Role of the PWWP Domain of Lens Epithelium-derived Growth Factor (LEDGF)/p75 Cofactor in Lentiviral Integration Targeting*§

Rik Gijsbers†, Sofie Vets‡, Jan De Rijck†, Karen E. Ocwieja*, Keshet Ronen†, Nirav Malani†, Frederic D. Bushman*, and Zeger Debyser†,‡

From the †Division of Molecular Medicine, KU Leuven, 3000 Leuven, Belgium and ‡Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

HIV LEDGF/p75 is a chromatin-interacting, cellular cofactor of essential to direct integration in the host cell chromatin to sites infected with virus; mRFP, monomeric red fluorescent protein; MRC, matched random control; ROC, receiver operating characteristic; Fluc, firefly luciferase; ΔN96 deletion of the first 93 amino acids; BC, back-complementation; eGFP, enhanced green fluorescent protein.

Stable integration of the viral DNA into the host genome is one of the hallmarks of retroviral replication and has profound consequences for both the virus and the host. For the virus it is

Studies of yeast retrotransposons demonstrated that cellular cofactors determine the integration site preference (7–9). Likewise, interactions between HIV integrase (IN)§ and the host cell protein lens epithelium-derived growth factor (LEDGF/p75) (10) direct HIV integration targeting. LEDGF/p75 specifically binds IN via its integrase binding domain (LEDGE347–429) (Fig. 1) and functions as a molecular tether during lentiviral integration, bridging the viral pre-integration complex (PIC) with the host cell chromatin (11–13). Cells depleted for LEDGF/p75 (13, 14) or somatic knock-out cells (15, 16) inhibit productive HIV infection. In addition, fusion proteins in which the LEDGF/p75 DNA binding portion is replaced with an alternative chromatin binding domain efficiently tether the PIC to the host cell chromatin and support lentiviral vector transduction (17–20). We and others demonstrated that these fusion proteins retarget proviral integration to the regions bound by the specific chromatin binding domain (17, 18).

Defects in lentiviral infection are only observed after potent knockdown of LEDGF/p75 because residual protein levels are sufficient to support integration (13, 14, 21). This initially blurred the interpretation causing controversy with two studies that failed to observe a reduction in HIV replication after partial

This article has been withdrawn by the authors. The authors have become aware of a duplication of two lanes (the PWWP<sub>H</sub>LEDGF and HDGF-LEDGE<sub>325–530</sub>) in Fig. 2 of this manuscript and withdraw the article in the interests of maintaining their publication standards and those of the journal.

Received for publication, April 28, 2011, and in revised form, September 10, 2011 Published, JBC Papers in Press, October 10, 2011, DOI 10.1074/jbc.M111.255711

* This work was supported, in whole or in part, by National Institutes of Health Grants AI52845 and Al66230 (to F. D. B.), T32 AI007324 (to K. E. O.), and T32 AI-0724-17 (to K. R.). This work was also supported by KU Leuven Research Council Grant OT/09/047, Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) Cell-CoViR SBO Grant 60813, Flanders Research Foundation (FWO) Grant G.0530.08, European Commission THINC Grant HEALTH-F3–2008-201032 (to Z. D.), a predoctoral fellowship from Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) (to S. V.), and a Mathilde-Krim postdoctoral fellowship from Amfar (to J. D. R.). The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1 and Figs. S1–S9.

† To whom correspondence may be addressed: Kapucijnenvoer 33, VCTB 33, 3000 Leuven, Belgium. Fax: 32-16-33-21-31; E-mail: rik.gijsbers@med.kuleuven.be.

‡ Present address: Fred Hutchinson Cancer Research Center, Dept. of Human Biology, 1100 Fairview Ave N., MS: C3–168, P. O. Box 19024, Seattle, WA 98109-1924.

§ To whom correspondence may be addressed: Kapucijnenvoer 33, VCTB 33, 3000 Leuven, Belgium. Fax: 32-16-33-21-31; E-mail: zeger.debyser@med.kuleuven.be.

© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.
knockdown of LEDGF/p75 (22, 23). Although weak knockdown fails to demonstrate clear effects on HIV replication, the genomic distribution of HIV integration sites is already significantly affected under these circumstances (16, 24). The recent development of LEDGINs, small molecules that disrupt the LEDGF/p75-IN interaction, potently block HIV-1 replication and prove the requirement of LEDGF/p75 for HIV replication (25). The requirement of potent LEDGF/p75 knockdown (>90% depletion) to obtain a phenotype for HIV replication is the reason why LEDGF/p75 was not detected as a cofactor in any of the four genome-wide RNAi-based screens (26–29).

Detailed mapping of the chromatin binding profile of LEDGF/p75 by DamID technology revealed an association with markers of active chromatin and a disfavoring of promoter regions, a profile paralleling that of HIV-1 integration (30). Whereas the interaction between LEDGF/p75 and IN is well studied (12, 31–35), the molecular basis underlying the LEDGF/p75 chromatin interaction is largely unknown. Elements in the N-terminal portion of LEDGF/p75 have been shown to be necessary for chromatin binding, including a PWWP domain (amino acids 1–93) and two AT hook-like motifs (amino acids 178–197) (19, 36–38) (Fig. 1).

