study by Gary Wang in our laboratory (Wang *et al.* in press). Large integration data sets collected using pyrosequencing (see Aim 1a) enabled the deeply annotated ENCODE (Encyclopedia of DNA Elements) regions (see Aim 1b) to be queried with statistical significance. Fig. 2 summarizes the enrichment of experimental integration sites over matched random controls in sites with increasing levels of each histone or DNA modification. A positive correlation with marks of transcriptional activation is shown (histone 3 and 4 acetylation, histone 3 lysine 4 methylation) and a negative correlation with marks of transcriptional repression (histone 3 lysine 27 methylation, DNA CpG methylation).

Studies of HIV integration site selection in murine embryonic fibroblasts knocked out for the histone methyltransferases Suv39h1 and Suv39h2^[65] also support this finding. These enzymes mediate methylation of histone 3 at lysine 9, another mark of transcriptional repression. These cells were used to ask whether modulating epigenetic marks influences HIV's integration site selection. Preliminary analysis of these data (see Fig. 3) suggests loss of H3K9 methylation

activity leads to a reduction in HIV's preference for integration in genes. These data also show that HIV integration can be studied in murine cells and shows similar trends as in human cells.

D. Experimental Design

Specific Aim 1: To determine the contribution of LEDGF/p75 to integration site selection

Rationale Based on its interaction with IN and ability to target it to chromatin, LEDGF has been proposed as a molecular tether for integration, directing PICs to its genomic binding sites. Ciuffi *et al.*^[43] showed that knockdown of LEDGF expression by siRNA led to significant reduction in the targeting of integration to transcription units and regions expected to be enriched in LEDGF. Though this provided strong evidence for LEDGF's role as a tether, some important questions remain.

a. Where in the genome is LEDGF bound? How well do these sites correlate with integration sites? We will perform ChIP to establish sites of LEDGF occupancy, and use these in analyzing integration site selection.

b. Why did LEDGF knockdown by Ciuffi *et al.* only partially reduce HIV's preference for transcription units? Residual LEDGF^[44] may be the explanation. Alternatively, integration targeting may be mediated by other factors in cooperation with LEDGF. We will analyze integration targeting in intensified LEDGF knockdown SupT1 cells generated by the Poeschla laboratory^[44] to explore these questions.

c. If LEDGF acts as a tether, will fusions of LEDGF's integrase-binding domain with other chromatin-targeting domains redirect integration to new sites?

1a. Genome-wide analysis of LEDGF/p75 binding and transcriptional activation

1a.i Sites of LEDGF/p75 binding

In order to define the genomic regions bound by LEDGF, chromatin immunoprecipitation (ChIP) will be performed to isolate LEDGF-bound chromatin and its distribution in the genome mapped. This could be done by hybridization of immunoprecipitated DNA to a microarray (ChIP-chip), but the expense of such experiments makes coverage of more than a small fraction of the human genome unrealistic. Though ChIP-chip remains an option, we will employ a recently developed technique, ChIP-PET^[51,52], which enables high throughput sequencing of immunoprecipitated fragments across the whole genome. This has been employed to identify genome-wide p53-binding sites^[51]. Cells will be treated with formaldehyde to cross-link DNA

and associated proteins; chromatin will be sonicated to shear DNA and immunoprecipitated with anti-LEDGF antibody; cross-linking will be reversed, and DNA purified. The genomic fragments will then be ligated into a vector (see ^[52]) that incorporates the recognition site of a type IIS restriction endonuclease such as MmeI at each side of the ChIP insert. Such restriction enzymes cleave some distance away from their recognition site (for MmeI, 20bp away), so in this plasmid, cleavage occurs within the inserted ChIP fragment. Thus, upon digestion of the vector with such an enzyme, most of the ChIP fragment insert is removed, leaving only a pair of short (20bp) tag sequences, one from each end of the fragment, known as a paired-end ditag (PET). The PET plasmid will then be end-polished, self-ligated and transformed into bacteria to generate a library. From this library, PET plasmids will be purified and PETs cut out of the vector. These can then be concatenated (for efficiency) and sequenced, using newly-developed high-throughput pyrosequencing ^[54]. In this method, linkers are ligated onto templates, which are then bound to beads by virtue of the linker, with one template per bead on average. They are then amplified by PCR, attached to the beads, within individual droplets of an emulsion, generating a homogeneous population on each bead. The beads are loaded into a picotiter plate, one bead per well, approximately 400,000 beads per plate. Sequencing by synthesis takes place in each well: the four dNTPs are washed over the plate sequentially and release of pyrophosphate from nucleotide incorporation results, through the action of an enzyme system in the buffer, in luciferase activation and light emission. A CCD camera records each flash, building up the sequence in each well one nucleotide at a time, up to 200-300bp.

ChIP can be notoriously difficult with some proteins. We will begin by trying commercially available LEDGF/p75 antibodies. A number exist that bind to the 3' region of LEDGF, absent from the p52 splice variant. Some of these have been tested in immunoprecipitation but none is ChIP qualified. In order to test antibodies, Q-PCR of ChIP products will be performed to verify enrichment of regions of the genome previously shown to be bound by LEDGF/p75, for example the promoters of antioxidant protein 2 ^[47] and involucrin ^[48]. As a reference sample, chromatin from knockdown cells will be immunoprecipitated with the same antibody: little or no DNA is expected to be immunoprecipitated. As a negative control, a gene-poor region not expected to be bound by LEDGF/p75 will be tested. If no commercially available antibody is successful, we will express Flag-tagged LEDGF (resistant to siRNA) in intensified knockdown cells, and use anti-Flag antibody. This construct (without siRNA resistant mutations) has been made in the

laboratory and expressed well in HOS cells. Use of tagged proteins for ChIP experiments has been successful in the past^[49]. Effective immunoprecipitation throughout the ChIP protocol will be monitored by Q-PCR on known LEDGF binding sites (as above) to check enrichment. Three biological replicates of the experiment will be performed.

