Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-derived Growth Factor/Transcriptional Co-activator p75 (LEDGF/p75) That Binds HIV-1 Integrase

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation
Cherepanov, Peter, Eric Devroe, Pamela A. Silver, and Alan Engelman. 2004. “Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-Derived Growth Factor/Transcriptional Co-Activator p75 (LEDGF/p75) That Binds HIV-1 Integrase.” Journal of Biological Chemistry 279 (47): 48883–92. https://doi.org/10.1074/jbc.m406307200.

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:41482992

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) protein was recently identified as a binding partner for HIV-1 integrase (IN) in human cells. In this work, we used biochemical and systematic approaches to define the domain organization of LEDGF/p75. Using limited proteolysis and deletion mutagenesis we show that the protein contains a pair of evolutionarily conserved domains, assuming about 35% of its sequence. Whereas the N-terminal PWWP domain had been recognized previously, the second domain is novel. It is comprised of ~80 amino acid residues and is both necessary and sufficient for binding to HIV-1 IN. Strikingly, the integrase binding domain (IBD) is not unique to LEDGF/p75, as a second human protein, hepatoma-derived growth factor-related protein 2 (HRP2), contains a homologous sequence. LEDGF/p75 and HRP2 IBDs avidly bind HIV-1 IN in an in vitro GST pull-down assay and each full-length protein potently stimulated HIV-1 IN activity in vitro. LEDGF/p75 and HRP2 are predicted to share a similar domain organization and have an evident evolutionary and likely functional relationship.

Human immunodeficiency virus type 1 (HIV-1) integrase (IN) accomplishes the joining of the reverse-transcribed viral genome to a host cell chromosome (for reviews see Refs 1–3). Its activity is essential for viral replication and spread in primary cells and contributes to the persistence of the viral infection in vivo (4, 5). Blocking HIV-1 IN activity by specific inhibitors was shown to arrest viral spread in cell culture (6). HIV-1 IN, born to the transposase family of DNA transposases, is structurally and mechanistically similar to Mu phage and Tn5 transposases. Its active site formed by the three acidic residues Asp84, Asp116, and Glu152 (known as the “DDE motif”) is located within the structurally conserved catalytic core domain. The core domain is contained within residues 50–212 of the protein and flanked by the N-terminal HHCC-type zinc finger and the C-terminal DNA binding domains. Similar to the related bacterial transposases, retroviral INs form multimers, although the true stoichiometry of IN within the retroviral preintegration complex (PIC) is not known (7).

Mutations in HIV-1 IN display a wide range of phenotypes, affecting viral replication at the integration step (class I mutants), or causing various pleiotropic effects on virion morphogenesis and reverse transcription (class II) (8). Pleiotropic phenotypes of many IN mutants advocate that IN might have additional functions in viral replication. Thus, a role for IN in reverse transcription has been proposed (9). The complex phenotypes of class II mutants could potentially be explained by failure of the mutant INs to interact with viral reverse transcriptase and/or a host cell factor(s). A number of cellular and viral proteins were suggested to participate in retroviral integration (for a review see Ref. 10). Furthermore, several proteins were reported to directly interact with HIV-1 IN, including viral reverse transcriptase (9), a component of the SWI-SNF chromatin-remodeling complex INI1 (11), uracil DNA glycosylase UNG2 (12), heat shock protein HSP60 (13), a DNA repair protein Rad18 (14), a Polycyth group protein EED (15) and lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) (Ref. 16, for a review see Ref. 17). The exact roles of these proteins and their importance to viral replication have yet to be determined. However, when HIV-1 or feline immunodeficiency virus (FIV) INs are expressed separately from other viral proteins, endogenous host-cell LEDGF/p75 appears to be the dominant interactor, accounting for their nuclear/chromosomal accumulation (16, 18, 19). LEDGF/p75 protein markedly stimulated HIV-1 IN activity in vitro and was recently reported to be associated with functional HIV-1 PICs (16, 19). These data cumulatively suggest that LEDGF/p75 and possibly its homologs pose as cellular host factors in retroviral replication likely acting at the levels of chromosomal targeting, and/or integration of viral cDNA (17). LEDGF/p75 belongs to a family of hepatoma-derived growth factor (HDGF)-related proteins (HRPs). Five mammalian HRPs are known: HDGF, HRP1, HRP2, HRP3, HRP4, and...
Isolation and Sequence Analysis of LEDGF/p75 and HRP2 cDNAs—A wealth of expressed sequence tags (ESTs) representing fragments of cDNAs encoding homologs of human LEDGF/p75 from various vertebrate sources were readily identified by searching the NCBI sequence data base with translating basic local alignment search tool (BLAST) (www.ncbi.nlm.nih.gov/BLAST). ESTs with the following GenBank accession numbers gave sufficient sequence information to design primers for PCR amplification of the complete coding regions of Gallus gallus LEDGF/p75 cDNA—BU112216, AJ394255, BU323557, BU129859, and CN231179. ESTs for the Xenopus laevis ortholog were: BJ042207, BJ055881, BX528242, BU912001, and BQ792240. Total RNA isolated from G. gallus-b-globin-P cell line DT40 or kidney tissue from an adult male X. laevis specimen was reverse-transcribed using random hexamers and digested it with BglII, filling-in using Pfu polymerase, and religation