A PWWP domain contains a relatively well conserved Pro-Trp-Trp-Pro signature that is related to the Tudor domain Royal Family (39, 40). The domain is present in diverse chromatin-associated proteins involved in DNA repair, histone modification, transcriptional regulation, and DNA methylation (41–43). Although the function of the domain is not yet understood, the PWWP module has been reported to have sequence specificity (41, 44) and to have methyllysine binding ability (39, 40). The largest homolog of containing proteins is related to the histone lysine methyltransferase HDGF (classified as HDGF-related proteins (HRPs) (41). Deletion of the PWWP domain disrupts association of LEDGF/p75 with condensed mitotic chromatin (37). We generated an N-terminal LEDGF/p75 truncation that lacks the PWWP module together with a set of chimeric proteins in which the PWWP domain of LEDGF/p75 was swapped for that of two other HRPs: HDGF and HRP-2. In addition, we fused the HDGF D366A was amplified from pCP Nat LEDGF D366A (31) using LEDGF-KZ and Stu325_as primers (supplemental Table S1). Deletion of the PWWP domain was obtained using LEDGF_dPWWP_fwd_Xmal and Stul_325_as with pLNCHDGF BC-Ires-Bsd as a template. The latter ampiclon was digested by Xmal-Stul and cloned in pLNC_FLAG-AR-DBD-LEDGF325–530 BC-Ires-Bsd digested with Agel and Stul, resulting in pLNC AN93 LEDGF BC-Ires-Bsd. The latter construct AT-hook mutations were introduced by site-directed mutagenesis as described earlier (47) to generate pLNC AN93 LEDGF AT1 + 2 BC-Ires-Bsd. PWWP chimeras pLNC_PWWPHRF2-LEDGF BC-Ires-Bsd and pLNC_PWWPHDVDF-BC-Ires-Bsd were constructed using HsHDGF_s Agel and HsHRP2_s Agel. LEDGF_PWWP146as and HsHRP2 LEDGF-PWWP domains were amplified as PCR products encoding the PWWP domain from pLNC_FLAG-HRPHDF-325–530 BC-Ires-Bsd (17). All cloning steps were verified by sequencing.

Cell Culture—HeLaP4-CCR5 cells (a gift from Dr. Charneau, Institut Pasteur, Paris, France) were grown in Dulbecco’s modified Eagle’s medium (DEM) with Glutamax (Invitrogen) and supplemented with 5% heat-inactivated fetal calf serum (FCS; Sigma) and 50 μg/ml gentamicin and 500 μg/ml Geneticin (Invitrogen). LEDGF/p75-depleted cells were grown as described earlier (17). Transduced cells were selected with 200 μg/ml zeocin (Invitrogen) and/or 3 μg/ml blasticidin (Invitrogen). 293T cells were obtained from the ATCC and grown in DMEM supplemented with 8% FCS and 50 μg/ml gentamicin. Viral Vector Production—Vesicular stomatitis virus G (VSV-G) pseudotype vector particles were produced by PEI transfection of 293T producer cells with pCHMWS_eGFP-T2A-fluc (49), the packaging construct p8.91 (50), and the VSV-G encoding pMD.G construct. EIAV-based vector particles were produced likewise after triple-transduction and HIV virus replication. Ultimately, proviral integration sites were identified, and the genomic distribution of proviral integration sites was analyzed.

EXPERIMENTAL PROCEDURES

Construction of MLV-based Retroviral Vector—pLNC_FLAG-LEDF BC-Ires-Bsd was described earlier (17). pLNC_FLAG-LEDF BC D366A-Ires-Bsd was constructed by replacing the 5’ end of the LEDGF/p75 cDNA after Xhol-Stul digest. LEDGF D366A was amplified from pCP Nat LEDGF D366A (31) using LEDGF-KZ and Stu325_as primers (supplemental Table S1). Deletion of the PWWP domain was obtained using LEDGF_dPWWP_fwd_Xmal and Stul_325_as with pLNCHDGF BC-Ires-Bsd as a template. The latter ampiclon was digested by Xmal-Stul and cloned in pLNC_FLAG-AR-DBD-LEDGF325–530 BC-Ires-Bsd digested with Agel and Stul, resulting in pLNC AN93 LEDGF BC-Ires-Bsd. The latter construct AT-hook mutations were introduced by site-directed mutagenesis as described earlier (47) to generate pLNC AN93 LEDGF AT1 + 2 BC-Ires-Bsd. PWWP chimeras pLNC_PWWPHRF2-LEDGF BC-Ires-Bsd and pLNC_PWWPHDVDF-BC-Ires-Bsd were constructed using HsHDGF_s Agel and HsHRP2_s Agel. LEDGF_PWWP146as and HsHRP2 LEDGF-PWWP domains were amplified as PCR products encoding the PWWP domain from pLNC_FLAG-HRPHDF-325–530 BC-Ires-Bsd (17). All cloning steps were verified by sequencing.

Cell Culture—HeLaP4-CCR5 cells (a gift from Dr. Charneau, Institut Pasteur, Paris, France) were grown in Dulbecco’s modified Eagle’s medium (DMEM) with Glutamax (Invitrogen) and supplemented with 5% heat-inactivated fetal calf serum (FCS; Sigma) and 50 μg/ml gentamicin and 500 μg/ml Geneticin (Invitrogen). LEDGF/p75-depleted cells were grown as described earlier (17). Transduced cells were selected with 200 μg/ml zeocin (Invitrogen) and/or 3 μg/ml blasticidin (Invitrogen). 293T cells were obtained from the ATCC and grown in DMEM supplemented with 8% FCS and 50 μg/ml gentamicin. Viral Vector Production—Vesicular stomatitis virus G (VSV-G) pseudotype vector particles were produced by PEI transfection of 293T producer cells with pCHMWS_eGFP-T2A-fluc (49), the packaging construct p8.91 (50), and the VSV-G encoding pMD.G construct. EIAV-based vector particles were produced likewise after triple-transduction and HIV virus replication. Ultimately, proviral integration sites were identified, and the genomic distribution of proviral integration sites was analyzed.
sient transfection of p6.1G3CeGFPw,5 the packaging construct pEV53B (51) and pMD.G plasmid. HIV_NL4–3.Luc single round virus was prepared by transient transfection with pNL4–3.Luc_R–E− (NIH AIDS Research and Reference Reagent Program) and pMD.G. Supernatant was harvested, filtered through a 0.22-μm filter and concentrated by Vivaspin15 columns (Vivascience, Hannover, Germany), separated into aliquots, and stored frozen at −80 °C. Titers were quantified by p24 ELISA (Innotest HIV Antigen mAb, Innogenetics, Ghent, Belgium) or by functional titration on 293T cells as described earlier (52).

