Integration, catalyzed by the viral integrase protein, is an essential step in the replication cycles of all retroviruses. Integration into cellular chromatin provides an optimal environment for gene expression and ensures that the viral genetic material is inherited by daughter cells upon division. Integration proceeds via the following steps: (i) integrase binding to the cDNA end regions that are synthesized during reverse transcription; (ii) hydrolysis adjacent to invariant CA sequences near both 3' ends (3' processing); (iii) transfer of the reactive 3'-OH ends to the 5'-phosphates of a double-stranded cut in cellular chromatin (DNA strand transfer); and (iv) repair of the resulting DNA recombination intermediate, which is likely accomplished by host cell enzymes. (See reference 61 for a detailed overview of retroviral integration.)

Although all retroviruses rely on integrase 3' processing and DNA strand transfer activities, significant differences exist in the way the various viral genera select their chromosomal integration sites. These differences manifest themselves at the level of local DNA sequence (20, 68) and genetic structure (reviewed in reference 2). Lentiviruses, for example, favor integration into active transcription units, targeting genes fairly equally along their lengths (37, 50). Moloney murine leukemia virus, a γ-retrovirus, more modestly favors genes and transcriptional activity but in stark contrast to lentiviruses displays a marked preference for promoter regions and associated CpG islands (19, 37, 67). Simian foamy virus, a spumaretrovirus, slightly disfavors genes though promoters and CpG islands are targeted significantly over random (43, 60). Other profiled genera, including α-, β-, and δ-retroviruses, display less overall preferences for genes, promoter regions, and CpG islands than their lenti-, γ-, and spumaviral cousins (12, 16, 36, 37, 40). Analyses of Moloney murine leukemia virus/HIV-1 chimera viruses revealed that the cognate integrase protein principally determines local DNA sequence and genetic structure specificities during integration (25). Cell factor(s) that may help guide recombination complexes (PICs) to promoter regions for integration are unknown. Recent findings by contrast clarify that lens epithelium-derived growth factor (LEDGF)/p75 is a key lentivirus-specific gene targeting factor (9, 21, 34, 55).

LEDGF/p75 behaves as a bifunctional molecular tether during integration (see references 14 and 46 for recent reviews). An α-helical integrase-binding domain (IBD) (8) that encompasses amino acids 347 to 429 of the 530 residue human protein (6, 65) binds directly to integrase, whereas two regions...
within its N-terminal half, the PWWP domain (residues 1 to 92) and two copies of the AT-hook (ATH) DNA-binding motif, mediate binding to chromatin (29, 62) (Fig. 1A). Accordingly, the expression of integrase- or chromatin binding-defective mutants in cells depleted for endogenous LEDGF/p75 by RNA interference (27) or gene knockout (55) fails to confer sensitivity to human immunodeficiency virus type 1 (HIV-1) infection. The molecular basis for the interaction between LEDGF/p75 and HIV-1 integrase is well understood (3, 5, 8, 13, 21, 49), whereas the mechanism of LEDGF/p75 chromatin binding function is for the most part unknown. Recent results have revealed that the PWWP domain plays a more important role than the AT-hook motifs during HIV-1 infection (55) and chromatin-dependent integration in vitro (1).

The PWWP domain is a ca. 90- to 135-amino acid sequence module present in approximately 60 eukaryotic proteins, many of which, like LEDGF/p75, display an affinity for chromatin (39, 58). Based on amino acid sequence and three-dimensional (3D) structural similarities, the PWWP domain is related to a number of other protein interaction domains that include Tudor, Agenet, malignant brain tumor, and Chromo, which together comprise the Tudor domain “Royal Family” or clan

FIG. 1. LEDGF/p75 schematic and PWWP domain sequence alignments. (A) LEDGF/p75 domain organization and previously described mutants used in the present study. Binding of recombinant protein to DNA in vitro is predominantly mediated by the NLS and dual copy of the AT-hook (ATH) DNA-binding motif, with the PWWP domain (gray shade) contributing a minor role (1, 56, 62). Chromatin binding is largely mediated by the PWWP domain and AT-hooks, with charged regions (CRs) 1 to 3 making secondary contributions (29, 62). MutL1 harbors 12 missense mutations: 6 in the NLS and 3 in each copy of the AT-hook (1, 62). (B) Amino acid alignment of the N-terminal PWWP domains from select human HRP family members. Secondary structural elements elucidated from the solution structure of the HDGF domain (31) are shown above the alignment. Residues targeted for mutagenesis are noted below by triangles; numbers refer to LEDGF/p75 positions. The coloring scheme denotes the following levels of sequence conservation: black, amino acid identity at three or more positions; dark gray, amino acid similarity at all four positions; light gray, chemically similar side chains at three of four positions. (C) Alignment of DNMT3B and LEDGF/p75 PWWP domains. Secondary elements elucidated from the X-ray crystal structure of the murine DNMT3B domain (47) are shown above the corresponding human sequence, whereas HDGF structural elements are drawn below the LEDGF/p75 sequence. The substitution of Pro for Ser-270 (boxed) in DNMT3B is associated with ICF syndrome (53). Black denotes identical amino acids, whereas gray indicates homologous side chains as defined by the following chemical groupings (51): Gly, Ala, Ser, Thr, and Pro; Val, Leu, and Ile; Phe, Tyr, and Trp; Asp, Glu, Asn, and Gln; Lys, Arg, and His; Cys. PR, proline-rich.
(17, 35). Some of these modules, for example, the heterochromatin protein 1 (HP1) chromodomain (23, 41), interact specifically with posttranslationally modified histone tails. The LEDGF/p75 domain displays affinity for humanized chromatin templates in vitro (1), and the homologous zebra fish Brp1 domain was recently shown to bind unmodified core histones H2A and H2B (24). To investigate the mechanism of LEDGF/p75 function during integration, we have delineated PWWP domain amino acid residues critical for chromatin binding and HIV-1 infection.