Sequences obtained by this method will be processed as described by Ng *et al.*^[52]: the PET will be separated into the 5' and 3' tags and aligned onto the genome by BLAT, with quality control to ensure stringent matches and appropriate juxtaposition of the two tags. Genome hits will be dereplicated to prevent artefactual enrichment due to amplification in cloning. In this way, unique blocks of genome covered by immunoprecipitated fragments will be generated. Regions of the genome represented by 2 or more clustered PETs will be considered sites of enrichment due to LEDGF immunoprecipitation, as described in^[52]. Such sites of LEDGF occupancy will be used in analyzing integration site targeting (see Aim 1b).

Given its reported interaction with the basal transcription machinery, we expect LEDGF binding to be enriched in transcription units, particularly those identified as LEDGF-dependent by transcriptional profiling (see below and^[43]). A/T-rich regions have been hypothesized to be enriched by virtue of LEDGF's AT-hook^[43] and there have been reports of LEDGF binding to heat shock and stress-related elements^[50] so such sites may be enriched.

1a.ii Transcriptional profiling of LEDGF/p75 knockdown and wild-type cells

The knockdown used by Ciuffi *et al.* to identify LEDGF-responsive genes has since been shown to be incomplete, so it is likely the documented LEDGF regulon is similarly incomplete. We will carry out transcriptional profiling of intensified knockdown cells from the Poeschla laboratory to generate an exhaustive account of LEDGF-dependent genes. Total RNA will be extracted from scramble and LEDGF shRNA cells and submitted to the Penn Microarray Core Facility for transcriptional profiling using the Affymetrix HU133 plus 2.0 microarray. Probes on this array cover transcription units in the human genome based on GenBank, dbEST, and RefSeq databases. mRNA will be reverse transcribed, cDNA amplified by T7-promoter-mediated transcription, then the resulting cRNA biotin-labeled, hybridized to the microarray and stained. Three biological replicates of each cell type will be prepared. Hybridization data for all probe sets from knockdown and control cells will be compared and differentially expressed genes extracted using a permutation t-test implemented in the SAM software package, assuming a 5% false-discovery rate. As a control for the procedure, reduced hybridization of probes in the exons

of the LEDGF gene should be seen in knockdown cells.

These data will be used to analyze preference for integration in active genes in Aim 1b.

1b. To determine if LEDGF/p75 is an essential and exclusive integration tether

In order to establish definitively whether LEDGF is essential for directing lentiviral integration away from a random distribution, and whether it functions in cooperation with other factors, we will study the integration site distribution in the genomes of wild-type, scramble shRNA and LEDGF-knockdown SupT1 cells ^[44]. These cells already contain integrated HIV-based proviruses expressing the shRNAs, which would interfere with cloning of HIV integration sites from *de novo* infection. However, LEDGF is known to interact with IN proteins of other lentiviruses such as Equine Infectious Anemia Virus (EIAV) and Feline Immunodeficiency Virus (FIV) ^[39], and these show the same integration site distributions as HIV ^[24,25], so their integration will be studied. As a control, the integration in each cell type of MLV, a gammaretrovirus whose IN does not interact with LEDGF ^[39] and which showed no infectivity defect in the knockdown cells ^[44], will also be studied, with the expectation that LEDGF knockdown will not affect its integration site distribution.

The shRNA vectors in these cells contain eGFP and mCherry reporter genes, so prior to infection cells will be sorted by FACS for eGFP and mCherry expression to ensure no outgrowth of cells lacking shRNA expression has occurred. Cells will be infected with VSV-G-pseudotyped EIAV, FIV or MLV-based vectors containing a luciferase reporter gene (available in the laboratory). 20 independent infections of each cell type, each of 10^6 cells, will be performed in parallel to maximize the diversity of proviruses analyzed. Cells will be harvested 48 hours after infection. A sample of the cells will be analyzed by luciferase assay to quantify the infection efficiency. Sites of integrated proviruses will be isolated by linker-mediated nested PCR ^[53]. Briefly, genomic DNA will be extracted from the harvested cells and cleaved with a frequent-cutting restriction enzyme. Linkers will be ligated onto the fragments, and integration sites amplified using nested PCR with primer pairs that bind to the viral LTR and the linker. These amplicons will then be sequenced by pyrosequencing (see Aim 1a.i and ^[54]).

Sequences will be processed using a computational pipeline developed in the laboratory. This identifies sequences containing the appropriate viral LTR, trims this off, and uses BLAT to align the remaining sequence to the host genome. After quality control, annotations of aligned regions will be analyzed using the Chi-square statistical test to compare the proportion of

sequences falling inside or outside of a particular feature. Annotations used include in vs. out of gene, in vs. out of exon, gene expression level, and distance from CpG islands, DNase I hypersensitive sites and transcription start sites. In addition, we will employ a more sophisticated statistical analysis to test for associations with genomic features. This is based on a linear regression model developed by our collaborator Charles Berry, based on the ability of various genomic features to distinguish between integration sets ^[55]. All integration data sets are compared with computationally generated matched random controls in order to correct for possible biases introduced by use of sequence-specific restriction enzymes in recovering sites. For each experimental integration site three random control sites are generated that are the same distance from the appropriate restriction site as the experimental integration.

We expect that in wild-type and scramble shRNA cells, lentiviral vectors will integrate preferentially inside transcription units, as previously reported ^[24, 25]. It is expected that active genes, as obtained in Aim 1a.ii, it will be favored for integration. If LEDGF is an essential integration tether responsible for the reported lentiviral integration pattern, we would expect that sites of LEDGF occupancy obtained in Aim 1a.i will correlate strongly with integration sites in control cells. In the knockdown cells, the preference of lentiviruses for active transcription units, especially LEDGF-regulated genes, is expected to be lost. Alternatively, we may find that intensified LEDGF knockdown does not abolish integration site targeting to active genes further than in previous knockdowns, suggesting independent factors contribute to targeting. It is also possible that sites of LEDGF occupancy will turn out to be poor predictors of integration site selection, calling into question the action of LEDGF as a tether. Comparison of integration preference in regions bound by LEDGF (Aim 1a.i) with that in regions regulated by it (Aim 1a.ii) may shed light on the extent to which LEDGF is acting strictly as a tether. Genes regulated by LEDGF are not necessarily bound by it, so their impact on integration site selection may result also from effects on chromatin accessibility or the presence of other transcription-related marks.