Vector Transduction and Analysis—For transduction experiments 20,000 cells were plated per well in a 96-well plate and transduced. 72 h post-infection, cells were reseeded in two new 96-well plates, one for qFACS and one for luciferase activity assay. For qFACS, transduced cells were fixed in a final concentration of 2% paraformaldehyde. Overall qGFP expression (mean fluorescence intensity) multiplied by the percentage of gated cells was measured with a FACSCalibur flow cytometer (BD Biosciences) and analyzed using the CellQuest software package provided with the instrument. For luciferase activity measurement, transduced cells were washed with 1X PBS and subsequently lysed with 70 μl of lysis buffer (50 mM Tris, pH 7.5, 200 mM NaCl, 0.2% Nonidet P-40, 10% glycerol). The lysate was assayed according to the manufacturer’s protocol (ONE-Glo™ Luciferase Assay System, Promega, Madison, WI). Data were normalized for total protein (BCA Protein Assay, Pierce) or by functional titration on 293T cells as described earlier (52).

Vector Transduction and Analysis—For transduction experiments 20,000 cells were plated per well in a 96-well plate and transduced. 72 h post-infection, cells were reseeded in two new 96-well plates, one for qFACS and one for luciferase activity assay. For qFACS, transduced cells were fixed in a final concentration of 2% paraformaldehyde. Overall qGFP expression (mean fluorescence intensity) multiplied by the percentage of gated cells was measured with a FACSCalibur flow cytometer (BD Biosciences) and analyzed using the CellQuest software package provided with the instrument. For luciferase activity measurement, transduced cells were washed with 1X PBS and subsequently lysed with 70 μl of lysis buffer (50 mM Tris, pH 7.5, 200 mM NaCl, 0.2% Nonidet P-40, 10% glycerol). The lysate was assayed according to the manufacturer’s protocol (ONE-Glo™ Luciferase Assay System, Promega, Madison, WI). Data were normalized for total protein (BCA Protein Assay, Pierce) or by functional titration on 293T cells as described earlier (52).

Integration Site Cloning—Integration sites were amplified by nested PCR using bar-coded primers. This enabled pooling of PCR products into one sequencing reaction. Products were gel-purified and sequenced on the 454 GS-FLX instrument at the University of Pennsylvania.

Bioinformatic Analysis—Integration sites were analyzed essentially as described earlier (17). Briefly, sites were judged to be authentic when the sequences contained a proper bar code and long terminal repeat (LTR) primer and had a best unique hit when aligned to the human genome as appropriate (hg18) using BLAT, and the alignment began within 3 bp of the viral LTR end and had >98% sequence identity. Statistical methods

5 M. Patel and J. Olsen, unpublished information.

DECEMBER 2, 2011
RESULTS

Generation of LEDGF/p75 Chimeras—To investigate the role of the N-terminal DNA binding elements of LEDGF/p75 in HIV integration targeting, we generated a series of chimeric proteins, specifically deleting, replacing, or altering these elements (Fig. 1). In a first step we targeted the PWWP domain of LEDGF/p75. We deleted the complete domain (amino acids 1–93) to generate ΔNWP-LEDGF. To study the role of the PWWP domain, we replaced the PWWP domain of LEDGF/p75 with that of HDGF (amino acids 1–146) and HRP-2 (amino acids 1–106), generating PWWP_{HDGF}-LEDGF and PWWP_{HRP2}-LEDGF, respectively. Although the PWWP domains of LEDGF/p75, HDGF, and HRP-2 are 80% identical, essential amino acids believed to be involved in the interaction with chromatin are different among the different proteins (41) (supplemental Fig. S1). HDGF has been shown to act as a transcriptional repressor by binding a conserved element in the promoter of target genes via its N-terminal PWWP domain (45). Additionally, we replaced the chromatin binding elements of LEDGF/p75 (amino acid 1–324) with full-length HDGF (HDGF-LEDGF{325–530}). Subsequently, constructs encoding the chimeric proteins were stably introduced into LEDGF/p75-depleted cells (17) using MLV-based viral vectors and selected with blasticidin. Control cell lines expressing RNAi-resistant LEDGF/p75 (LEDGF BC) or the HIV integrase interaction-deficient D366A mutant (LEDGF_{D366A}) were generated in parallel (11, 35). Growth rates were comparable with that of the parental HeLaP4-CCR5 cell line for all generated cell lines (data not shown). Expression of the fusion proteins was verified by Western blotting (Fig. 2). Although not all proteins were expressed to the same extent, all fusion proteins migrated at the correct molecular weight.