MATERIALS AND METHODS

Plasmids. Animal cell LEDGF/p75 expression vectors were based in pIREs2-eGFP, and proteins were expressed untagged or as fusions to a C-terminal hemagglutinin (HA) tag as described previously (55). Bacterial expression plasmids were based in pET-1-LEDGF (63), pET-1-MutL1 (carrying six missense mutations in the LEDGF/p75 nuclear localization signal [NLS] and three in each mids) were generated by mating E6(f/f)/H11001 with E5(f/f)/H11002 and E6(f/f)/H11001 with E5(f/f)/H11002, respectively, were prepared from embryos following two rounds of mouse mating that began with f/f and Sox2Cre animals as described previously (55). This same approach was repeated with the E17(f/f)/H11001 and E17(f/f)/H11003 littermates matched E17(f/f)/H9262 and E17(f/f)/H9252, respectively, were prepared from embryos following two rounds of mouse mating that began with f/f and Sox2Cre animals as described previously (55). A variety of control and LEDGF/p75 knockout MEFs were used in the course of these studies. The majority of infectivity measurements were conducted with E6(+/−) knockout cells derived from E6(+/−) control MEFs ex vivo via bacteriophage P1 Cre protein expression (55). E6(+/+) and E6(−/−) littermate control and knockout MEFs, respectively, were prepared from embryos following two rounds of mouse mating that began with f/f and Sox2Cre animals as described previously (55). This same mating scheme was used to generate the following control and knockout cell sets from two independent pregnancies: E4(+/+) and E4(−/−); E3(+/+) and E3(−/−), and E6(+/−). Heterozygous (+/−) mice lacking Cre were generated by mating Sox2Cre+/− to C57BL/6 animals; littermate matched E17(+/−) mice were analyzed by Western blotting with anti-HA 3F10 antibodies. The results were quantified by using a FluorChem FC2 imager (Alpha Innotech Corp., San Leandro, CA).

Expression and purification of recombinant proteins. Wild-type and full-length LEDGF/p75 missense mutant proteins expressed from pFRT-1-LEDGF were purified from soluble extracts of Escherichia coli cells as described previously (63). Glutathione S-transferase (GST) fusions to the wild-type or mutated PWWP domains were expressed from pGEX-4T-1-LEDGF(D1001) and purified from soluble E. coli extracts essentially as previously described (1). In brief, protein eluted from glutathione-Sepharose beads (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) was further purified by using a Superdex 75 column (GE Healthcare). Two of the mutations, W21A and A51P, negatively impacted protein purity and yield. Fractionation of GST-PWWP/W21A and GST-PWWP/A51P on Hi-Trap heparin columns (GE Healthcare) prior to gel filtration chromatography increased the purity of these final preparations to the same level (>90% as assessed by Coomassie blue staining) as the other GST-PWWP proteins. His-tagged HIV-1 integrase was expressed and purified from a soluble E. coli extract as described previously (1).

In vitro integration and chromatin and DNA binding assays. LEDGF/p75-dependent integration assays using naked target DNA or matched chromatinized templates were performed as described previously (1). Reaction product formation was quantified by real-time PCR as described therein. Radiolabeled DNA and polynucleosome (PN) binding to wild-type and mutant GST-PWWP domain fusion proteins were performed as described previously (1).