The resulting PCR fragments were subcloned into the pCR4-TOPO vector (Invitrogen). The complete coding region of chicken LEDGF/p75 cDNA was amplified using Expand DNA polymerase (Roche Applied Science) and primers: 5'-CAGCGCCGCGGCAACAG-3' and 5'-AGATTTCAATGGAATCACTTTG-3'. The primers used to amplify the frog cDNA were: 5'-TGGCGTGTATCTCCGAGGAGAAACTGCGAG-3' and 5'-AATGCAGAATTCAATTCCTTCTCTTCC. The resulting PCR fragment was digested with XhoI and cloned between NdeI and XhoI sites of pGEX-6P3, resulting in pCP-GSTHRP2. For expression of non-tagged full-length HRP2, the entire HRP2 ORF was amplified using primers 5'-CGGGTATCCGAGGAGAAACTGCGAG-3' and 5'-CCCTCCGAGGAGAAACTGCGAG-3' directly by the first codon of the relevant LEDGF/p75 fragment; anti

The BamHI/XhoI fragment of pCP-GSTHRP2 carrying the entire HRP2 ORF was PCR-amplified using 5'-Ultra DNA polymerase (Stratagene). Sense primers were designed to incorporate a BamHI restriction site followed by a Pfu primer with a restriction site and a Pfu primer incorporating a BamHI restriction site: 5'-CTGAGCATAAAAGGAG-3' and 5'-GTGGACGGATAGAAACG-5'. DNA Constructs for Bacterial Protein Expression—All gluthatone S-transferase (GST)-LEDGF/p75 fusion constructs used for protein expression in this work were based on the pGEX-4T1 vector (Amer sham Biosciences). The full-length LEDGF/p75 ORF and its fragments were PCR-amplified using Pfu-Ultra DNA polymerase (Stratagene). Sense primers were designed to incorporate a BamHI restriction site fol lowed by a Pfu primer with a restriction site and a Pfu primer incorporating a BamHI restriction site: 5'-CTGAGCATAAAAGGAG-3' and 5'-GTGGACGGATAGAAACG-5'. The protein secondary structure prediction and sequence analysis—Secondary structure prediction was done using the PROSsec and NetNucPred algorithms available through the PredictProtein server (cubic.bioc.columbia.edu) (31, 32). Hydrophobicity profiles (33) were analyzed using BioAnnotator software (Informax Inc.). Multiple sequence alignments were done with AlignX (Informax Inc.) using BLO SUM62 or GONNET matrices (34, 35). Homology between IBs and the N-terminal domain of transcription factor II S (TFIIS) was found using InterProScan release 7.2 (www.ebi.ac.uk/InterProScan/) (36) and SMART version 4.0 (smart.embl-heidelberg.de) (37).
Results

Conservation of LEDGF/p75 Protein—Sequences of several mammalian LEDGF/p75 orthologs were available in public sequence databases. We identified ESTs representing partial cDNA sequences of *G. gallus* and *X. laevis* LEDGF/p75 cDNAs, which allowed us to clone and sequence complete LEDGF/p75 cDNAs from these species. On the basis of the obtained cDNA sequences, chicken and frog LEDGF/p75 were predicted to be which larger than the 530-residue human ortholog. Alignment of the predicted amino acid sequences revealed ~48% identity between mammalian, avian, and amphibian LEDGF/p75 proteins (Supplementary Fig. S1). The plot in Fig. 1A between mammalian, avian, and amphibian LEDGF/p75 proteins (Supplementary Fig. S1). The plot in Fig. 1A

**N-terminal Microsequencing and Mass Spectrometry**—To determine the N-terminal residues of tryptic-resistant (TR) 1 and TR2 peptides, a mixture of digestion products was separated on a 2.1 × 250 mm Vydac C8 column. Fractions containing the TR1 and TR2 fragments were analyzed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS).

**GST Pull-down Assay**—Purified GST fusion proteins were adsorbed onto glutathione-Sepharose beads (Amersham Biosciences) in 200 mM NaCl, 5 mM Tris-HCl, pH 7.3, using 125 μl (settled volume) beads per 40 μg of protein. After 4 h at 4 °C, the beads were washed in excess buffer and stored on ice. To test for IN binding, 10 μl of glutathione-Sepharose beads carrying GST fusion proteins were resuspended in 200 μl of cold PD buffer (150 mM NaCl, 5 mM MgCl₂, 5 mM DTT, 0.1% Nonidet P-40, 25 mM Tris-HCl, pH 7.4) containing 10 μg of BSA. After addition of 3.8 μg of His₆-tagged HIV-1 IN the samples were gently rocked for 1.5–2 h at 4 °C and left for an additional 15–30 min without mixing. After careful aspiration of the supernatant, the settled beads were resuspended in 700 μl of fresh PD buffer, and allowed to sediment without centrifugation. The wash was repeated twice and bound proteins were eluted in SDS-containing sample buffer and analyzed by SDS-PAGE. In certain cases IN pull-down was confirmed by Western blotting using polyclonal anti-IN serum (42).