LEDGF/p75 Chimeras Tether HIV IN to the Nucleus—Whereas endogenous LEDGF/p75 typically appears as dense fine speckles in the interphase nucleoplasm, it localizes to condensed chromatin during mitosis (59). Immunocytochemistry for LEDGF/p75 demonstrated no fluorescence in the LEDGF/p75-depleted cells (Fig. 3a) (17). Complementation of the depleted cells with LEDGF BC or the D366A mutant (LEDGF_{D366A}) restored both the fine speckled nuclear pattern
During interphase and the binding to mitotic chromatin, in line with earlier reports (46, 59, 60) (Fig. 3, b and i, and Fig. 3, c and j, respectively). Truncation of the PWWP domain (ΔN93-LEDGF) resulted in a more diffuse subnuclear distribution with loss of nucleolar exclusion (Fig. 3d) and a loss of interaction with condensed chromatin (Fig. 3k), indicating that the PWWP domain is a major determinant for chromatin association. Swapping the PWWP domain of LEDGF/p75 with that of either HDGF or HRP-2 recovered the speckled nuclear distribution with nucleolar exclusion (Fig. 3e and f, respectively), comparable with wild-type LEDGF/p75, and restored interaction with mitotic chromosomes (Fig. 3, l and m, respectively). Replacing the N-terminal end of LEDGF/p75, containing all chromatin binding elements, with full-length HDGF resulted in similar staining (Fig. 3, g and n, respectively), demonstrating that the N-terminal domain of LEDGF/p75 can be replaced with alternative chromatin binding elements, in line with earlier reports (17, 19).

In addition to nuclear localization, LEDGF/p75 hybrids should support chromatin tethering of HIV-1 IN. Transient expression of HIV-1 integrase fused to monomeric red fluorescent protein (mRFP-IN*) resulted in a diffuse fluorescent signal throughout the cytoplasm and the nucleus in the absence of LEDGF/p75 (Fig. 3, a and h, red fluorescence), as previously reported (17, 22, 59). Expression of LEDGF BC relocated mRFP-IN* to the nucleus and to condensed chromatin (Fig. 3, b and i), whereas complementation with the interaction-deficient LEDGFD336A did not (Fig. 3, c and j), corroborating the requirement of a direct interaction of the integrase binding domain with IN for both the nuclear and condensed chromatin binding. Although deletion of the LEDGF/p75 PWWP domain (ΔN93-LEDGF) supported nuclear localization of IN during interphase, interaction with condensed chromosomes was lost (Fig. 3, d and k, respectively). However, substitution of the PWWP domain of LEDGF/p75 with that of HDGF or HRP-2 rescued nuclear localization of IN, resembling the wild-type...
LEDGF/p75 phenotype (Fig. 3, e and f and Fig. 3, f and m, respectively). Likewise, fusion of LEDGF<sub>325–530</sub> to full-length HDGF relocated IN to the nucleus of interphase cells and to condensed chromatin (Fig. 3, g and n, respectively).

**LEDGF/p75 Chimeras Promote Lentiviral Transduction**—
LEDGF/p75 plays a pivotal role in lentiviral integration and viral infection (11–15, 24). After determining the abilities of LEDGF/p75 chimeric proteins to interact with HIV-1 IN and thereby to tether IN to chromatin, we asked whether each of the chimeras promotes lentiviral transduction in LEDGF/p75-depleted cells. We previously generated potent knockdown cell lines and matched control cell lines (17). Complemented cell lines were challenged with HIV-based lentiviral vectors encoding firefly luciferase (Fluc) as a reporter protein (49). Transduction efficiency was evaluated by assaying Fluc activity (relative light units/μg of protein) relative to that of LEDGF BC. Transduction efficiency of LEDGF/p75-depleted cells (KD) was 16.8% that of LEDGF BC (Fig. 4A). Complementation of KD cells with interaction-deficient LEDGF<sub>D366A</sub> rescued transduction only to 25.2% that of LEDGF BC (p > 0.1 compared with KD, two-tailed t test). Introduction of LEDGF/p75 lacking the PWWP domain (∆N<sub>93</sub>-LEDGF) rescued vector transduction to 48.5% that of the LEDGF BC level (p = 0.0268 compared with KD, two-tailed t test). Substitution of the LEDGF/p75 PWWP domain with that of HDGF or HRP-2 restored lentiviral vector transduction to near wild-type levels (95.2 and 82.4% of LEDGF BC, respectively). Likewise, fusion of LEDGF<sub>325–530</sub> to full-length HDGF restored vector transduction to 87.9% of LEDGF BC. Similar data were obtained when employing eGFP as a reporter (supplemental Fig. S2) or when using a near-complete HIV reporter virus (HIV-1NL4.3.LucR<sup>e</sup>) (supplemental Fig. S3).

Next we quantified the number of integrated proviral copies in the genomic DNA of the respective cell lines with previous data (17), introduction of LEDGF BC in KD cells resulted in a 3.9-fold increase in integrated copies, with levels comparable with those obtained in wild-type cells (data not shown), whereas complementation with LEDGF<sub>D366A</sub> or ∆N<sub>93</sub>-LEDGF did not significantly increase the number of integrated copies over KD cells (p > 0.1, two-tailed t test). Complementation of LEDGF/p75-depleted cells with the PWWP chimeras rescued integration to near wild-type levels (87 and 73% that of LEDGF BC for the PWWP<sub>HDGF</sub>- and PWWP<sub>HRP2</sub>-LEDGF, respectively). Similarly, introduction of HDGF-LEDGF<sub>325–530</sub> recovered 75.4% of the LEDGF BC levels.

In addition to HIV-1N, LEDGF/p75 interacts with other lentiviral integrases (22, 61, 62). To evaluate the potency of the LEDGF/p75 chimeras to complement other lentiviruses, we transduced the respective cell lines and control cells with an EIAV-based viral vector, engineered to encode eGFP and Fluc reporters. Introduction of the respective LEDGF/p75 fusions in knockdown cells rescued EIAV transduction and integration, paralleling the results obtained for the HIV-based viral vectors (supplemental Fig. S4).