RESULTS

Mutagenesis strategy. LEDGF/p75, expressed from the PSIP1 gene, is a member of the hematopoietic-derived growth factor (HDGF)-related protein (HRP) family that also contains HDGF, HRP1, HRP2, HRP3, and LEDGF/p52 (an alternative mRNA splice variant of PSIP1). The PWWP domain is the most conserved region among these proteins (22) and sequence alignments of protein orthologs (6, 39, 58) and paralogs (22) (Fig. 1B), as well as the 3D structures of the HRP3 (39) and HDGF (31) (Fig. 2) domains, were consulted for mutagenesis. The heart of the PWWP domain is a five-stranded antiparallel β barrel with a 310 helix preceding the fifth strand (Fig. 1B and 2), followed by various numbers of α helices (31, 39, 47, 57). The barrel presents a solvent exposed hydrophobic cavity that is the discerning structural feature of Tudor clan members (35).
Lys-14, Trp-21, Pro-22, and Phe-45 are the most conserved residues among PWWP domain orthologs (39). Because Pro-22 and Phe-45 are both solvent inaccessible (31, 39), they were not targeted. In addition to Trp-21, LEDGF/p75 residues Met-15 and Tyr-18 from within the β-α arch, Phe-43 from β3, Thr-47 from the β3-β4 loop, and Glu-49 and Ala-51 from β4 (Fig. 1B) line the putative ligand-binding cavity (31, 39) (Fig. 2A). Each of these residues, with the exception of Ala-51, was changed to Ala and/or Glu to test their role in LEDGF/p75 PWWP domain function. The substitution of Ser-270 by Pro within the PWWP domain of DNA methyltransferase (DNMT) 3B is associated with immunodeficiency, centromeric instability, and facial anomaly (ICF) syndrome in humans (53). Since Ser-270 in DNMT3B is analogous to Ala-51 in LEDGF/p75 (boxed in Fig. 1C), Pro was tested in place of Ala-51.

The binding of recombinant LEDGF/p75 protein to DNA in vitro is predominantly mediated by conserved Arg and Lys residues within the NLS and AT-hook motifs (Fig. 1A), with a more minor contribution from the PWWP domain (1, 56, 62). HDGF also binds DNA in vitro (31, 69) but, unlike LEDGF/p75, this protein lacks discernible AT-hook DNA-binding motifs (38). Instead, a positively charged face of the HDGF PWWP domain, comprised of LEDGF/p75 analogous residues Arg-3, Lys-14, Lys-16, Lys-56, Lys-67, Lys-70, Lys-73, Arg-74, and Lys-75 (Fig. 2B), has been implicated in DNA binding (31). These residues were therefore targeted individually or in groups to assess the role of the presumed positively charged LEDGF/p75 face in HIV-1 infection and integration. Other solvent-exposed residues, including Leu-10, Ile-11, Pro-19, His-20, Val-28, Leu-40, Pro-41, and Ile-42, were targeted due to their relative degrees of sequence conservation or proximities to the presumed DNA binding face (Fig. 2B) or hydrophobic cavity (Fig. 2C). Table 1 lists the 24 targeted residues, the reasons for their selection, and resulting 37 novel mutant proteins that were tested for their abilities to support HIV-1 function.

**Virus infectivity.** LEDGF/p75 plays a critical role in mediating lentiviral integration and hence virus infection (11, 21, 27, 34, 55, 64). We previously generated knockout and matched control LEDGF/p75-expressing mouse cell lines to analyze the role of the host factor in HIV-1 infection. MEFs were infected with VSV-G-pseudotyped single-round HIV-Luc carrying either wild-type or D64N/D116N active-site mutant integrase. At 2 days postinfection, cells were processed for Luc activity, and integrase-dependent levels of HIV-1 infectivity were determined by subtracting the low levels of active site mutant activities from matched wild-type viral infections (55). Accordingly, HIV-Luc infected E6(−/−) knockout cells at 3.4% ± 0.0% (n = 7) of the level of control E6(+/+) cells. These results moreover defined a valuable reverse genetic system, as transiently expressed LEDGF/p75 protein sensitized E6(−/−) cells to HIV-1 infection in an IBD- and PWWP domain-dependent manner (55). Each of the novel PWWP domain mutant proteins was therefore tested for its ability to support E6(−/−) cell infection alongside wild-type LEDGF/p75 and two previously described N-terminal deletion mutants: ΔPWWP, lacking the PWWP domain and ΔPWWPΔATh, which lacked the dual copy of the AT-hook DNA-binding motif in addition to the PWWP domain (Fig. 1A).

As previously established (55), the double ΔPWWPΔATh mutant failed to stimulate the basal level at which E6(−/−) cells became infected, whereas the ΔPWWP deletion mutant supported ca. 19% of the level of wild-type LEDGF/p75 function (Fig. 3A). A perusal of the novel PWWP domain mutant proteins (Fig. 3 and Table 1) revealed a full activity spectrum that spanned from the wild-type level (for example, V28E) to undetectable (for example, K14A/K16A or W21A). Consistent with its high degree of sequence conservation among PWWP domain orthologs (39, 58), Lys-14 appeared the most important of the targeted basic residues. K14A and K14E functioned at ca. 58 and 17% of the level of wild-type LEDGF/p75, respectively, while K14A/K16A was unable to support HIV-1