In order to determine whether LEDGF acts exclusively or in cooperation with other factors, we will analyze integration sites that fall in the ENCODE region. This is the 1% of the genome selected by the ENCODE project for detailed annotation of transcription factor binding sites, post-translational histone modifications, DNA methylation etc. It is expected that in control cells EIAV and FIV integration will show the same associations with marks such as histone and DNA modifications as HIV (see Preliminary Data). We will then ask whether the correlations still

exist in LEDGF-knockdown cells. Statistical analysis of the ability of such epigenetic marks to predict the distribution of HIV integration sites (Wang *et al.* in press) suggested that these marks are independent of other annotations. Similarly, our preliminary data that loss of Suv39h alters integration distribution support the idea that factors other than LEDGF play a role in integration targeting. Therefore, we expect that the correlation with these features will persist in the absence of LEDGF. If this is found, it will suggest LEDGF is not exclusive in its role as an integration tether, and other factors or chromatin accessibility are involved in integration site selection. If, however, the correlations are abolished, this would suggest LEDGF/p75 is the only determinant of lentiviral integration site selection, and other correlations are only secondary to its tethering.

A potential pitfall in this approach is that the infection block in the LEDGF-knockdown cells may be so strong that we have difficulty isolating successful integration events. As an additional system, we have obtained from the Bickmore laboratory murine embryonic fibroblasts whose LEDGF locus is disrupted such that the C-terminal integrase-binding domain is not expressed^[56]. The susceptibility of these cells to HIV infection has not yet been determined, but given the high degree of sequence conservation of LEDGF, particularly at the integrase-binding domain^[35], these cells are expected to be a good model. We have sequenced HIV integration sites from murine cells and reproduced previously documented integration trends (see Fig.2).

1c. To assess the ability of LEDGF-based fusions to act as integration tethers

If LEDGF acts as an integration tether by virtue of its integrase- and chromatin-binding motifs, it would be expected that replacement of LEDGF's chromatin-binding domain with other known chromatin-binding domains could similarly mediate integration targeting, but at different sites. If this were the case, such an approach could potentially be taken to exert control over integration site selection of lentiviral gene therapy vectors. Indeed, viruses containing IN fused to DNA binding domains of λ repressor^[30] and the zinc finger protein E2C^[31] show some integration preference for the appropriate target sequences. Ciuffi *et al.*^[57] created fusions of LEDGF's integrase binding domain and the λ repressor DNA binding domain and found increased strand transfer activity near λ repressor binding sites *in vitro*. We will express similar fusions of LEDGF in knockdown cells and test their ability to rescue and redirect integration.

Alternative DNA-binding domains will be fused N-terminally to the following (see Fig. 4): A. full-length LEDGF/p75; B. chromatin-binding-domain deleted (Δ -CBD) LEDGF (residues 226-530); C. LEDGF's C-terminal domain (CTD, 347-530); D. the integrase-binding domain

(IBD, 347-429). The outcomes of Aim 2c may influence to some extent the portions of LEDGF we retain in the fusion proteins for functional integration. All LEDGF sequences will contain synonymous siRNA-resistance mutations. Initially, we will use the sequence-specific DNA-binding domain of λ repressor and engineered zinc-finger proteins (ZFPs) designed to bind specific target sequences. Constructs will be expressed in knockdown SupT1 cells by use of an MLV-based vector.

Appropriate expression, nuclear localization, and DNA-binding of the constructs will be verified. Western blot with a number of LEDGF antibodies will be performed to check for expression. Subcellular localization of LEDGF constructs expressed in knockdown cells will be verified by immunofluorescence using LEDGF antibodies. The constructs, like endogenous LEDGF, should be nuclear. Chromatin binding can also be assayed by fractionation – chromatin-associated proteins are resistant to extraction by Triton X-100, but become released from this fraction by treatment with DNase and salt^[34]. DNA-binding of purified constructs will be tested by gel-shift assay with oligos of the appropriate target sequence and binding to sequence elements *in vivo* will be assayed by ChIP-PCR (using LEDGF or Flag antibodies as described above). λ repressor fusions are expected to bind human sequence elements matching the λ operator consensus, and ZFP fusions to bind the target sequence chosen in their design.

Once expression, localization and DNA-binding have been confirmed in intensified LEDGF-knockdown cells, lentiviral integration site targeting will be examined. As an initial test, rescue of the infection block will be tested with HIV. It is expected that complementation of the knockdown with all constructs bearing the IBD and a chromatin-binding domain will rescue the infection defect in the same way as wild-type LEDGF (assuming all necessary domains of LEDGF are present, as discussed in Aim 2c). Integration site selection of EIAV and FIV will then be analyzed as described in Aim 1b (ASLV will be used as a control, since the cells will contain an integrated MLV vector). It is expected that modified DNA binding specificity will redirect integration to sequences targeted by the engineered domain. Depending on the outcome of Aim 1b, such retargeting could affect the vast majority of integration sites (if LEDGF is the sole integration tether) or could have a more minor effect (if other factors play a strong role), but the findings of Aims 1b and 1c are expected to be in agreement.

Specific Aim 2: To determine the effect of LEDGF on preintegration complexes

Rationale Initial reports of LEDGF/p75 in the HIV field were of its role in enhancing IN activity in *in vitro* reactions^[11], mediating chromatin association and nuclear localization of recombinant IN^[39,41] and protecting it from proteasomal degradation^[38]. However, at first these effects were not validated in the context of a cellular infection, where no effect was seen on the susceptibility of LEDGF-knockdown cells to HIV infection^[40,43]. In the intensified knockdown cells generated by Llano *et al.*^[44] integration was greatly impaired. The question remains however whether in these cells the basis for the integration and infection defect is LEDGF's function as an integration tether (as described in Specific Aim 1) or whether the reported effects on IN activity and stability *in vitro* might also have an impact on cellular infection.

In order to investigate these effects in a context more relevant than purified IN protein, we propose to use pre-integration complexes isolated from intensified knockdown and control cells to study the effect of LEDGF on PICs. Specifically, we will investigate the effect of LEDGF on PIC activity *in vitro* and PIC stability, and add back modified LEDGF constructs to identify the domains necessary for any effects observed. As an additional experimental system, embryonic fibroblasts from mice disrupted at the LEDGF locus^[56] will be also used.