**Rescue of Spreading HIV Infection by LEDGF/p75 Chimeras**—
We next evaluated the capacity of the LEDGF/p75 chimeras to rescue HIV-1 replication. Cell lines stably expressing LEDGF/p75 chimeras were infected with HIV-1<sub>NL4.3</sub> virus (28,000 pg of p24/ml; multiplicity of infection 0.01). Viral replication was monitored by sampling the culture medium at regular intervals and subsequent determination of the p24 concentration by ELISA. Experiments were repeated independently at least three times. A representative HIV replication experiment is shown (Fig. 5A). As expected, HIV replication was significantly inhibited in KD cells (compared with HeLaP4-CCR5 control cells (WT). Complementation of the KD cells with LEDGF BC restored HIV replication to near wild-type levels, whereas expression of the interaction-deficient LEDGF<sub>D366A</sub> did not restore HIV replication. Cells expressing ∆N<sub>93</sub>-LEDGF, lacking the PWWP domain, supported HIV replication only marginally but significantly above the level observed for the KD cells. By comparison, exchanging the PWWP of LEDGF/p75 with that of either HDGF or HRP-2 restored HIV replication to near

**FIGURE 4. Rescue of lentiviral vector transduction by PWWP chimeras and LEDGF/p75 hybrids.** WT, LEDGF BC, and the interaction-deficient LEDGF<sub>D366A</sub> cells were included as controls. A, shown is relative luciferase activity (RLU/μg of protein) after HIV-based vector transduction (LV CMV eGFP-T2A-Fluc). Data compile at least four independent experiments and represent percentages relative to LEDGF BC cells (mean ± S.D.). B, vector integration was estimated by quantitative PCR analysis in the same cell lines. Data are normalized and are represented as the mean ± S.D.
wild-type levels, with a delay compared with LEDGF BC cells. Likewise, introduction of HDGF-LEGDF325–530 in LEDGF/p75-depleted cells resulted in rescue of HIV replication (Fig. 5A). Comparable results were obtained when cells were infected at higher multiplicity of infection (supplemental Fig. S5). Estimation of integrated HIV copies by quantitative PCR (Fig. 5B) demonstrated rescue of viral integration to wild-type levels upon introduction of LEDGF BC (8.3-fold more than KD), whereas expression of LEDGF_{D366A} or ΔN_{93}-LEDGF did not. The number of integrated proviral copies for LEDGF_{D366A} complemented cells was not different from that in KD cells (p = 0.2857, two-tailed t test), whereas ΔN_{93}-LEDGF reached significance (p = 0.0193, two-tailed t test). Swapping the LEDGF/p75 PWWP domain with that of HDGF or HRP-2 or fusion of LEDGF^{325–530} to HDGF restored HIV proviral integration to 61.6, 65.3, and 39.8% of LEDGF BC, respectively.

The Integration Site Consensus Sequence Is Not Affected by LEDGF/p75 Hybrids—We next asked whether altering the DNA binding elements of LEDGF/p75 affects integration site distribution. Integration sites were determined for an EIAV-based viral vector, as HeLaP4 cells contain an integrated HIV long terminal repeat that might interfere with the isolation of HIV proviral integration sites. Both EIAV and HIV-IN interact with LEDGF/p75. The same integration site preference around the site of integration. Consistent with previous reports (15–17), the characteristic palindrome at the site of integration was conserved (supplemental Fig. S6) even when integration sets were significantly different from their respective MRC sites and allows for more accurate statistical analysis (3, 54).

Genomic Distribution of Integration Sites in Cells Containing LEDGF/p75 Hybrids—Lentiviruses favor integration into transcription units and gene-dense regions (1, 3, 63). Depletion of LEDGF/p75 reduces this preference, and a preference for integration close to CpG islands and gene 5' ends emerges instead (15, 16). As an initial survey of the proviral integration site distribution in the different complemented cell lines, we first analyzed the integration frequency relative to these features. All integration sets were significantly different from their respective MRC sites for the integration into RefSeq genes in a Mann-Whitney test (p < 0.001; Table 1, statistics not shown). Integration frequency in RefSeq transcription units and gene-dense regions (1, 3, 63). EIAV integration sites were analyzed as a total of 4923 integration sites and allows for more accurate statistical analysis (3, 54). Retroviral integrases display a virus-specific target sequence preference around the site of integration. Consistent with previous reports (15–17), the characteristic palindrome at the site of integration was conserved (supplemental Fig. S6) even when the genomic distribution of integration sites is altered, underscoring that IN determines the local sequence preference in the target DNA independent of the tethering mechanism.
pared integration site distributions of the different sets for a selection of genomic features. The heat map (Fig. 6) summarizes relationships between integration sites and specific genomic features (54, 55). Enriched associations compared with random (MRC) are displayed as increasing shades of red, and negative associations are displayed as increasing shades of blue, with no difference as gray tiles. Statistical significance compared with the KD data set is determined by regression and represented by asterisks overlaid on the tile (54). Again, integration site distribution in LEDGFD366A cells was similar to that of KD cells (Fig. 6).

**FIGURE 6. Heat map of integration frequency relative to genomic features.** A heat map summarizes the relationships of proviral integration site data sets to genomic features. Integration data sets are indicated above the columns. Genomic features analyzed are shown to the left of the corresponding row of the heat map (55). Tile color indicates whether a genomic feature is favored (red, enrichment compared with random) or disfavored (blue, depletion compared with random) for integration for the respective data sets relative to their MRCs, as detailed in the colored ROC area scale at the bottom of the panel. p values show significance of departures from the KD data set are shown with asterisks (*, p < 0.05; **, p < 0.01; ***, p < 0.001, Wald statistics referred to $\chi^2$ distribution). The naming of the genomic features is described in Brady et al. (55); TSS, transcription start site.