**FIG. 2.** 3D representations of targeted LEDGF/p75 residues. (A) The solution structure of the HDGF PWWP domain (Protein Data Base accession code 2B8A) is shown highlighting the residues that form its hydrophobic cavity (31, 39). (B) Cluster of basic Lys and Arg residues implicated in HDGF DNA binding (31). His-20 of the PWWP motif locates in this general vicinity; the extended N-terminal tail that harbors Arg-3 is not shown. (C) Remaining targeted residues. Ile-11, Pro-19, Val-28, Leu-40, Pro-41, and Ile-42 contribute to the hydrophobicity of the central cavity. Leu-10, Ile-11, Val-28, and Ile-42 are conserved among Tudor clan members (35), whereas Pro-19 is conserved among PWWP domain orthologs (39, 58). Note that Ile-11, Val-28, Leu-40, and Pro-41 are Val, Met, Tyr, and Gln in HDGF, respectively (Fig. 1B); the mutagenesis option is not shown. (C) Remaining targeted residues. Ile-11, Pro-19, Val-28, Leu-40, Pro-41, and Ile-42 contribute to the hydrophobicity of the central cavity. Leu-10, Ile-11, Val-28, and Ile-42 are conserved among Tudor clan members (35), whereas Pro-19 is conserved among PWWP domain orthologs (39, 58). Note that Ile-11, Val-28, Leu-40, and Pro-41 are Val, Met, Tyr, and Gln in HDGF, respectively (Fig. 1B); the mutagenesis option is not shown. (C) Remaining targeted residues. Ile-11, Pro-19, Val-28, Leu-40, Pro-41, and Ile-42 contribute to the hydrophobicity of the central cavity. Leu-10, Ile-11, Val-28, and Ile-42 are conserved among Tudor clan members (35), whereas Pro-19 is conserved among PWWP domain orthologs (39, 58). Note that Ile-11, Val-28, Leu-40, and Pro-41 are Val, Met, Tyr, and Gln in HDGF, respectively (Fig. 1B); the mutagenesis option is not shown. (C) Remaining targeted residues. Ile-11, Pro-19, Val-28, Leu-40, Pro-41, and Ile-42 contribute to the hydrophobicity of the central cavity. Leu-10, Ile-11, Val-28, and Ile-42 are conserved among Tudor clan members (35), whereas Pro-19 is conserved among PWWP domain orthologs (39, 58). Note that Ile-11, Val-28, Leu-40, and Pro-41 are Val, Met, Tyr, and Gln in HDGF, respectively (Fig. 1B); the mutagenesis option is not shown.
infection despite efficient expression of the double mutant protein (Fig. 3A). The only other inactive charge-to-alanine mutant, 5K/R>A, harbored six overall amino acid changes (Fig. 3A and Table 1).

Of the residues predicted to comprise the hydrophobic cavity (Fig. 2A), Trp-21 and Ala-51 were the most critical with important roles determined for Met-15, Thr-47, and Glu-49 (Fig. 3B and Table 1). Ile-42 is near the cavity (Fig. 2C), and changing it in concert with Phe-43 also ablated cofactor function (Fig. 3B). Because Y18A conferred the wild-type level of HIV-1 infection, the phenolic side chain was dispensable under these assay conditions. Surprisingly, the activity of the double mutant T47E/E49A protein exceeded the level of either single point mutant (Fig. 3B and Table 1).

The His within the signature PHWP motif appeared unimportant, as H20A and H20E mutants functioned at ca. 95 and 81% of the level of wild-type LEDGF/p75, respectively (Table 1). The first Pro within the motif was also not critical, as P19A functioned at the wild-type level; substituting Glu for Pro-19 though reduced cofactor function approximately fivefold (Fig. 3C and Table 1). The combined L10A/I11A mutant functioned at ca. 8% of wild-type LEDGF/p75, similar to the residual level of activity displayed by the double L40E/P41E mutant protein (Fig. 3C and Table 1).

The observation that a number of point mutant proteins functioned at levels that were significantly less than that observed for the ΔPWWP deletion, which lacked the domain, was unexpected. Because of this, ΔPWWP was tested alongside wild-type LEDGF/p75 and the empty expression vector in expanded sets of knockout and control MEFs. The ΔPWWP mutant functioned at 20.6 ± 1.7% of wild-type in E16(−/−) cells, similar to the level observed in E6(−/−) cells (55) (Fig. 3A), whereas about half this level (11.5% ± 0.6%) was seen using E5(−/−) cells. In contrast, the mutant displayed only residual function (<0.1 to 4.2%) in three other knockout cell lines. A subset of mutant proteins, chosen because they displayed the gamut of activities in E6(−/−) cells, was therefore retested using E2(−/−) cells where the ΔPWWP mutant functioned at 1.7% ± 2.2% (n = 4) of wild-type LEDGF/p75 (Fig. 4A and B). The ΔATH deletion mutant (Fig. 1A), which functioned similar to wild-type LEDGF/p75 in E6(−/−) cells (55), importantly maintained this phenotype in E2(−/−) cells (Fig. 4A and 4B). Levels of K14E, K14A, K14A, and K16A mutant protein function, which ranged from 17 to 91% of wild-type LEDGF/p75 in E6(−/−) cells (Fig. 3A and Table 1), spanned 8 to 82% when tested in E2(−/−) cells (Fig. 4A). The relative levels of K14E, K14A, K16E, and K14A/K16A mutant protein function in E6(−/−) cells were furthermore maintained in E2(−/−) cells (compare Fig. 3A and 4A). Due to these results, additional point mutants were not tested in E2(−/−) cells. Of note, the expression level of the ΔPWWP deletion protein in the two knockout cell types did not account for its differential function during infection, since it was expressed just as well if not somewhat better in E2(−/−) versus E6(−/−) cells (Fig. 4C, compare lane 2 to lane 6).