**Transfection and Immunoprecipitation**—293T cells were maintained in Dulbecco’s modified Eagle’s medium containing 10% fetal calf serum (Invitrogen), 5 units/ml penicillin and 5 μg/ml streptomycin. 293T cells grown in 6-well dishes to 30–50% confluency were transfected with 0.5 μg of pCypHA, pBHA-p75, or pCPHA-HRP2 along with 0.5 μg of pED-FLAG-IN per well using FuGENE 6 transfection reagent (Roche Applied Science). Twenty-four hours post-transfection, cells were washed in cold phosphate-buffered saline, and lysed in 400 μl of lysis buffer (50 mM NaCl, 0.5% Triton X-100, 50 mM HEPES pH 7.9, 5% glycerol, 25 mM β-glycerophosphate, 1 mM sodium orthovanadate, supplemented with complete protease inhibitor mixture (Roche Applied Science)). The extracts were centrifuged at 19,000 × g to remove cell debris and preclarified by incubation with 4 μl (settled volume) of protein G-Sepharose beads (Amersham Biosciences). Preclarified supernatants were incubated with 4 μg of mouse anti-HA 12CA5 antibody (Roche Applied Science) at 4 °C, 4 μl of protein G-Sepharose beads were added, and the samples were left rocking for an additional hour. The beads were washed three times in cell lysis buffer, four times in reduced salt buffer (cell lysis buffer modified to contain 150 mM NaCl, 0.1% Triton X-100, and 0.1% Nonidet P-40). Whole cell extracts and immunoprecipitated proteins were resolved in 4–20% gradient gels, transferred to nitrocellulose membranes, HA-tagged CypA, LEDGF/p75, and HRP2 proteins were detected by Western blotting using anti-HA 3F10 antibody conjugated to horseradish peroxidase (Roche Applied Science) and Western Lightning chemiluminescent reagent plus (PerkinElmer Life Sciences). FLAG-tagged IN was detected with anti-FLAG M2 antibody (Sigma-Aldrich) and goat anti-mouse IgG horseradish peroxidase conjugate (Jackson ImmunoResearch Laboratories).

**RESULTS**

Conservation of LEDGF/p75 Protein—Sequences of several mammalian LEDGF/p75 orthologs were available in public sequence databases. We identified ESTs representing partial cDNA sequences of *G. gallus* and *X. laevis* LEDGF/p75 cDNAs, which allowed us to clone and sequence complete LEDGF/p75 cDNAs from these species. On the basis of the obtained cDNA sequences, chicken and frog LEDGF/p75 were predicted to be composed of 579 and 564 amino acids, respectively, both somewhat larger than the 530-residue human ortholog. Alignment of the predicted amino acid sequences revealed ~48% identity between mammalian, avian, and amphibian LEDGF/p75 proteins (Supplementary Fig. S1). The plot in Fig. 1A summarizes this alignment by showing the degree of conservation along the protein sequence. Three regions of homology were evident (highlighted as shaded boxes in Fig. 1A). The most conserved fragment spanning residues 1–94 (conserved region 1), which showed about 89% identity between human, chicken, and frog, corresponded to the PWWP domain (22). A 105-residue region spanning residues 351–455 displayed about 87% identity (re-
In addition, a short fragment involving residues 178–197 (region II) showed significant homology. Intuitively, these most conserved regions likely represent functional and/or structural determinants within the protein. The most variable regions encompassed an internal fragment flanking the PWWP domain (residues 94–177 in human LEDGF/p75, showing only about 13% identity) and the 60 C-terminal residues of the protein (20% identity). The single conserved feature of the first hypervariable region was a 7-residue sequence, 146RRGRKRK152, which partially overlaps the NLS in human LEDGF/p75 (residues 148–156) (43). Both chicken and frog LEDGF/p75 contain an insertion of 39 amino acids within the first hypervariable region (Supplementary Fig. S1).

Secondary Structure Prediction—We used the protein structure prediction program package available through the PredictProtein server (31) to analyze possible structural elements in LEDGF/p75. The protein is uncommonly rich in charged amino acids, accounting for about 42% of its sequence. Thus, it was not surprising that two extensive loop regions were predicted by the NORSp program (residues 94–177 in human LEDGF/p75, showing only about 13% identity). The single conserved feature of the first hypervariable region was a 7-residue sequence, 146RRGRKRK152, which partially overlaps the NLS in human LEDGF/p75 (residues 148–156) (43). Both chicken and frog LEDGF/p75 contain an insertion of 39 amino acids within the first hypervariable region (Supplementary Fig. S1).