2a. The effect of LEDGF/p75 on the integration activity of PICs

To isolate PICs, wild-type, LEDGF-knockdown, and scramble shRNA SupT1 cells^[44] will be infected with viral supernatant produced by SkSM2 cells^[58], a tetracycline-inducible VSV-G pseudotyped lentivirus packaging cell line which can generate virus particles at titers greater than 10⁶ IU/ml. In previous work with PICs^[6] peak detection of cytoplasmic reverse-transcribed DNA (the form in PICs) was reported 4-6hr after infection. At this time, cells will be washed in isotonic buffer and permeabilized with digitonin. The cytoplasmic fraction, which contains the PICs, will be isolated, concentrated by precipitation in low ionic strength buffer (75mM KCl) and resuspended in isotonic buffer (150mM KCl) for subsequent use.

To measure integration activity, isolated PICs will be incubated with ϕ X174 nicked circular DNA^[7] as an integration target. DNA will be extracted from the reaction mixture and analyzed by Southern blot with radiolabeled LTR probes. Integration of the viral genome into the target DNA will generate a slower-migrating product, and integration activity will be quantified by densitometric scanning of the blots (using a PhosphorImager and ImageQuant software) to measure the relative abundance of unintegrated and integrated viral DNA. Usually cytoplasmic

extracts are assayed without quantifying the viral PIC content – instead the amount of integration product is normalized to the unintegrated cDNA in the Southern blot. However, as a control, a representative extract from each cell type will be used to quantify the viral cDNA yield by Q-PCR. If it is found that PIC yield is reduced in knockdown cells, this could indicate effects on PIC stability (see Aim 2b), but given the wild-type 2-LTR circle data of Llano *et al.* [44], PIC formation and cytoplasmic stability are not expected to be reduced.

If the integration defect reported in the intensified knockdown cells is due to LEDGF's enhancement of PIC activity, in accordance with its reported enhancement of purified IN activity, we would expect greater integration in reactions using PICs from wild-type and scramble shRNA cells than LEDGF-knockdowns. An important control will be to use MLV PICs isolated from the three cell types in the *in vitro* assay. It is expected that MLV PIC activity will be equal in all cell types.

2b. The effect of LEDGF/p75 on the stability of PICs

A possible explanation for the observed defects in HIV infectivity in the absence of LEDGF is that PICs become destabilized before integration can take place. There is evidence for this kind of role for LEDGF from work with purified IN *in vitro*: LEDGF was shown to enhance IN binding to DNA substrates resembling the viral LTR and increase IN solubility [37]. It has also been argued that the effect of LEDGF on IN nuclear localization is due to increased nuclear retention or protection from degradation rather than import [42, 59]. Protection of IN from the proteasome [38] was argued to be unimportant in the phenotype of the intensified knockdown cells by the report of Llano *et al.* that the infection defect was not rescued by mutant LEDGF that retained the ability to protect IN from the proteasome but lacked chromatin binding [44]. Nonetheless, physical stabilization of the PIC by LEDGF remains a possibility and has not been studied. The effect of LEDGF on the biochemical properties of PICs will be tested by extracting PICs as described in Aim 2a from LEDGF-deficient and -sufficient cells and subjecting them to increasing salt concentration and temperature to disrupt them. PICs will be incubated in various concentrations of KCl between 150mM (isotonic) and 1.2M, at which PIC activity is known to be lost [10]. Similarly, PICs will be incubated at temperatures ranging from 4°C to 95°C. After disruption, PICs will be spun through a CL4B gel filtration column. To assay PIC integrity after treatment, activity in *in vitro* integration assays will be determined as above. Another test of integrity will be the accessibility of the viral cDNA to exonucleases exoIII and λ exo, which is

reduced in the intact PIC^[7]. PICs having undergone destabilizing conditions will be treated with each exonuclease and the DNA extracted and analyzed by Southern blot with probes directed to the LTR ends: exonuclease digestion results in a lack of probe hybridization.

If LEDGF stabilizes PICs, we expect PICs extracted from knockdown cells to be destabilized at lower salt concentration and temperature than those extracted from wild-type and control shRNA cells, resulting in reduced *in vitro* activity and increased sensitivity to exonucleases. Again, as a control, MLV PICs will undergo the same treatments, with the expectation that there will be no difference between the stability of PICs extracted from wild-type and deficient cells.

Alternative approaches Llano *et al.* reported normal nuclear import in these knockdown cells, suggesting PIC production and cytoplasmic PIC stability should not be reduced. However, the 2-LTR circle assay has its limitations in measuring nuclear import^[62], and if PIC production was found to be reduced in these cells relative to wild-type, the stability and trafficking of PICs in cells would merit some attention. This could be studied using immunofluorescence of PICs in producer cells, using fluorescently labeled Vpr and MA available in the laboratory. Cytoplasmic colocalization of labeled proteins could be monitored to test for increased degradation in LEDGF-deficient cells. Since SupT1s have a small amount of cytoplasm, such an approach may be more applicable in the knockout murine fibroblasts from the Bickmore laboratory^[56]. Another approach would be to use the bis-arsenical fluorescein derivative FIAsh to label tetracysteine-tagged IN in PICs, as has been used to image PICs trafficking through the cytoplasm and even into the nucleus of infected cells^[63]. Such insight into PIC degradation in the cytoplasm or nucleus could also extend and lend support to findings of the *in vitro* assays of biochemical stabilization proposed here.

2c. Rescue of LEDGF/p75 defects

It will be of scientific and clinical interest to identify the parts of LEDGF responsible for any effects identified in Aims 2a and 2b. Much work has gone into understanding the domain structure of LEDGF/p75^[33, 36, 60], and attempting to target the regions relevant for interaction with HIV IN or enhancement of HIV integration with pharmacological inhibitors^[61, 64].