Association with cellular chromatin. The PWWP domain, which is present in approximately 60 eukaryotic proteins, plays a central role in LEDGF/p75 chromatin binding (29, 62). We therefore determined the binding properties of a subset of the novel point mutant proteins; K14A/K16A, W21A, I42A/F43A, and A51P were highlighted because each of these failed to sensitize mouse knockout cells to HIV-1 infection despite containing only one or two amino acid changes. LEDGF/p75 interacts intimately with the condensed chromosomes that form during mitosis (7, 29, 33, 42, 44, 62, 65), so the distribution of each mutant in interphase and mitotic cells was compared to the wild type. To maintain consistency with previous mutational studies (29, 33, 62, 65), human cells were used for these analyses.

We previously visualized green fluorescent protein (GFP)-LEDGF/p75 fusion proteins in live HeLa cells by using confo-
To increase the sensitivity of mutant protein detection, we more recently turned to immunodetection of HA-tagged variants in fixed cells using epifluorescence microscopy. To gauge this approach, we first documented the phenotypes of a number of previously analyzed mutant proteins. HeLa TZM-bl cells sorted to enrich for transient transfectants were treated with anti-HA antibodies at 2 days posttransfection, and these results were compared to DAPI-stained images. As determined for the analogous GFP fusion protein (62), deleting the PWWP domain did not significantly alter the localization of the mutant proteins.

FIG. 3. Mutant protein activities and expression profiles. (A) The indicated untagged mutant proteins were tested for their abilities to sensitize E6(f/f) knockout cells to HIV-1 infection in comparison to wild-type LEDGF/p75 and the empty pIRES2-eGFP expression vector. Protein expression levels are indicated to the right. Lane 1, lysate prepared from cells transfected with empty vector DNA. (B) Additional mutants from the experiment shown in panel A. (C) Same as in panels A and B except that the noted mutant and control proteins were analyzed in a separate Western blotting experiment. Infectivity data minimally compile the results of eight Luc assays (duplicate assays from two to four infections, each conducted in duplicate). In general, only mutants that displayed less than half of the level of wild-type LEDGF/p75 activity are shown (see Table 1 for the complete data set). The levels of endogenous LEDGF/p75 protein in E6(f/f) cells, shown in Fig. 4C below, were below the detection limit of the panel A and C experiments.
FIG. 4. LEDGF/p75 mutant protein activities and expression profiles in E2(−/−) knockout cells. (A) The indicated mutants were tested in comparison to wild-type LEDGF/p75 and empty pIRRES2-eGFP vector DNA; data are averages and standard deviations of minimally eight Luc assays from two or more independent transfection and infection experiments. (B) Expression levels of proteins analyzed in E2(−/−) and E6(−/−) cells, respectively. HIV-Luc infected parental E1(+/+) and E2(−/−) cells, respectively; HIV-Luc infected parental E1(+/+) and E2(−/−) cells, respectively. Lanes 3 and 4, endogenous mouse LEDGF/p75 levels in E2(−/−) and E6(−/−) cells, respectively. Lanes 5 to 8, endogenous protein levels in transfected E2(−/−) and E6(−/−) cells, respectively.

Calization of HA-tagged LEDGF/p75 during interphase or mitosis (Fig. 5A). Deleting the dual copy of the AT-hook motif marginally affected the chromosome binding capacity of LEDGF/p75, whereas the combined ΔPWWPΔATh deletion mutant, as expected (29, 62), failed to engage chromatin (Fig. 5A, compare rightward merged images). Due to the functional NLS at residues 146 to 152, the double mutant nevertheless was strictly nuclear during interphase (Fig. 5A, left center images). We previously reported GFP-MutL1 (Fig. 1A) as pan cellular in interphase and mitotic cells, concluding that the mutant lost chromatin binding (62). In contrast, the MutL1-HA construct retained partial chromatin binding activity: it was predominately nuclear in interphase cells despite the critical K150A NLS mutation (32) (Fig. 5A, left), indicating that the mutant functionally engaged chromatin during reformation of postmitotic nuclei to acquire its karyophilic phenotype in an NLS-independent manner (65). Accordingly, MutL1-HA displayed affinity for condensed mitotic chromatin, although this appeared somewhat intermediary compared to the wild-type and defective ΔPWWPΔATh-HA mutant phenotypes (Fig. 5A, rightward merged images). Because MutL1 fully sensitized mouse knockout cells to infection, we conclude that direct DNA binding as mediated by the NLS and AT-hooks is dispensable for LEDGF/p75-dependent HIV-1 integration ex vivo (Table 1), as well as in vitro (1).