Limited Proteolysis of LEDGF/p75—We used limited proteolysis (45) to probe the domain organization of LEDGF/p75. As the protein is rich in charged amino acids, a cleavage site for trypsin is predicted on average every 4–5 residues. Considering all Lys and Arg residues, the largest hypothetical LEDGF/p75 tryptic peptide was just 25 residues (Thr477–Lys501) with a molecular mass of about 2.6 kDa. We found that recombinant human LEDGF/p75 was indeed very sensitive to trypsin. A mass ratio of 250:1 of LEDGF/p75:protease yielded final proteolyzed products as well as semi-stable intermediates (Fig. 2A). The protease was quenched at different time points by addition of PMSF and reaction products were analyzed using Tris-glycine or Tricine SDS-PAGE. As quantified by densitometry of Coomassie-stained gels, 60–70% of the protein became extinct after a relatively short exposure to trypsin (compare lanes 1 and 6 in Fig. 2A). As proteolysis proceeded, two distinct polypeptides TR1 and TR2 with apparent molecular masses close to 10 kDa gradually accumulated at the expense of the intermediate cleavage products (Fig. 2A). Both TR1 and TR2 fragments persisted even after overnight digestion under these conditions (data not shown).
N-terminal sequencing of TR1 and TR2 fragments revealed that TR1 was derived from the N terminus part of LEDGF/p75, having the same N-terminal sequence as the full protein, i.e. NH2-Met-Thr-Arg-Asp-Phe. TR2 originated from the C-terminal portion of the protein and contained two overlapping N termini: NH2-Lys-Arg-Glu-Thr-Ser-Met- and NH2-Glu-Thr-Ser-Met-Asp-Ser- corresponding to trypsin cleavage at peptide bonds Lys342-Lys343 and Arg344-Glu345, respectively. To identify the C termini of the fragments, TR1 and TR2 were purified by reverse phase high performance liquid chromatography and their masses were determined by MALDI-MS. The molecular mass of the TR1 fragment was 11,429 ± 10 Da. The TR2 product represented a mixture of fragments of 11,348 ± 10 and 11,632 ± 10 Da. These data allowed us to unambiguously map the C termini of the TR fragments to LEDGF/p75 residues Lys100 for TR1 and Lys442 for TR2. Indeed, the calculated molecular mass of Met1-Lys100 was 11,423.3 Da, whereas the masses of Lys342-Lys343 and Arg344-Glu345-Lys346 fragments were 11,641.2 and 11,356.8 Da, respectively, which matched the experimentally determined masses well within confidence intervals. When a deletion mutant retaining the 206 C-terminal residues of LEDGF/p75 (residues 326–530) was exposed to trypsin, only the TR2 fragment was obtained (Fig. 2B, lanes 3–7). This result confirmed that TR1 and TR2 resided in the N- and C-terminal regions of LEDGF/p75, respectively. Although more than a dozen potential trypsin cleavage sites exist within fragments 1–100 and 345–442 of LEDGF/p75, both appeared to resist proteolysis, indicating that both are involved in stable structures.

In addition to trypsin, we tested protease K, thrombin, chymotrypsin, and Arg C proteases (data not shown). Unlike trypsin, digestion with protease K did not result in stable proteolytic products, however transient fragments of about 10 kDa in size were observed. Incubation of GST-LEDGF/p75 with thrombin resulted in multiple cuts within the putative loop region adjoining the PWWP domain. Chymotrypsin and Arg C proteases appeared less active than trypsin and although the fragments obtained confirmed the trypsin map, the cleavage patterns were more complex and longer incubation times were necessary to allow for accumulation of final products.

**TR2 Is the Functional LEDGF/p75 IBD**—To identify region(s) of LEDGF/p75 involved in the interaction with HIV-1 IN, we prepared a series of LEDGF/p75 deletion mutants. Mutants were expressed and purified as GST fusions, pre-adsorbed onto glutathione-Sepharose beads, and tested for their ability to pull-down recombinant HIV-1 IN. As can be seen from Fig. 3A, both the full-length protein (residues 1–530) and the mutant lacking the variable 99 C-terminal residues (1–471) readily bound HIV-1 IN (Fig. 3A, lanes 9 and 12). However, a more extended deletion from the C terminus disrupted interaction with IN, as LEDGF(1–325) lacking 205 residues failed to pull down IN (lane 10). This result corroborates the previous finding that LEDGF/p52, an alternative splice form containing a unique 8-residue tail in place of LEDGF/p75 residues 326–530, did not bind HIV-1 IN (18). Furthermore, the C-terminal fragment of LEDGF/p75 containing residues 326–530 was sufficient to pull down HIV-1 IN (lane 11). By making another set of deletions, the IN binding function of LEDGF/p75 was mapped to just 83 amino acids, spanning residues 347–429 (Fig. 3B, lane 15; see also Supplementary Fig. S1). Importantly, this fragment lies within conserved region III of LEDGF/p75 (Fig. 1A) and the TR2 fragment (Fig. 2; see also Fig. 1B for summary). We found that further truncations from the N terminus of 347–429 abolished the interaction with IN (lane 16) and reduced the solubility of the recombinant protein (data not shown). Deletions from the C terminus of this fragment, on the other hand, profoundly affected stability of GST fusion proteins in E. coli (data not shown). These observations indicated that residues 347–429 of LEDGF/p75 span the IBD and comprise the minimal sequence required for its proper folding.

Of note, full-length LEDGF/p75, as well as deletion mutants containing the interdomain region (residues 150–325) were only marginally stable when expressed in bacteria, even when the temperature of induction was reduced. The bulk of the GST fusions recovered by adsorption to glutathione-Sepharose represented various proteolytic fragments. Due to dimerization of GST, it was not feasible to completely remove proteolytic fragments from preparations of GST-LEDGF/p75.
or fusions with LEDGF-(1–325) or LEDGF-(1–471) even after additional heparin affinity and cation exchange chromatography (Fig. 3A).

Identification of HRP2 as a Second IBD-containing Protein—Using translated BLAST to search for human cDNAs encoding polypeptides with homology to the LEDGF/p75 IBD we found that a second HDGF-related protein, HRP2, contains a very similar sequence within its C-terminal region. Because this region of homology is relatively short and occurs within largely divergent sequences, the similarity within C-terminal regions of LEDGF/p75 and HRP2 remained unnoticed until now. Fig. 4A presents an alignment of the human LEDGF/p75 IBD with the related sequence from HRP2 and includes their respective orthologs from different species. Human LEDGF/p75 and HRP2 proteins are about 49% identical within this region, and, considering conservative amino acid substitutions, the similarity exceeds 70%. Furthermore, predicted secondary structural elements within the two putative IBDs matched very well, with both domains demonstrating high α-helical content (Fig. 4A).