As an initial control for the effect of LEDGF on PIC activity, complementation of the system with full-length LEDGF will be tested. Though there is much evidence for the direct interaction between IN and LEDGF^[11, 35], and some for the presence of LEDGF in PICs^[39], we will confirm the effect of LEDGF replacement at the level of both the PIC-producer cell, and the PIC

extract. For the former, intensified LEDGF-knockdown cells will be transduced with a retroviral vector expressing shRNA-resistant LEDGF and PICs extracted from these back-complemented cells and tested as outlined above. For the latter, purified full-length LEDGF will be incubated at various concentrations with PICs prior to testing stability and activity. As a control, a non-specific protein such as BSA will be added in the same way, with the expectation that it will not rescue any defects. It is expected that both back-complementation of the knockdown cell line and replacement of purified LEDGF/p75 protein in *trans* in the PIC extracts will mediate rescue.

Assuming complementation by full-length LEDGF, rescue by addition of purified protein to the PIC reaction mix will be tested for the following truncations of LEDGF to identify important domains (illustrated in Fig. 5): A. the N-terminal chromatin-binding domain (CBD, residues 1-226); B. a C-terminal truncation lacking the IBD (Δ IBD, residues 1-347); C. an N-terminal truncation lacking the CBD (Δ CBD, residues 226-530); D. the C-terminal domain containing the IBD (CTD, residues 347-530).

Our predictions about these PIC experiments are based on the following published *in vitro* experiments with purified IN. For stimulation of IN activity, the IBD and CBD synergize, with neither being sufficient alone, and loss of either reducing the effect^[33, 40]. Interestingly, Turlure *et al.* reported that a charged region between the CBD and IBD (residues 226-347) played a role in IN stimulation, and suggested the chromatin-tethering and IN-stimulating functions of LEDGF may be independent^[33]. The IBD but not the PWWP or AT-hooks were shown to be necessary for protection of IN from proteasomal degradation^[38, 44], so similar results could be expected for biochemical stabilization. Based on these reports, we hypothesize that addition of C-terminus alone (construct D) to PICs would not rescue any activity defect but may improve stability. Construct C, which contains the IBD and additional charged region described by Turlure *et al.*, may additionally rescue activity defects. Rescue of either defect is unlikely by constructs lacking the IBD (A and B). Additionally, a site-directed mutagenesis approach can be taken to disrupt particular domains that seem to be important based on the truncations. In addition to these truncation mutants, the fusion proteins constructed in Aim 1 will be tested by extracting PICs from cells expressing these fusions, and also by adding purified constructs in *trans* to PIC extracts from knockdown cells. Fusion proteins would be expected to rescue all defects, analogously to full-length LEDGF.

E. References

1. Coffin JM, Hughes SH, Varmus HE. (1997) *Retroviruses*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
2. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, et al. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288: 669-672.
3. Levine BL, Humeau LM, Boyer J, MacGregor RR, Rebello T, et al. (2006) Gene transfer in humans using a conditionally replicating lentiviral vector. *Proc Natl Acad Sci U S A* 103(46): 17372-17377.
4. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, et al. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348: 255-256.
5. Brown PO, Bowerman B, Varmus HE, Bishop JM. (1987) Correct integration of retroviral DNA in vitro. *Cell* 49: 347-356.
6. Farnet CM, Haseltine WA. (1990) Integration of human immunodeficiency virus type 1 DNA in vitro. *Proc. Natl. Acad. Sci USA* 87: 4164-4168.
7. Miller MD, Farnet CM, Bushman FD. (1997) Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. *J. Virol.* 71: 5382-5390.
8. Farnet C, Bushman FD. (1997) HIV-1 cDNA integration: Requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* 88: 1-20.
9. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 266(5193): 2002-6.
10. Chen H, Engelman A. (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc. Natl. Acad. Sci. USA* 95: 15270-15274.
11. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278: 372-381.
12. Beitzel B, Bushman FD. (2003) Construction and analysis of cells lacking the HMGA gene family. *Nuc. Acids Res.* 31: 5025-5032.
13. Yung E, Sorin M, Pal A, Craig E, Morozov A, et al. (2001) Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. *Nat. Med.* 7: 920-926.
14. Jacques JM, Stevenson M. (2006) The inner-nuclear-envelope protein emerlin regulates HIV-1 infectivity. *Nature* 441(7093): 641-645.

15. Shun MC, Daigle JE, Vandegraaff N, Engelman A. (2007) Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular emerlin protein. *J Virol* 81(1): 166-172.
16. Bor Y-, Miller M, Bushman F, Orgel L. (1996) Target sequence preferences of HIV-1 integration complexes in vitro. *Virology* 222: 238-242.
17. Carteau S, Hoffmann C, Bushman FD. (1998) Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. *J. Virol.* 72: 4005-4014.
18. Shih C-, Stoye JP, Coffin JM. (1988) Highly preferred targets for retrovirus integration. *Cell* 53: 531-537.
19. Schroder A, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110: 521-529.
20. Mitchell R, Beitzel B, Schroder A, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2: E234.
21. Wu X, Li Y, Crise B, Burgess SM. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300: 1749-1751.
22. Barr SD, Ciuffi A, Leipzig J, Shinn P, Ecker JR, et al. (2006) HIV integration site selection: Targeting in macrophages and the effects of different routes of viral entry. *Mol. Ther.* 14: 2 (218-225).
23. Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD. (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* 79: 12035-12044.
24. Hacker CV, Vink CA, Wardell TW, Lee S, Treasure P, et al. (2006) The integration profile of EIAV-based vectors. *Mol Ther* 14(4): 536-45.
25. Kang Y, Moressi CJ, Scheetz TE, Xie L, Tran DT, et al. (2006) Integration site choice of a feline immunodeficiency virus vector. *J Virol* 80(17): 8820-3.
26. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. (1987) Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61: 336.
27. Carteau S, Gorelick R, Bushman FD. (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.
28. Zhu Y, Zou S, Wright DA, Voytas DF. (1999) Tagging chromatin with retrotransposons: Target specificity of the *Saccharomyces* Ty5 retrotransposon with the chromosomal localization of Sir3p and Sir4p. *Genes Dev.* 13: 2738-2749.