Each of the novel PWWP domain point mutant proteins predictably retained the karyophilic behavior of wild-type LEDGF/p75 in interphase cells (Fig. 5B, left sets of images). Significantly, each of these mutants failed to effectively engage condensed mitotic chromatin (Fig. 5B, right image sets). The results of Western blotting revealed that each mutant protein was expressed at a level that equaled or exceeded that of wild-type LEDGF/p75 (Fig. 5C). The interphase and mitotic distributions of the 5K/R−A mutant, which also failed to sensitize mouse knockout cells to HIV-Luc infection (Fig. 3A), was indistinguishable from other defective PWWP domain point mutant proteins (data not shown).

Chromatin association was also assessed following biochemical fractionation of transiently transfected human 293T or mouse knockout cells essentially as previously described (29, 59). Isolated nuclei (fraction P1 in Fig. 6A) were subsequently treated with DNase and then high salt (250 mM NH2SO4) to extract LEDGF/p75 from insoluble cytoskeletal and nuclear matrix materials. Accordingly, the chromatin-bound fraction is defined as the percentage of total LEDGF/p75 protein that partitions to fractions P1 and S2. As expected (29, 59), the vast majority of wild-type LEDGF/p75 (96% ± 1% for n = 4 experiments) was chromatin bound under these conditions (Fig. 6B, lanes 1 to 4). Deleting the PWWP domain marginally impaired association with chromatin (89% bound in Fig. 6B, lanes 5 to 8; 77% ± 7% for n = 4 experiments), whereas approximately half of the ΔPWWPΔATh mutant protein (51% ± 4% for n = 4) was unbound since it partitioned to the initial S1 fraction (Fig. 6B, lanes 9 to 12). The PWWP point mutant proteins behaved similar to the ΔPWWP deletion under these conditions: 76% ± 4% (n = 3) of the W21A mutant partitioned to fractions P1 and S2 (Fig. 6C, lanes 1 to 4), whereas 85% ± 7% (n = 3) of the I42A/F43A mutant remained chromatin bound (Fig. 6C, lanes 5 to 8). Since the ΔATh deletion significantly affected the fractionation phenotype of the ΔPWWP mutant (panel B), the AT-hook motifs were similarly removed from PWWP domain point mutant proteins. In these cases, though, the deletion effected at best a modest influence: 79% ± 5% and 63% ± 2% (n = 2) of the W21A-ΔATh and I42A/F43A-ΔATh mutants, respectively,
partitioned to the P1 and S2 fractions (Fig. 6D, lanes 1 to 8). Similar results were obtained using K14A/K16A, A51P, K14A/K16A/ΔATh, and A51P-ΔATh mutant proteins (data not shown).

**In vitro integration and DNA/chromatin-binding activities.** Purified LEDGF/p75 protein potently stimulates the activities of recombinant lentiviral integrase proteins (4, 6, 7, 62), and the nature of the in vitro reaction conditions can influence the requirements for the different host factor functions. Direct binding to HIV-1 integrase, for example, is crucial under all conditions (6, 8, 45, 48, 62, 71), whereas N-terminal deletion mutants of LEDGF/p75 defective for chromatin binding retained about half the level of wild-type activity when integration was performed with naked target DNA (1, 62). Recombinant ΔPWWP/MutL1 protein by contrast functioned at only ca. 4% of the level of the wild-type when reconstituted PNs were used in place of naked DNA (1). Recombinant K14A/K16A, W21A, I42A/F43A, and A51P proteins, with or without added MutL1 changes, were purified following their expression in *E. coli* to assess the effects of PWWP domain point mutations on the ability for LEDGF/p75 to stimulate HIV-1 integrase activity in vitro. Three single amino acid changes that effected intermediary phenotypes in the infectivity assay, K14A, I42A, and F43A, were also analyzed.

As previously noted (1), integrase favored PNs ~10-fold over naked DNA in the absence of added LEDGF/p75 protein (Fig. 7A, inset). Maximal levels of PN-dependent stimulation occurred at 500 nM LEDGF/p75, whereas twice as much protein was required to max out integration into naked DNA (Fig. 7A). Because of this, mutant protein activities were compared to the wild-type at 500 nM and 1 μM LEDGF/p75 when using chromatinized and naked target DNAs, respectively.

The mutants functioned at ca. 16% (K14A/K16A/MutL1) to 76% (K14A and I42A) of the level of wild-type LEDGF/p75 for stimulating integration into naked target DNA (Fig. 7B, black bars). Of those that failed to sensitize mouse knockout cells to HIV-1 infection (first four entries in Fig. 7B), W21A supported the most integration (51% of the wild type), whereas K14A/K16A and I42A/F43A were the least active at ca. 23%. The activity of each mutant protein relative to wild-type was moreover reduced when PNs were used in place of naked target DNA (Fig. 7B, compare gray bars to black bars; summarized in panel C as the percent naked DNA stimulatory activity). These changes were most evident for W21A and K14A/K16A. W21A and W21A/MutL1 stimulated PN-dependent integration at ca. 30 and 18% of the levels seen with naked DNA, respectively, whereas the K14A/K16A and K14A/K16A/MutL1 proteins failed to detectably influence integrase activity under these assay conditions (Fig. 7B and C). Titrations revealed that chromatin-binding defective mutants supported more integrase activity at a higher (1.0 or 1.5 μM) LEDGF/p75 concentration (data not shown). These values, however, were invariably less than the suboptimal levels of wild-type activity observed under these conditions (Fig. 7A).

