Zinc-finger endonuclease targeting PSIP-1 inhibits HIV-1 integration

Roger Badia, Eduardo Pauls, Eva Riveira-Munoz, Bonaventura Clotet, José A. Esté* and Ester Ballana

IrsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, 08916 Badalona, Spain.

Running title: ZFN targeting PSIP1 inhibits HIV-1 integration

* Corresponding author mailing address:
José A. Esté
IrsiCaixa, Hospital Germans Trias i Pujol, C. Canyet s/n, 08916 Badalona, Spain
Phone: 34 934656374
FAX: 34 934653968
E-mail: jaeste@irsicaixa.es

Key words
Zinc finger endonuclease, Ledgf/p75, integrase, HIV-1
Genome editing using zinc-finger nucleases (ZFN) has been successfully applied to disrupt CCR5 or CXCR4 host factors, inhibiting viral entry and infection. Gene therapy using ZFN to modify PSIP1 gene, encoding for LEDGF protein, might restrain an early step of viral replication cycle at the integration level. ZFNs targeting the PSIP1 gene (ZFN_{LEDGF}) were designed to specifically recognize the sequence after the integrase binding domain (IBD) of LEDGF/p75 protein. ZFN_{LEDGF} successfully recognized the target region of the PSIP1 gene in TZM-bl cells by heteroduplex formation and DNA sequence analysis. Gene editing induced a frame shift of the coding region and resulted in the abolishment of LEDGF expression at mRNA and protein level. Functional assays revealed that infection with HIV-1 R5 BaL or X4 NL4-3 viral strains was impaired in LEDGF/p75 knock-out cells regardless of entry tropism, due to a blockade in HIV-1 provirus integration into host genome. However, residual infection was detected in LEDGF knock-out cells. Indeed, LEDGF knock-out restriction was overcome at high multiplicity of infection, suggesting alternative mechanisms for HIV-1 genome integration rather than LEDGF/p75. Observed residual integration was, however, sensitive to the integrase inhibitor Raltegravir. These results demonstrate that the described ZFN_{LEDGF} effectively targets PSIP1 gene which is involved at early steps of viral replication cycle and thus ZFN_{LEDGF} may become a potential antiviral agent to restrict HIV-1 integration. Moreover, LEDGF knock-out cells represent a potent tool to elucidate the role of HIV integration cofactors in virus replication.
INTRODUCTION

Human immunodeficiency virus (HIV) requires the host cellular machinery to successfully replicate (1). Developing of genome editing tools, such as zinc finger endonucleases (ZFNs) (2), transcription activator-like [TAL] effector nuclease (TALEN) (3) or clustered regulatory interspaced short palindromic repeat (CRISPR) (4-6) become a promising alternative to modify essential host factors along the replication cycle of HIV (7, 8). ZFNs have demonstrated their applicability to reproduce in vitro the CCR5Δ32 phenotype by successfully cleaving the CCR5 gene, generating human CD4+ T cells refractory to HIV-1 infection (9-12). Similarly, ZFNs approach successfully cleaved the alternative HIV-1 coreceptor CXCR4 in CD4+ T cells from humanized mice model resulting in impaired HIV-1 infection (13). Genome editing as anti-HIV therapy is currently under study in at least 2-3 clinical trials using ZFNs targeting CCR5. However, similar strategies targeting host cellular factors affecting later steps of the virus replication cycle have not been evaluated.

A crucial step of the viral replication cycle is exerted by the lens epithelium-derived growth factor (LEDGF/p75), a member of the hepatoma-derived growth factor (HDGF) related protein (HRP) family. HRPs are characterized by a conserved N-terminal PWWP domain, an 90– to 135–amino acid module found in a variety of nuclear proteins (14). Six human HRP family members have been described: HDGF, HRP1, HRP2, HRP3, LEDGF/p75, and LEDGF/p52. Two of them, LEDGF/p75 and HRP2, possess affinity for HIV-1 integrase (IN), given by a second evolutionary conserved domain within their C-termini that mediates the interaction with HIV-1 IN, hence the term “IN-binding domain (IBD)”(15). Initially identified as IN associated protein (16), LEDGF/p75 was revealed as a lentivirus-specific cellular cofactor required for HIV integration into the host genome (see references (17, 18) and (19) for review). LEDGF/p75 directly interacts with viral HIV-IN, tethering viral preintegration complex into active transcription units of the cellular chromatin. The role of LEDGF/p75 in HIV-1 replication was studied using RNA interference (RNAi) targeting LEDGF/p75 and LEDGF KO murine embryonic fibroblasts (MEF). Although both strategies potently downregulate or completely abolished LEDGF/p75 expression, residual replication was observed. Thus, all studies point to a key but not essential role for LEDGF/p75 in lentiviral replication and suggested that the existence of alternative cellular cofactors, such as HRP2, were responsible for the residual replication observed in the absence of LEDGF/p75 (20, 21).
Nevertheless, the LEDGF/p75 interaction with HIV-IN has been suggested as valid target for antiviral therapy (22-24). In that sense, recently developed allosteric LEDGF/p75-IN interaction inhibitors (LEDGINs and ALLINIS) have been proved to target the LEDGF/p75 binding pocket of HIV-IN and to inhibit the catalytic activity of the IN. Moreover, LEDGINs and ALLINIS also exert antiviral activity by promoting IN multimerisation. Aberrant IN complexes lead to the formation of defective regular cores during the maturation process, resulting in an impaired the infectivity of the new viral particles (25-27).