We identified several ESTs encoding an HRP2 ortholog from D. rerio, which allowed us to clone and sequence its complete coding region. In addition, HRP2 cDNA from X. laevis could be completely reconstructed from available ESTs (see “Experimental Procedures”). Sequence alignment of human, frog, and fish HRP2 revealed high degrees of sequence conservation within the PWWP and IBD-like regions (regions I and III, Fig. 4B) (for a complete alignment see supplementary Fig. S2). An approximate 20-amino acid region of homology (region II) was similar to homology region II in LEDGF/p75 and HRP2 IBDs, respectively, the N-terminal domain of both containing stretches of highly conserved Pro, Arg, and Lys residues. HRP2 region IV, however, appears unique to this protein. In addition, we also identified a hypothetical 475-residue protein CG7946 from Drosophila melanogaster (GenBank™ accession NP_651768, UniGene cluster Dm.4512) that contains an IBD-related sequence. This fragment, spanning CG7946 residues 318–400, shared about 21% identical and 48% similar residues with the HRP2 IBD (not shown). Intriguingly, since this protein is also predicted to possess an N-terminal PWWP domain, it likely represents an insect ortholog of HRP2. Additional searches using InterProScan and SMART (Simple Modular Architecture Research Tool) revealed homology between the IBDs and the N-terminal domain of TFIIS (SMART accession SM00509). Although the E-values reported by SMART for these hits were relatively high, equating to 3.1 and 1.2 for human LEDGF/p75 and HRP2 IBDs, respectively, the N-terminal domain of TFIIS seems to represent their closest relative among known protein domains. The TFIIS domain family includes four-helix bundle domains of TFIIS, elongin A, and CRSP70 (46).

To find out whether the putative IBD of HRP2 has affinity for HIV-1 IN, we fused a fragment spanning HRP2 residues 470–593 to GST and tested it in our pull-down assay. As seen in Fig. 5A, the HRP2 fragment readily interacted with recombinant HIV-1 IN, suggesting that HRP2 contains a functional IBD. To determine if HRP2 protein interacted with HIV-1 IN in human cells, 293T cells were transiently transfected with a plasmid encoding FLAG-tagged HIV-1 IN and a second plasmid encoding either HA-tagged LEDGF/p75, HRP2, or CypA. Cell extracts prepared 24 h post-transfection were immunoprecipitated with anti-HA 12CA5 antibody, and recovery of FLAG-IN was readily co-immunoprecipitated with HA-tagged LEDGF/p75 (lane 5). Significantly lower, but detectable amounts of FLAG-IN were recovered with HA-HRP2 (lane 6).

Significantly lower, but detectable amounts of FLAG-IN were recovered with HA-HRP2 (lane 6).

In contrast, only negligible binding of FLAG-IN was detected with HA-CypA, which served as negative control (lane 4). We estimated that recovery of FLAG-IN upon co-immunoprecipitation with HA-HRP2 was 5–10-fold lower than with HA-LEDGF/p75, suggesting that at least under the conditions of this experiment, LEDGF/p75 had a greater affinity for HIV-1 IN than did HRP2.