29. Kirchner J, Connolly CM, Sandmeyer SB. (1995) In vitro position-specific integration of a retroviruslike element requires pol III transcription factors. *Science* 267: 1488-1491.
30. Bushman F, Miller MD. (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J. Virol.* 71: 458-464.
31. Tan W, Dong Z, Wilkinson TA, Barbas CF,3rd, Chow SA. (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J Virol* 80(4): 1939-1948.
32. Ge H, Si Y, Roeder RG. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J.* 17: 6723-6729.
33. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. (2006) A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nuc. Acids Res.* 34: 1663-1675.
34. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, et al. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J Mol Biol* 360(4): 760-773.
35. Cherepanov P, Devroe E, Silver PA, Engelman A. (2004) Identification of an evolutionarily-conserved domain of LEDGF/p75 that bind HIV-1 integrase. *J. Biol. Chem.* 279: 48883-48892.
36. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, et al. (2005) Solution structure of the HIV integrase-binding domain in LEDGF/p75. *Nat. Struct. Mol. Biol.* 12: 526-532.
37. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, et al. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* 280(18): 17841-17847.
38. Llano M, Delgado S, Vanegas M, Poeschla EM. (2004) LEDGF/p75 prevents proteasomal degradation of HIV-1 integrase. *J. Biol. Chem.* 279: 55570-55577.
39. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, 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.
40. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. (2006) Biochemical and genetic analyses of integrase-interacting protein 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.

41. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* 278: 33528-33539.
42. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, et al. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J Biol Chem* 280(27): 25517-23.
43. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* 11: 1287-1289.
44. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, et al. (2006) An essential role for LEDGF/p75 in HIV integration. *Science* 314: 461-464.
45. Li L, Olvera JM, Yoder K, Mitchell RS, Butler SL, et al. (2001) Role of the non-homologous DNA end joining pathway in retroviral infection. *EMBO J.* 20: 3272-3281.
46. De Rijck J, Vandekerckhove L, Gijssbers R, Hombrouck A, Hendrix J, et al. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J Virol* 80(23): 11498-11509.
47. Fatma N, Singh DP, Shinohara T, Chylack LT, Jr. (2001) Transcriptional regulation of the antioxidant protein 2 gene, a thiol-specific antioxidant, by lens epithelium-derived growth factor to protect cells from oxidative stress. *J Biol Chem* 276(52): 48899-48907.
48. Kubo E, Fatma N, Sharma P, Shinohara T, Chylack LT, et al. (2002) Transactivation of involucrin, A marker of differentiation of keratinocytes, by lens epithelium-derived growth factor (LEDGF). *J. Mol. Biol.* 320: 1053-1063.
49. Lee TI, Johnstone SE, Young RA. (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 1(2): 729-748.
50. Singh DP, Fatma N, Kimura A, Chylack LT, Shinohara T. (2001) LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes. *Bio. Biophys. Res. Com.* 283: 943-955.
51. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell* 124(1): 207-219.
52. Ng P, Tan JJ, Ooi HS, Lee YL, Chiu KP, et al. (2006) Multiplex sequencing of paired-end ditags (MS-PET): A strategy for the ultra-high-throughput analysis of transcriptomes and genomes. *Nucleic Acids Res* 34(12): e84.
53. Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, et al. (2005) Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Micro.* 3: 848-858.
54. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057): 376-380.

55. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. *PLoS Comp. Biol.* 2(11): 1450-1462.
56. Sutherland HG, Newton K, Brownstein DG, Holmes MC, Kress C, et al. (2006) Disruption of *Ledgf/Psip1* results in perinatal mortality and homeotic skeletal transformations. *Mol Cell Biol* 26(19): 7201-7210.
57. Ciuffi A, Diamond T, Hwang Y, Marshall H, Bushman FD. (2006) Fusions of *LEDGF/p75* to lambda repressor promote HIV DNA integration near lambda operators in vitro. *Human Gene Therapy* 17: 1-8.
58. Kafri T, van Praag H, Ouyang L, Gage FH, Verma IM. (1999) A packaging cell line for lentiviral vectors. *J. Virol.* 73: 576-584.
59. Vanegas M, Llano M, Delgado S, Thompson D, Peretz 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.
60. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. (2005) Biochemical and genetic analysis of integrase-interacting protein 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.
61. V M, Greenwald J, Rhodes D, Hwang Y, Kwiatkowski W, et al. (2001) Identification of a small molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr. D. Biol. Crystallogr.* 57: 536-544.
62. Yamashita M, Emerman M. (2006) Retroviral infection of non-dividing cells: Old and new perspectives. *Virology* 344(1): 88-93.
63. Arhel N, Genovesio A, Kim KA, Miko S, Perret E, et al. (2006) Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. *Nat Methods* 3(10): 817-824.
64. Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, et al. (2007) Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc Natl Acad Sci U S A.* 2007 [Epub ahead of print]
65. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, et al. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107(3):323-37.