**PWWP Domain and Integration Targeting**

DECEMBER 2, 2011 • VOLUME 286 • NUMBER 48 • JOURNAL OF BIOLOGICAL CHEMISTRY

**WITHDRAWN**

December 13, 2017

**PWWP Swaps Rescue the Knockdown Phenotype**—Next, the role of the PWWP domain of LEDGF/p75 in integration site targeting was investigated. In a first step we deleted the PWWP domain ($\Delta N_{93}$-LEDGF). Introduction of $\Delta N_{93}$-LEDGF into LEDGF/p75-depleted cells failed to rescue the KD phenotype (Table 1) and showed diminished integration in RefSeq genes and a preference for integration near CpG islands as in LEDGF/p75 KD ($p = 0.929$ and $p = 0.234$ compared with KD, respectively; Mann-Whitney test). There was no additional alteration in the genomic features heat map (Fig. 6, no statistical difference compared with KD). Swapping the PWWP domain of LEDGF/p75 with that of HDGF or HRP-2 rescued integration frequency in RefSeq genes with the concomitant reduction of integration near CpG islands to levels observed in LEDGF BC cells (Table 1). The correlation of integration sites with several genomic features was restored to levels observed in LEDGF BC cells (supplemental Fig. S8, no statistical difference compared with LEDGF BC). Replacing all DNA binding elements in LEDGF/p75 with full-length HDGF (HDGF-LEDGF325–530) recovered integration into transcription units (84.1% in RefSeq genes; Table 1), favoring integration in the body of genes significantly more than LEDGF BC ($p = 1.84e-6$, Mann-Whitney test). Integration in HDGF-LEDGF325–530-expressing cells was still preferred near CpG islands as observed in KD cells (4.8%; $p = 0.5477$ compared with KD, Mann-Whitney test). In addi-
tion, the genomic heat map demonstrated favored integration in GC-rich regions for HDGF-LEDGF325–530-expressing cells (Fig. 6 and supplemental Fig. S8).

**HDGF Fusion Does Not Retarget to HDGF Binding Sites**—We and others demonstrated that fusion of alternative DNA-binding proteins to the C-terminal end of LEDGF/p75 allows retargeting of integration in the neighborhood of the respective protein binding sites (17, 18). HDGF has been demonstrated to bind specifically to SMYD1 promoter-like sequences (45). We analyzed the number of the HDGF binding sites around each lentiviral integration site using various window sizes (5 and 10 kb). The frequency of the SMYD1-like sequence near integration sites of KD and ΔNαβ-LEDGF-expressing cells was not different compared with their respective MRCs (Table 2; p > 0.05, Fisher’s exact test). However, the frequency was significantly different for all other sets (p < 0.001, Fisher’s exact test). Pairwise comparison of the frequency of the SMYD1-like sequences of the different integration site sets to that of LEDGF BC only reached significance for KD cells, which showed reduced frequency (Mann-Whitney test). Thus all the functional fusion proteins targeted integration near SMYD1-like sequences, likely a consequence of targeting integration to gene-rich regions. Elevated frequency of targeting SMYD1-like sites was seen for HDGF-LEDGF325–530 (52.8% for the fusion versus 47.8% for the LEDGF BC; 5-kb window), but the trend did not achieve significance given the size of the dataset; when combining the LEDGF BC and WT data sets, a significant difference was observed for HDGF-LEDGF325–530 (p < 0.01 for 5- and 10-kb windows).

**PWWP Domain and Integration Targeting**

Recently, it was demonstrated that PWWP modules in several other proteins recognize methyllysine residues on histone tails (57, 64–66). We thus evaluated the role of the PWWP domain in the integration site distribution relative to histone modifications for LEDGF BC, PWWPHDGF-LEDGF, and PWWPHRP2-LEDGF, as they are the only domain that differs among these proteins. For analysis we used high resolution maps from a study in HeLa cells, where H3K4 mono- and trimethylation ChIP-seq profiles were identified (58), supplemented with the genome-wide distribution of 39 histone modifications in human CD4⁺ T cells (56, 57). Associations of integration and histone methylation/acetylation were quantified (Fig. 7, 10-kb window). Histone modifications are grouped into clusters reported to co-localize and associate with chromatin states (58). The heat map of integration frequency relative to epigenetic marks is shown. ChIP-seq data from HeLaS3 cells (58) and human CD4⁺ T cells (56, 57) were used. Detailed information on the epigenetic marks can be found in Refs. 56 and 57. Aspects of integration and histone methylation/acetylation were quantified using ROC areas, comparing the association of integration site data sets with the frequency in corresponding MRC sets. A ROC area scale is shown along the bottom of the panel. Tile color indicates whether a chosen feature is favored (blue, enrichment relative to random) or disfavored (yellow, negative correlation compared with random) for integration (10-kb window) in the respective data sets relative to their MRCs. p values showing significance of departures from WT cells are shown with asterisks (**, p < 0.01; ***, p < 0.001, Wald statistics referred to χ² distribution), dashes overlay control tiles.

### TABLE 2

| Cell line          | No sites | No./5-kb window | No./10-kb window |
|--------------------|----------|-----------------|------------------|
| EIAV sites         |          |                 |                  |
| WT                 | 717      | 45.3            | 65.8             |
| KD                 | 448      | 40*             | 59.8#            |
| LEDGF BC           | 862      | 47.8            | 70.2             |
| ΔNαβ-LEDGF         | 157      | 43.9            | 62.4#            |
| PWWPHDGF-LEDGF    | 610      | 44.6            | 67.4             |
| PWWPHRP2-LEDGF    | 802      | 48.6            | 68               |
| HDGF-LEDGF325–530 | 873      | 52.8            | 71.7             |
| MRC sites          |          |                 |                  |
| MRC KD             | 1344     | 34.4            | 53.6             |
| MRC LEDGF BC       | 2586     | 34              |                  |
| MRC ΔNαβ-LEDGF     | 471      | 36.5            |                  |
| MRC PWWPHDGF-LEDGF| 1827     | 34              |                  |
| MRC PWWPHRP2-LEDGF| 2406     | 34              |                  |
| MRC HDGF-LEDGF325–530| 2619 |                  |                  |

*p < 0.001.