Stimulation of HIV-1 IN Activity by LEDGF/p75 and HRP2—In accordance with previously reported results (16), recombinant LEDGF/p75 potently stimulated HIV-1 IN strand transfer.
activity in the absence of organic solvents and polyethylene glycol (lanes 1–7, Fig. 6A). The DNA substrate was a linearized plasmid containing HIV-1 U3 and U5 sequences at its termini.
FIG. 6. Stimulation of HIV-1 IN strand transfer activity by LEDGF/p75 and HRP2. A, substrate DNA was preincubated with (+) 0.8 μM or without (−) IN for 7 min at room temperature. LEDGF/p75 (lanes 3–7) or LEDGF-(347–471) (lanes 8–12) was added at the indicated final concentrations. The reaction mixtures (20 μl) contained 150 ng mini-HIV DNA, 110 mM NaCl, 5 mM MgCl₂, 10 mM DTT, 2 μM ZnCl₂, 10 mM Hepes, pH 7.45. The DNA substrate was prepared as in Ref. 47. Following incubation for 90 min at 37 °C the reactions were stopped by addition of 25 mM EDTA and 0.5% SDS. Products were treated with 0.5 μg/μl proteinase K (Roche Applied Sciences) for 45 min at 50 °C, ethanol-precipitated, re-dissolved in Tris-EDTA and separated in 0.8% agarose. The gel was stained with ethidium bromide. Migration of the substrate DNA, strand transfer products, and molecular mass markers (23.1, 9.4, 6.6, and 4.4 kb) are indicated. B, inhibition of LEDGF/p75-dependent integration by diketo acid L-731,988. L-731,988, diluted in 50% dimethyl sulfoxide, was added to samples 5–9 at the indicated final concentrations. IN (lanes 2 and 4–9) was preincubated with substrate DNA and drug for 7 min at room temperature. Next, LEDGF/p75 was added to samples in lanes 3–9, and the reactions proceeded for 90 min at 37 °C. The composition of the final reaction mixtures was as explained in Fig. 6A, however all reactions were adjusted to contain 1% dimethyl sulfoxide. Reaction products were treated and visualized as in Fig. 6A. C, IN (0.6 μM) was added to lanes 2, 4, and 6–14. LEDGF/p75 (0.2 μM) was present in lane 4. Full-length GST-LEDGF/p75 (lanes 5–8), GST-LEDGF-(1–325) (lanes 9–11) or GST-LEDGF-(326–530) (lanes 12–14) were added at the indicated concentrations. D, LEDGF-(347–471) (lanes 5–7) or GST (lane 9) were added together with full-length protein as competitors at the indicated concentrations. E, purified LEDGF/p75, HRP2, and LEDGF-(347–471) (∼7.5 μg each) were separated by 4–20% SDS-PAGE. The gel was stained with Coomassie R250. Positions of molecular mass markers are indicated in kDa. F, reactions were set up as in Fig. 6A, and supplemented with HRP2 or LEDGF/p75 at the indicated concentrations. Following 90 min at 37 °C the reactions were stopped by addition of 700 mM NaCl, 20 mM EDTA and extracted twice with phenol/chloroform (1:1). The products precipitated with ethanol were dissolved in Tris-EDTA and separated in 0.8% agarose. G, LEDGF-(347–471) (lanes 5–8) or GST (lane 9) were added as competitors. Reaction products were treated and visualized as in Fig. 6F.
Sequence alignments, in silico secondary structure prediction, and limited proteolysis collectively suggest that LEDGF/p75 contains a pair of small structural domains: an N-terminal PWWP domain (residues 1–90), the existence of which had been recognized on the basis of sequence homology, and a novel domain that mediates interaction with HIV-1 IN. Remarkably, these two domains encompass only about 35% of the protein sequence. Recombinant LEDGF/p75 displays high sensitivity to proteolysis suggesting that a large portion of the protein exists as flexible regions or loops. Of note, we did not detect a stable interaction between the PWWP and IBD domains in a GST pull-down assay (data not shown), suggesting that the domains are relatively independent in the full-length protein. We think that such flexibility might be related to the function of the protein in vivo, allowing the domains to associate with and link together components of various complexes. Two putative loop regions, with no regular secondary structure were suggested by in silico analysis of LEDGF/p75 (Fig. 1B). Interestingly, proteins containing extended loops are statistically associated with transcription regulatory functions (48). In addition to the PWWP and the IBD domains, an internal 20-residue fragment of LEDGF/p75 (residues 178–197) displayed significant sequence conservation (region II, Fig. 1A). This 20-amino acid fragment contains five Pro residues and is thus unlikely to adopt an independent secondary structure. High Pro and Arg content makes it similar to the AT hook motif of the HMGA proteins. Due to the recognized sequence conservation of region II, we speculate that it is important for LEDGF/p75 function. One likely possibility is that it represents a part of the DNA binding determinant of LEDGF/p75.

The LEDGF/p75 IBD is comprised of about 80 residues and is predicted to fold into four or five a-helices (Figs. 1B and 4). The minimal fragment that bound HIV-1 IN via GST pull-down spanned residues Ser347–Val429. This is in agreement with a minimal fragment that bound HIV-1 IN via GST pull-down assay (data not shown), suggesting that the domains are relatively independent in the full-length protein. We think that such flexibility might be related to the function of the protein in vivo, allowing the domains to associate with and link together components of various complexes. Two putative loop regions, with no regular secondary structure were suggested by in silico analysis of LEDGF/p75 (Fig. 1B). Interestingly, proteins containing extended loops are statistically associated with transcription regulatory functions (48). In addition to the PWWP and the IBD domains, an internal 20-residue fragment of LEDGF/p75 (residues 178–197) displayed significant sequence conservation (region II, Fig. 1A). This 20-amino acid fragment contains five Pro residues and is thus unlikely to adopt an independent secondary structure. High Pro and Arg content makes it similar to the AT hook motif of the HMGA proteins. Due to the recognized sequence conservation of region II, we speculate that it is important for LEDGF/p75 function. One likely possibility is that it represents a part of the DNA binding determinant of LEDGF/p75.

Sequence alignments, in silico secondary structure prediction, and limited proteolysis collectively suggest that LEDGF/p75 contains a pair of small structural domains: an N-terminal PWWP domain (residues 1–90), the existence of which had been recognized on the basis of sequence homology, and a novel domain that mediates interaction with HIV-1 IN. Remarkably, these two domains encompass only about 35% of the protein sequence. Recombinant LEDGF/p75 displays high sensitivity to proteolysis suggesting that a large portion of the protein exists as flexible regions or loops. Of note, we did not detect a stable interaction between the PWWP and IBD domains in a GST pull-down assay (data not shown), suggesting that the domains are relatively independent in the full-length protein. We think that such flexibility might be related to the function of the protein in vivo, allowing the domains to associate with and link together components of various complexes. Two putative loop regions, with no regular secondary structure were suggested by in silico analysis of LEDGF/p75 (Fig. 1B). Interestingly, proteins containing extended loops are statistically associated with transcription regulatory functions (48). In addition to the PWWP and the IBD domains, an internal 20-residue fragment of LEDGF/p75 (residues 178–197) displayed significant sequence conservation (region II, Fig. 1A). This 20-amino acid fragment contains five Pro residues and is thus unlikely to adopt an independent secondary structure. High Pro and Arg content makes it similar to the AT hook motif of the HMGA proteins. Due to the recognized sequence conservation of region II, we speculate that it is important for LEDGF/p75 function. One likely possibility is that it represents a part of the DNA binding determinant of LEDGF/p75.