F. Appendix*Fig. 1. Schematic representation of LEDGF domain structure.***A. p75**

← Chromatin-binding domain (CBD) →

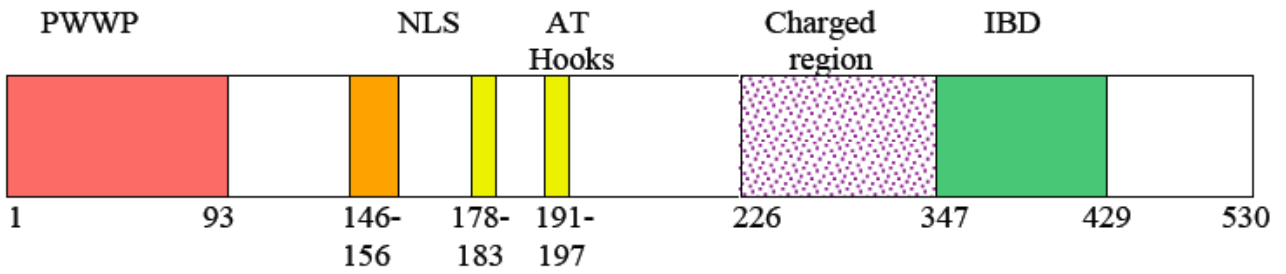
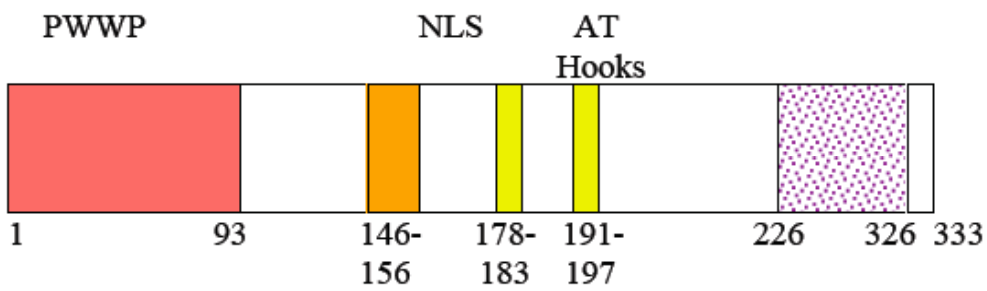
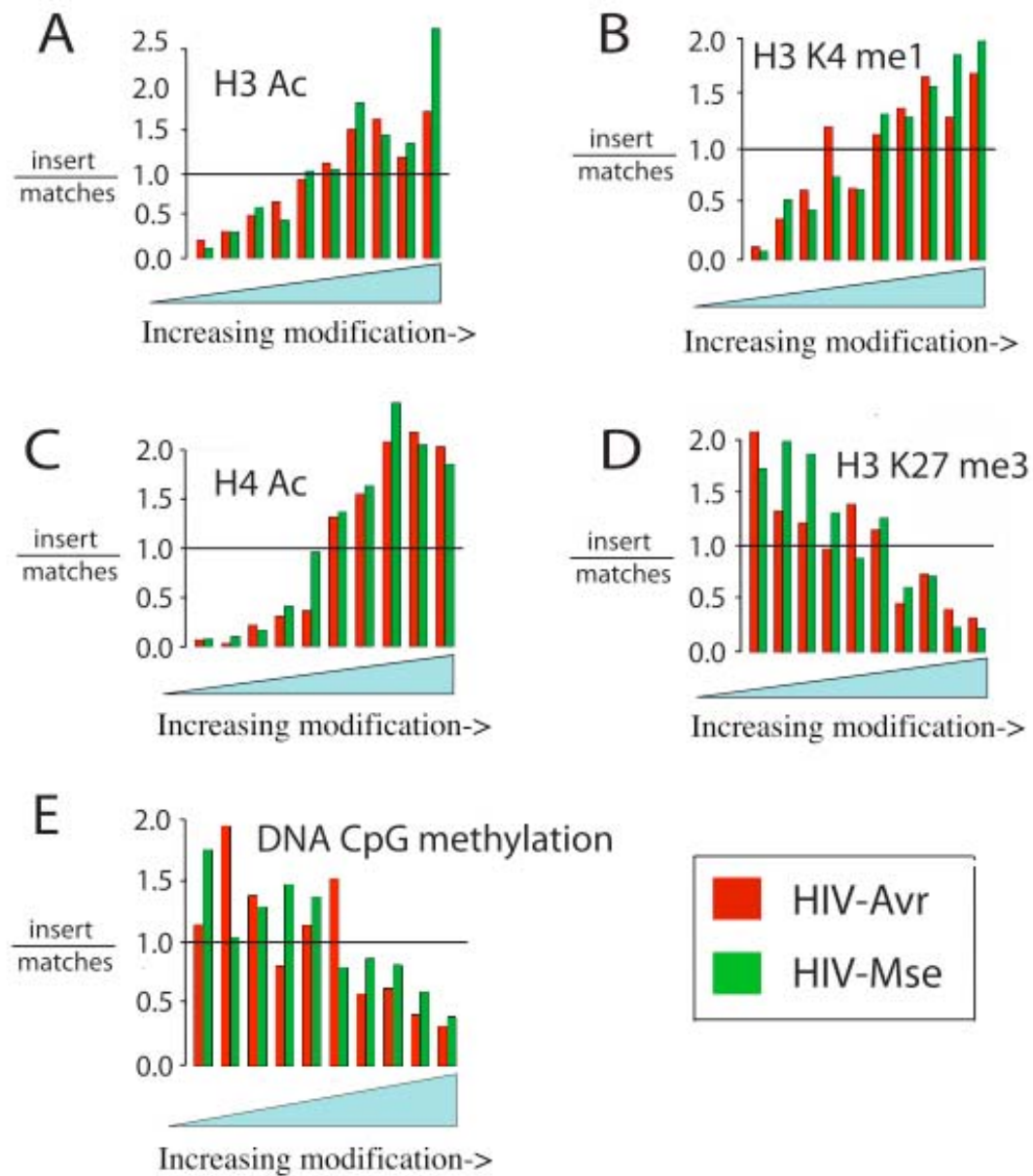
**B. p52**

Fig. 2. HIV integration correlates with epigenetic marks.

Jurkat T-cells were infected with HIV and integration sites isolated by linker-mediated PCR using MseI (green) or a cocktail of AvrII/NheI/SpeI (red) to fragment the genomic DNA (see Aim 1b). Amplicons were sequenced by pyrosequencing. The level of each epigenetic modification, as reported in ENCODE tracks analyzing GM06990 cells (lymphoid lineage), was divided into ten bins and the enrichment of experimental integration sites ('insert') over matched random control ('matches') determined for each bin.



(Gary Wang, Wang *et al.* in press)

Fig. 3. HIV integration in murine cells shows a preference for genes and this is reduced by knockout of Suv39h histone methyltransferase.

Wild-type (W8) and Suv39h1/h2 double knockout (D5) murine embryonic fibroblasts were infected with HIV and integration sites isolated and pyrosequenced as described in Aim 1b. Fold enrichment of experimental integration sites ('inserts') over matched random controls ('matches') inside and outside of RefSeq gene annotations is shown. Both genotypes show significant preference for integration in genes over random ($p < 2e-16$ according to a likelihood ratio statistic test implemented in the Berry regression model, see reference 55). Preference is significantly reduced in the knockout ($p = 1.9e-7$).