*p < 0.01.

*p < 0.05.
with classes of functional genomic elements (57). Tile color reflects whether a feature is favored (blue, enriched relative to random) or disfavored (yellow, depleted compared with random) for integration in the respective data set.

Integration frequency showed similar patterns for all four data sets, positively correlating with histone modifications generally associated with active transcription (acetylations and monomethylation of H3K27, H3K9, H4K20, H3K79, H2BK5, etc.) and negatively associating with markers common to transcriptionally repressed regions (e.g. H3K27me3, H3K9me3) and heterochromatin (e.g. H4K20me3 and H3K79me3). Significant departures from WT, indicated with asterisks (**, \( p < 0.01; ***\), \( p < 0.001 \), Wald statistics referred to \( \chi^2 \) distribution), are most prominent for PWWP_HRP2-LEDGF (Fig. 7). To deduce whether certain epigenetic marks are favored by a specific PWWP domain, we counted the epigenetic marks around integration sites for all PWWP data sets (Fig. 8). The plot compares the distribution of epigenetic counts of experimental PWWP data sets versus control (WT) for several window sizes (10 kb, 100 kb, 1 Mb windows). As expected, no significant difference was observed for LEDGF BC \( (p > 0.05; \text{Wilcoxon rank sum test}) \). However, when the PWWP of LEDGF/p75 was replaced with that of HDGF or HRP-2, significance was reached for several modifications (10-kb and 100-kb window). Integration was preferred near H3K36me3 for PWWP_HRDG-LEDGF \( (p < 0.001; \text{Wilcoxon rank sum test}) \), and a ratio >2 was obtained near H4R3me2 and H3K9me3 \( (p < 0.05; \text{Wilcoxon rank sum test}) \).

Integration in PWWP_HRP2-LEDGF-expressing cells was preferred near acetylated H2BK5, H3K27, H4K16 and near H2BK5me3, H3K36me3, and H4K20me1 \( (p < 0.001; \text{Wilcoxon rank sum test}) \). Taken together, these results indicate that the PWWP domain is involved in either nucleolar exclusion or bound by marks or else binds factors positively correlated in a fashion correlated with epigenetic marks around integration sites.

Role of the AT Hook-like Motif

Consistent with integrase site analysis indicated that deletion of the PWWP domain failed to rescue the KD integration site distribution (Table 1 and Fig. 6). In addition, we demonstrated that the PWWP domain is needed for efficient HIV integration and replication (Fig. 5). However, complementation of LEDGF/p75-depleted cells with \( \Delta N_{93} \)-LEDGF still rescued viral vector transduction and virus replication, albeit only marginally. One explanation could be that the AT-hook-like motifs support nonspecific binding to chromatin (36, 70, 47). In an effort to study the contribution of the AT-hook-like domains, we mutated both AT-hook-like motifs by site-directed mutagenesis (R183D, K192D, R196D) (47) and complemented LEDGF/p75-depleted cells \( (\Delta N_{93}-\text{LEDGF}^\text{AT}) \). \( \Delta N_{93} \)-LEDGF^AT locates to the nucleus and does not interact with mitotic chromatin (supplemental Fig. S9a) (36, 38). HIV-based lentiviral vectors showed a reduced transduction efficiency compared with \( \Delta N_{93} \)-LEDGF cells (supplemental Fig. S9b; \( p = 0.0113 \), two-tailed \( t \) test), in line with earlier reports (38, 67). Next, stable \( \Delta N_{93} \)-LEDGF and \( \Delta N_{93} \)-LEDGF^AT cells were infected with HIV-1 NL4-3 virus (supplemental Fig. S9c). Although HIV replication was already severely affected in cells complemented with \( \Delta N_{93} \)-LEDGF compared with LEDGF/p75 BC, additional mutation of the AT hook-like motifs reduced HIV replication to the background levels obtained in LEDGF\(_{D366A}\) or LEDGF/p75-depleted cells (supplemental Fig. S9c, inset).

**DISCUSSION**

To determine whether the PWWP domain plays a role in LEDGF/p75-mediated tethering and targeting of the lentiviral PIC, we employed stable knockdown cell lines to allow analysis in the absence of endogenous LEDGF/p75 protein (17). We deleted or replaced the PWWP domain of LEDGF/p75 and studied IN localization, HIV replication, and lentiviral integration site distribution in the host genome. Stable integration of the viral DNA links the fate of the invading virus with that of the host cell. For the virus, integration in active chromatin facilitates viral gene expression, which in turn will result in a successful infection cycle. Although the role of the PWWP domain has been studied before (19, 36–38, 68–73), we provide the first analysis of its impact on integration site selection.

Deletion of the PWWP domain in LEDGF/p75 resulted in a diffuse subnuclear distribution, loss of nucleolar exclusion (Fig. 3) (19, 36, 37, 69). Interaction with mitotic chromatin (19, 37) and binding to the nucleolus (19, 46, 68, 74). Nucleolar localization is consistent with current understanding suggests that the PWWP domain is required (75, 76). Although the PWWP domain of LEDGF/p75 locates to the nucleolus, it is not required by a “cryptic” nucleolar localization signal. Integration is almost completely unraveled by the presence of PWWP. Several AT-hook-like motifs have been identified that can target a protein to the nucleolus (76), such as the \((R/K)(R/K)\) motif of the human La protein (77), which occurs in multiple copies in \( \Delta N_{93} \)-LEDGF.