The LEDGF/p75 IBD is comprised of about 80 residues and is predicted to fold into four or five a-helices (Figs. 1B and 4). The minimal fragment that bound HIV-1 IN via GST pull-down spanned residues Ser347–Val429. This is in agreement with a previous report that LEDGF/p52 protein lacking residues 326–530 neither bound HIV-1 IN in vitro nor co-localized with it in live cells (18). Intriguingly, we identified a homologous sequence within another HDGF-related protein, HRP2, which likewise displayed affinity for HIV-1 IN. Thus, in addition to the N-terminal PWWP domains, LEDGF/p75 and HRP2 share conserved C-terminal domains, suggesting a close evolutionary and probable functional relationship between these proteins. Although we did not analyze susceptibility of HRP2 to proteases, analysis of its predicted amino acid sequence suggests that domain organization is similar to that of LEDGF/p75. Alignment of HRP2 orthologs from mammalian, amphibian, and fish sources showed a high degree of sequence conservation between orthologs compared with that of LEDGF/p75 (see Supplementary Table I and Figs. S1 and S2). In silico analysis of amino acid sequences of other HRPs suggest that although they do not possess IBD-like domains, a-helical elements are located within C-terminal regions of HDGF and HRP1 (data not shown), suggesting the presence of a second functional domain within these proteins as well.

Like HDGF, all HRPs seem to have mitogenic activity in cell culture (21, 25, 30). It is presently unclear whether the growth factor activity of such proteins that lack classical secretory signals is related to their functions in vivo (20). The original observation that LEDGF/p75 co-purified from HeLa nuclear extracts together with the transcription co-activator PC4 provided a clue that the protein might be involved in transcription regulation (50). More recently, LEDGF/p75 was reported to bind to heat shock and stress-related elements within promoters regions of the AOP2, Hsp27, and αB-crystallin genes and trans-activate their expression (28, 29). Although an earlier study isolated LEDGF/p75 from a lens epithelial cDNA library, expression of the protein is clearly not limited to lens. In contrast to the protein’s name, cDNA clones encoding LEDGF/p75 have been isolated from a wide range of primary and transformed mouse and human tissues at all stages of development (refer to EST collections associated with the UniGene entries from Supplementary Table I). Sequences derived from 215 cDNA clones suggesting several alternative LEDGF splice variants exist in the AceView data base (for up to date information consult www.ncbi.nlm.nih.gov/IEB/Research/Acembly). While the most abundant splice form, supported by 170 cDNA clones, encodes for LEDGF/p75, only 12 cDNAs are derived from p52 mRNA. Although a detailed expression analysis of individual splice forms will require a specialized study, it would appear that LEDGF/p75 is the dominant protein product of the PSIP1 gene in most tissues.

According to the large numbers of human and mouse ESTs corresponding to LEDGF/p75 and HRP2, these proteins are ubiquitously expressed at relatively high levels (see Supplementary Table I). Although the HRP2 IBD displayed an apparent high affinity for HIV-1 IN by GST pull-down (Fig. 5A), results of co-immunoprecipitation experiments suggested that LEDGF/p75 was a more potent IN interactor than was full-length HRP2 in human cells (Fig. 5B). This was not entirely unexpected, as depletion of endogenous LEDGF/p75 alone by siRNA efficiently disrupted the nuclear and chromosomal accumulation of HIV-1 and FIV IN in cells (18, 19). However, LEDGF/p75 and HRP2 proteins stimulated HIV-1 IN to a comparable degree in vitro (Fig. 6F). Based on this result we speculate that binding of IN to HIV-1 cDNA termini might stabilize the HRP2-IN interaction. HRP2 could potentially explain the failure of persistent siRNA-mediated knockdowns of LEDGF/p75 to reduce viral replication (19). It would also be interesting to determine if LEDGF/p75 and/or HRP2 modulate the enzymatic activity of FIV and other retro/lenti-viral INs (19).

It was demonstrated that HIV-1 displays a significant bias toward integration into active genes (51, 52). Somewhat similar, but not identical integration specificity was observed for murine leukemia virus, which prefers to integrate within transcription start regions in the human genome (52). On a practical level, specificity for integration within or near active genes poses a problem in developing retroviral vector-based gene therapies (53). Distant relatives of retroviruses, yeast retrotransposons present the best studied paradigm of targeted integration in eukaryotes (reviewed in Ref. 54). At least in the case of the Ty5 retrotransposon, a specific interaction between Ty5 IN and the chromosomal protein Sir4p determines the specificity of retrotransposition into silent chromatin (55, 56).
Integration of another yeast retrotransposon, Ty3, which has a preference for RNA polymerase III transcription start sites, is controlled by a TFIIB transcription factor complex, although the interacting determinant on the retrotransposon side is not known (57). Putative chromodomains were identified in the C-terminal regions of INs from many LTR retrotransposons, such as fungal Crl1 and Skippy, and were hypothesized to mediate the targeting of their integration (58). In this context, a model involving a chromatin binding protein as a targeting factor for reoviral integration seems quite plausible. LEDGF/p75, a chromosomal protein and a putative regulator of transcription that binds lentiviral INs in live cells, represents such a candidate factor (16–19). Identification of LEDGF/p75 as a candidate factor (16–19). Identification of LEDGF/p75 as a component of HIV-1 PICs encourages further research, as it remains to be seen whether LEDGF/p75 and/or its close relative HRp2 play role(s) in PIC formation or targeting during retroviral infection (19).