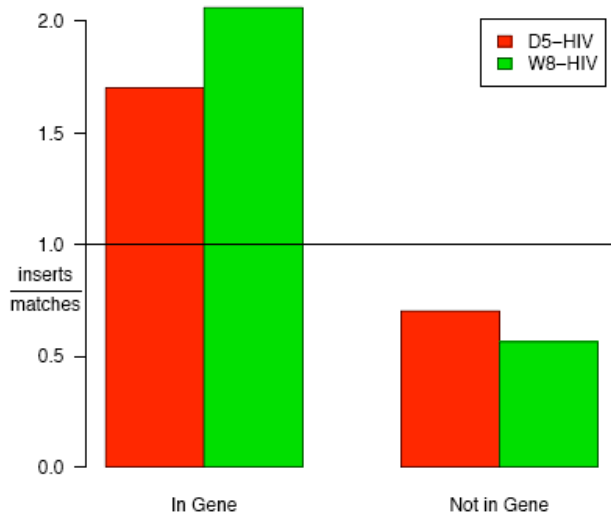
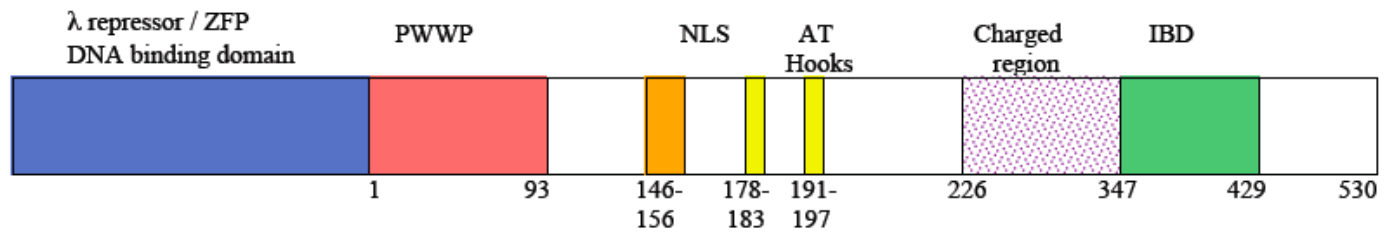
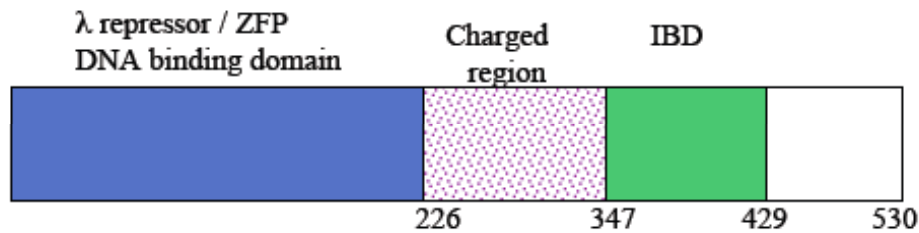


Fig. 4. Proposed tethering fusion constructs (see Aim 1c)

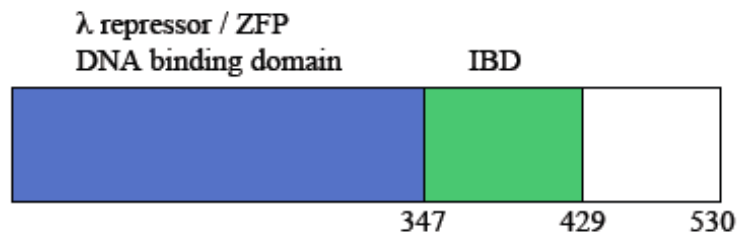
A. Full-length LEDGF/p75



B. Δ-CBD



C. CTD



D. IBD

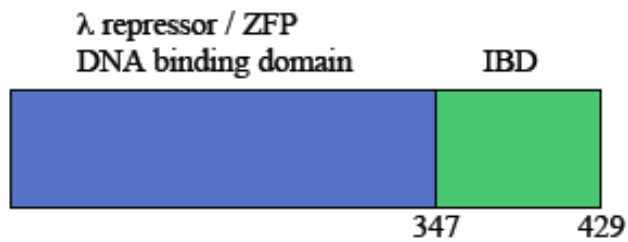
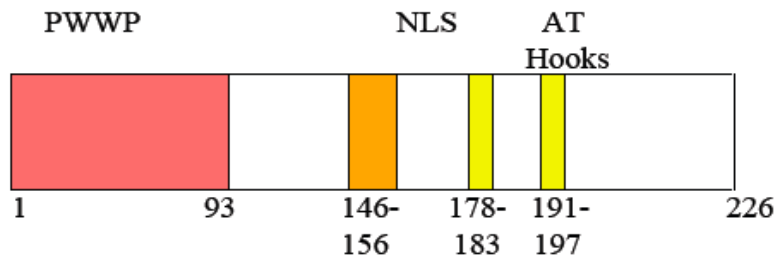
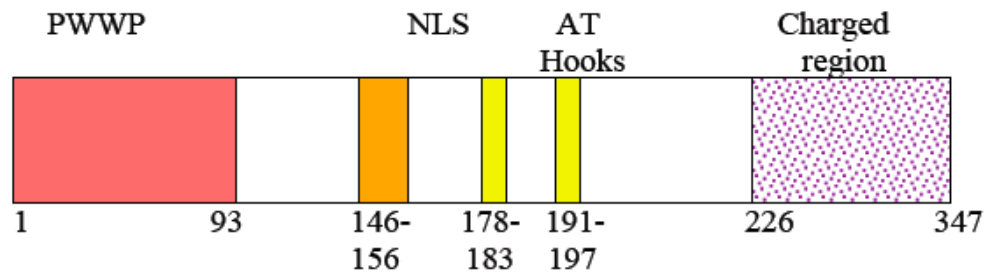


Fig. 5. Proposed truncations (see Aim 2c)

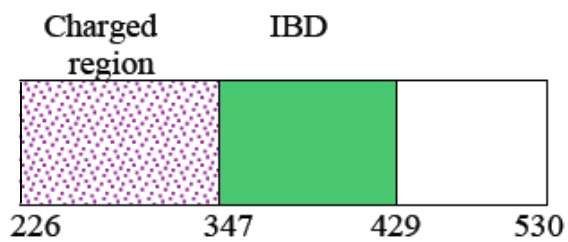
A. Chromatin-binding domain (CBD)



B. Δ -IBD



C. Δ -CBD



D. CTD