NLS and AT-hook residues primarily mediate the binding of LEDGF/p75 to DNA in vitro (1, 62). The isolated PWWP domain displays affinity for PNs and DNA as well, although this latter activity is comparatively weak due to salt hypersensitivity (1). To ascertain how the different mutations affect PWWP domain-specific chromatin and DNA binding, purified GST fusion proteins harboring the N-terminal 100 amino acid residues of LEDGF/p75 were prebound to glutathione-Sepharose beads and then exposed to radiolabeled DNA or PNs in the presence of low (50 mM) NaCl. After extensive washing, the levels of input substrate pulled down by the various bait proteins were quantified relative to those recovered by the wild-type GST fusion. Accordingly, neither the I42A nor F43A mutation significantly affected DNA or chromatin binding (Fig. 8A; the results of experimental replicates are quantified in panel B). Three of the four mutations that ablated function in the virus infection assay—K14A/K16A,
W21A, and I42A/F43A—by contrast dramatically reduced binding to both substrates. GST-PWWP/A51P interestingly recovered ca. 75% of the wild-type level of naked DNA yet only ca. 20% of the input chromatin substrate. In contrast, the K14A mutation significantly reduced binding to both substrates, with DNA capture somewhat more affected than chromatin binding (Fig. 8).

DISCUSSION

This study confirms and extends the model that LEDGF/p75 acts as a bifunctional molecular tether during HIV-1 integration (27, 55): its C-terminal IBD engages PIC-born integrase, whereas the N-terminal PWWP domain mediates interactions with cellular chromatin. Prior mutagenic (3, 8, 13, 49, 61) and structural biology (5, 8) approaches have significantly clarified the molecular basis of the LEDGF/p75-integrase interaction. Using primary sequence and 3D structural similarities, 24 PWWP domain amino acids were targeted here to elucidate those residues that play important roles during HIV-1 integration (Fig. 1 and 2 and Table 1). LEDGF/p75 PWWP domain amino acid residues critical for HIV-1 infection. Nineteen of the twenty-four targeted residues were analyzed via single amino acid substitution and of the 24 resultant mutant proteins, only W21A, W21E, and A51P failed to sensitize E6(H11002)/H11002 mouse knockout cells to HIV-1 infection (Table 1). These results highlight critical roles for Trp-21 and Ala-51 in mediating lentivirus infectivity. As projected onto the solution structure of the HDGF PWWP domain (31), each of these residues lines the inner surface of the hydrophobic cavity that is the signature structural motif of members of the Tudor clan (Fig. 2A and 9A) (35). Considering that ectopically expressed W21A and A51P proteins failed to engage condensed mitotic chromosomes (Fig. 5B) and bound chromatin less efficiently than wild-type LEDGF/p75 (Fig. 6 and data not.
Because GST-PWWP/K14A was recombinant GST-PWWP protein to DNA and chromatinized 5B) chromatin association in cells, as well as the binding of the LEDGF/p75 molecular tether. These combination mut- 

tively disrupted the interaction of the PWWP domain with chromatin without gross disruption of in vitro DNA binding (Fig. 8), indicating that the structural integrity of the domain was not overly compromised by this mutation.

By extension, we conjecture that the primary defects of the K14A/K16A and I42A/F43A mutant proteins in supporting HIV-1 infection likewise reside at the chromatin-binding side of the LEDGF/p75 molecular tether. These combination mutations reduced (Fig. 6 and data not shown) or ablated (Fig. 5B) chromatin association in cells, as well as the binding of recombinant GST-PWWP protein to DNA and chromatinized templates in vitro (Fig. 8). Because GST-PWWP/K14A was preferentially defective for binding to naked DNA (Fig. 8), it seems that the K14A/K16A chromatin binding defect could be partially due to a loss of DNA-binding function. This interpretation is consistent with a role for the charged outer face comprised of LEDGF/p75 analogous residues Arg-3, Lys-14, Lys-16, Lys-56, Lys-67, Lys-70, Lys-73, Arg-74, and Lys-75 in the in vitro DNA-binding activity of the HDGF PWWP domain (31). It is tempting to speculate that the hydrophobic cavity supplies the dominant binding interface for an unknown chromatin binding partner, whereas the positively charged outer face (Fig. 2B) could increase the overall affinity of the PWWP domain for chromatin for example, through nonspecific contacts with the DNA phosphodiester backbone. Given sufficient disruption of outer domain face charge-mediated interactions, for example, via the six changes in the 5K/R->A mutant, LEDGF/p75 nevertheless loses chromatin association and hence HIV-1 cofactor function.