Acknowledgments—We thank G. Maertens for constructing the pBHA-p75 plasmid and for critical reading of the manuscript, T. Ellenberger and O. Tsodikov (Harvard Medical School) for advice on LEDGF/p75 domain mapping and use of their fermentation facility, and T. Ellenberger and O. Tsodikov (Harvard Medical School) for advice on LEDGF/p75 domain mapping and use of their fermentation facility, and J. Obenchain (Merck Research Laboratories) for L-731,988. We are grateful to J. Walter and C. Cvejic (Harvard Medical School) for providing us with X. laevis tissues, T. Look and T. Liu (Dana-Farber Cancer Institute) for their generous gift of D. rerio RNA samples, J.-M. Buerstedde (Heinrich-Pette Institute for Experimental Virology and Immunology) for the G. gallus DT40 cell line and H. Göttinger (Dana-Farber Cancer Institute) for the HA-CypA expression construct. We thank J. Lee and E. Gillepiec from the Molecular Biology Core facility at the Dana-Farber Cancer Institute for HPLC, MALDI MS and Edman analyses of our samples and J. Gibson-Brown (Washington University) for sharing the supporting raw data file for the lizard CF775831 EST.

REFERENCES

1. Craigie, R. (2001) J. Biol. Chem. 276, 23213–23216
2. Asante-Appiah, E., and Shalika, A. M. (1997) Antiviral Res. 36, 139–156
3. Brown, P. O. (1997) in Retroviruses (Collins, J. M., Hughes, S. H., and Varma, H. E., eds) pp. 161–203, Cold Spring Harbor Laboratory, Cold Spring Harbor
4. Nakajima, N., Lu, R., and Engelman, A. (2001) J. Virol. 75, 7944–7955
5. Lefemina, R. L., Schneider, C. L., Robbins, H. L., Callahan, P. L., LeGrew, K., Roth, E., Schleif, W. A., and Emini, E. A. (1992) J. Virol. 66, 7414–7419
6. Harada, D. J., Felcik, P., Wimmer, M., Wolfe, A., Stillmack, K., Grobler, J. A., Espeel, A., Gubreskis, L., Schleif, W., Blaz, C., and Miller, M. D. (2000) Science 287, 646–650
7. Rice, P. A., and Baker, T. A. (2001) Nat. Struct. Biol. 8, 302–307
8. Engelman, A. (1999) Adv. Virus Res. 52, 411–426
9. Wu, X., Liu, H., Xiao, H., Conway, J. A., Hehl, E., Kelapana, G. V., Prasad, V., and Kappes, J. C. (1999) J. Virol. 73, 2126–2135
10. Engelman, A. (2003) Curr. Top. Microbiol. Immunol. 281, 239–238
11. Kelapana, G. V., Marmon, S., Wang, W., Crabtree, G. R., and Goff, S. P. (1994) Science 266, 2002–2006
12. Willetts, K. E., Rey, F., Agostini, I., Navarro, J. M., Baudat, Y., Vigne, R., and Muesing, M. A. (2002) J. Biol. Chem. 277, 25920–25928
13. Limon, A., Devroe, E., Lu, R., Ghory, H. Z., Silver, P. A., and Engelman, A. (2002) J. Virol. 76, 10598–10607
14. Mulder, L. C., Chakrabarti, L. A., and Muesing, M. A. (2002) J. Biol. Chem. 277, 27489–27494
15. Violet, S., Hone, S. S., Rakotobe, D., Petit, C., Gay, B., Moreau, K., Billand, G., Prieur, S., Sire, J., Schwartz, O., Moscatedel, J. F., and Boulanger, P. (2003) J. Virol. 77, 12507–12502
16. Cheparpan, O., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engleborghs, Y., De Clercq, E., and Debryzer, Z. (2003) J. Biol. Chem. 278, 377–383
17. Turlure, F., Devroe, E., Silver, P. A., and Engelman, A. (2004) Front. Biosci. 9, 3187–3208
18. Maertens, G., Cheparpan, O., Pluymers, W., Busschaets, K., De Clercq, E., Debryzer, Z., and Engleborghs, Y. (2003) J. Biol. Chem. 278, 33528–33539
Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-derived Growth Factor/Transcriptional Co-activator p75 (LEDGF/p75) That Binds HIV-1 Integrase

Peter Cherepanov, Eric Devroe, Pamela A. Silver and Alan Engelman

J. Biol. Chem. 2004, 279:48883-48892.
doi: 10.1074/jbc.M406307200 originally published online September 14, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M406307200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2004/11/15/M406307200.DC1

This article cites 58 references, 30 of which can be accessed free at
http://www.jbc.org/content/279/47/48883.full.html#ref-list-1