Conflicting results with regard to the relevance of the PWWP domain for the LEDGF/p75 cofactor activity during HIV replication have been reported. Although the domain was shown to be essential for HIV-based vector transduction in LEDGF/p75−/− mouse fibroblasts (38) (rescuing \( \leq 0.1 \sim 20\% \) of WT activity) and in LEDGF/p75-deficient human CD\(^+\) T cells (19, 67) (44.2 and 63% of WT activity, respectively), it also was found to be dispensable (73). In our hands complementation of LEDGF/p75-deficient cells with \( \Delta N_{93} \)-LEDGF restored transduction efficiency only partially (48% of WT activity; Fig. 4A), in agreement with the initial reports in the LEDGF/p75-depleted CD\(^+\) T cells (19, 67). The absence of a significant rescue of integrated proviral copies relative to LEDGF/p75-depleted cells indicates that \( \Delta N_{93} \)-LEDGF does support productive virus replication only poorly. This was confirmed in multiple-round HIV infection experiments (Fig. 5B). The residual activity compared with LEDGF/p75 KD may originate from non-integrated vector copies or an increased stability of the IN complex in the presence of \( \Delta N_{93} \)-LEDGF (60). One explanation could be that the AT-hook-like motifs support nonspecific binding to chromatin (36, 47, 70). Additional mutation of the AT hook-like motifs in \( \Delta N_{93} \)-LEDGF further reduced viral vector transduction and HIV virus replication to the background levels of LEDGF/p75-depleted or interaction-deficient LEDGF\(_{D366A}\) control cells (\( p > 0.05 \) compared with KD and LEDGF\(_{D366A}\);
Student’s t test), suggesting a role for the AT hook-like region in the marginal rescue of HIV replication by ΔN93-LEDGF.

We also generated LEDGF/p75 hybrids where we exchanged the PWWP domain of LEDGF/p75 with that of HDGF or HRP-2, and we replaced the chromatin binding N terminus of LEDGF/p75 with HDGF (HDGF-LEDGF325–530). Complementation of LEDGF/p75-depleted cells with these chimeras restored nuclear localization of HIV-1 IN and rescued trans-

FIGURE 8. Distribution of epigenetic marks near integration sites of PWWP chimeras. Epigenetic marks around integration sites were counted for all PWWP data sets at different window sizes (10 kb, 100 kb, 1 Mb). Significance of departure from controls is indicated by the size of the circle (Wilcoxon rank sum test compared with WT). The color reflects the ratio of average epigenetic marker counts between sample and WT control. The ratio indicates whether integration is favored (ratio >1; blue) or disfavored (ratio <1; red) near individual marks compared with control.
Our conclusion is that the PWPP domain in LEDGF/p75 is essential for HIV integration and replication and that different PWPP domains can fulfill this requirement. However, the expression of PWPP_{HDGF}-LEDGF and PWPP_{HRP2}-LEDGF proteins in LEDGF/p75-depleted cells did not fully rescue HIV_{NL4.3} replication to wild-type levels (Fig. 5), indicating that the natural LEDGF/p75 PWPP domain may have unique properties.

We analyzed lentiviral integration site distributions and confirmed the characteristic palindromes at the site of integration was conserved in all data sets (supplemental Fig. S6), consistent with data that the sequence is determined by integrase (78, 79). Deletion of the PWPP domain resulted in an integration pattern that was found in LEDGF/p75-depleted cells or cells containing interaction-deficient LEDGF_{D366A} (Table 1 and Fig. 6). The patterns in the absence of the PWPP domain showed reduced integration into transcription units and gene-dense regions and a preference for CpG islands and gene 5' ends. Integration frequency relative to genomic features was restored to wild-type levels when the PWPP domain of LEDGF/p75 was replaced with that of other HRP proteins (HDGF, HRP-2). Replacing all chromatin binding elements in LEDGF/p75 with full-length HDGF resulted in significantly more integration in genes compared with LEDGF BC combined with a preference for CpG islands and favored integration in GC-rich regions (Table 1 and Fig. 6). Even though integration into genes was favored in HDGF-LEDGF_{325–530} could not rescue HIV replication to fully wild-type levels. Lentiviral integration favors gene dense regions, predictably being the largest window around integration sites (80). When considering only small genomic intervals around integration sites, lentiviral integration associates with specific epigenetic marks. To deduce whether certain epigenetic marks are favored by a specific PWPP domain, we compared the distribution of epigenetic marks that are recognized, these data suggest that PWPP_{HDGF}-LEDGF-, PWPP_{HRP2}-LEDGF-, and LEDGF BC-tethered integration associates with specific epigenetic marks. To deduce whether certain epigenetic marks are favored by a specific PWPP domain, we compared the distribution of epigenetic counts of the PWPP sets versus WT cells (Fig. 8).

One explanation for our data might be that variety in the PWPP domains results in different integration targeting. Binding motifs differ among family members; the HRP-family, containing LEDGF/p75, HDGF, and HRP-2 among other proteins, typically contains a PHWP motif, whereas the NSD1 family of histone methyltransferases often contains a RWHP motif, and DNA methyltransferase 3a/b carries a SWHP motif. Direct models are also possible in which different PWPP domains favor binding to chromosomal sites that have differing associations with histone residues and modifications. For the virus, selection of integration sites adjacent to particular epigenetic marks through binding of a cellular cofactor that recognizes these marks may be beneficial for proviral gene expression and/or may affect the ratio of transcriptionally latent versus active proviruses. More detailed studies of PWPP domain biology and effects on integration should clarify the molecular mechanisms involved.

Acknowledgments—We thank the K. U. Leuven Cell Imaging Core for use of the confocal microscope. Paulien Van de Velde and Barbara Van Remoortel provided excellent technical assistance.