The ΔPWWP deletion mutant curiously maintained affinity for mitotic chromatin under conditions where one or two amino acid changes within the domain disrupted function (Fig. 5). These data therefore suggest that the presence of a non-functional domain can counteract the abilities for the AT-hook DNA-binding motifs (29, 62) (Fig. 5A) and/or other charged regions within the LEDGF/p75 protein (29) (Fig. 1A) to confer chromatin-binding affinity.

As expected, the K14A/K16A, L10A/I11A, M15E/Y18A, L40E/P41E, and I42A/F43A double-mutant proteins fared less well than their single amino acid constituents at sensitizing mouse knockout cells to HIV-1 infection (Fig. 3 and Table 1). We were therefore somewhat surprised that the T47E/E49A double mutant reproducibly outshone the activities of either T47E or E49A point mutant (Fig. 3B). Thr-47 and Glu-49 participate in hydrophobic cavity formation and moreover abut each other in three dimensions (Fig. 2A and 9A). Our results therefore indicate the requirement for Glu at position 47 or 49, but not both, for effective PWWP domain function during HIV-1 infection. Other Tudor clan members utilize an analogous Glu residue for binding their respective substrates (39). For example, the E134K spinal muscular atrophy mutation in the survival motor neuron (SMN) Tudor domain abrogates binding to spliceosomal Sm protein (52). The co-crystal structure of the *Drosophila* HP1 chromodomain in complex with a histone H3 N-terminal tail peptide moreover revealed that Glu-52 interacted with the dimethylated Lys-9 side chain required for binding (Fig. 9B) (23). The majority (32 of 42) of PWWP domains harbor Glu or Asp at LEDGF/p75 analogous position 47 or 49 (but never both) (58), indicating that a negatively charged side chain in this general vicinity of the hydrophobic cavity is likely relevant for the function of most PWWP domains.

**Conclusions.** We have identified a number of PWWP domain residues, highlighted by Trp-21 and Ala-51, which play critical roles in LEDGF/p75-dependent HIV-1 infection and integration. The results of numerous independent experiments, including sensitization of mouse knockout cells to infection (Fig. 3 and 4), association of ectopically expressed protein with chromatin (Fig. 5 and 6), PN-dependent stimulation of integrase activity in vitro (Fig. 7), and PN pull-down assays (Fig. 8), combine to support a model for the PWWP domain-binding function of LEDGF/p75. In this model, the hydrophobic cavity is likely relevant for the function of most PWWP domains.
domain hydrophobic cavity as a crucial chromatin interaction motif (Fig. 9A).

Point mutations in the PWWP (Fig. 3 and Table 1) or IBD (55) can abrogate LEDGF/p75 function during HIV-1 infection. Various PWWP point mutant proteins (for example, W21A or K14A/K16A) were coexpressed with the functionally inactive D366N IBD mutant in knockout cells to test for functional complementation via phenotypic mixing. This approach, however, failed to yield HIV-Luc infectivities beyond those observed with cells expressing sole mutant proteins. Although admittedly negative in nature, these results indicate that LEDGF/p75 may very well function as a monomer during HIV-1 infection. Consistent with this interpretation, purified LEDGF/p75 protein sedimented as a monomer during analytical ultracentrifugation (8).

It is anticipated that the distribution of lentiviral integration is in large part defined by the chromosomal distribution of LEDGF/p75 protein. Lentiviruses may very well utilize a LEDGF/p75 independent pathway to accomplish ca. 2 to 20% of their overall integrations (27, 34, 55), but we speculate that the lion’s share occurs through the PIC engaging chromatin-bound LEDGF/p75 via the IBD. Integrase is then encouraged of their overall integrations (27, 34, 55), but we speculate that is in large part defined by the chromosomal distribution of HIV-1 infection. Consistent with this interpretation, purified LEDGF/p75 protein sedimented as a monomer during analytical ultracentrifugation (8).

ACKNOWLEDGMENTS

HeLa TZM-bl cells were obtained through the NIH AIDS Research and Reference Reagent Program from John Kappes, Xiaoyun Wu, and Tranzyme, Inc. We thank Manuel Llano for advice with the fractionation technique used for chromatin binding analyses, Peter Cherepanov and Lavanya Krishnan for their comments on the manuscript, and L. Krishnan for advice with figure preparation. This study was supported by grants from the National Institutes of Health (NIH) (AI39394 and AI70042 [A.E.], AI45587 [J.L.], and AI60354 [Harvard Medical School Center for AIDS Research]) and the Agence Nationale de Recherche sur le SIDA (2005/004 and 2006/124 [M.L.]). N.Y. was supported by a HMS CFAR Scholar Award, and M.-C.S. was supported by NIH Training Grant AI07245.

REFERENCES


4. Cherepanov, P. 2007. LEDGF/p75 interacts with divergent lentiviral inte-
DENVIP®75 PWWP DOMAIN AND HIV-1 INTEGRATION 11567