On the other hand, a series of PSIP1 single nucleotide polymorphisms (SNP) were associated to HIV-1 disease progression in cohorts of African and Caucasian HIV-1 positive individuals (28) (29). In addition, two missense mutations were identified in two samples belonging to a LTNP cohort (30). All missense mutations identified are located in the helix-turn-helix (HTH) motifs at the C-terminal region of the protein, after the IBD domain. Although none of the mutations restricted HIV replication in vitro (29, 31, 32), these findings suggested that genetic variation in PSIP1 may influence susceptibility to HIV-1 infection and disease progression.

Here, we describe a novel genome editing ZFN that specifically disrupt the PSIP1 gene encoding LEDGF/p75 in its C-terminus, after all relevant functional domains and nearby the missense mutations described in patients (ZFNLEDGF). ZFNLEDGF was able to generate LEDGF/p75 cells expressing a truncated protein that become refractory to HIV-1 integration. Generated LEDGF/p75 knock-out cells represent a potent tool to further investigate the function of LEDGF/p75 protein and may help to elucidate the role of HIV integration cofactors in virus replication. Moreover, the ZFNLEDGF may become a potential antiviral strategy to restrict HIV-1 integration and virus replication in vivo.
MATERIALS AND METHODS

Vectors. CompoZr™ Knockout Zinc Finger Nucleases targeting the PSIP1 gene (ZFN_{LEDGF}) were obtained from Sigma-Aldrich Biotechnology (St. Louis, Missouri, USA). Briefly, ZFN_{LEDGF} were designed to target the sequence AACATGTTCTTGGTTGGTGAAGGAGATTCCGTG (Fig. 1A), according to Mussolino and Cathomen guidelines (33) to ensure minimal homology of ZFN DNA-binding domain to any other site in the genome. The FokI and zinc finger domains of each ZFNs pair were assembled as previously described (34). Next, ZFN_{LEDGF} pairs were cloned into the pLVX-IRES-ZsGreen1 vector (Clontech, Takara Bio Company), by replacing the EcoRI-XbaI fragment, to obtain ZFN_{LEDGF} tagged with the green fluorescent reporter gene (pLVX-ZFN_{LEDGF}-ZsGreen).

Cells. Human K562 cells were obtained from ATCC (CCL-243) and grown in Iscove’s Modified Dulbecco’s Medium, supplemented with 10% FBS and 2 mM L-glutamine. TZM-bl cell line (NIH AIDS Research and Reference Program), was grown in Dulbecco’s modified Eagle’s medium (DMEM, Gibco, Life Technologies, Madrid, Spain) supplemented with 10% of heat-inactivated foetal calf serum (FCS, Gibco, Thermofisher, Madrid, Spain) and antibiotics 100 U/ml penicillin, 100 μg/ml streptomycin (Life Technologies) and maintained at 37ºC in a 5% CO2 incubator.

ZFN Transfection. TZM-bl cells were transfected with ZFN_{LEDGF} expressing plasmids as described previously (35, 36). Briefly, 1.5x10^5 cells were seeded in 24 well plates. After overnight culture, 0.5μg of each ZFN_{LEDGF} plasmids was mixed with Lipofectamine 2000 reagent (Invitrogen) in serum-free medium OptiMEM (Invitrogen) and then added to previously washed cells. Media was replaced by fresh DMEM 4 hours after transfection and left in the incubator for 3 days, when ZsGreen positive cells were sorted were sorted using a FACSAria II (BD Biosciences) flow cytometer. Single cell clones were obtained by limiting dilution seeding of sorted cells in 96-well plates.

Analysis of PSIP1 disruption by Cel-I Assay. The effect of ZFN on PSIP1 alleles was assessed by performing PCR surrounding ZFN target site followed by digestion with the Surveyor (Cel-1) nuclease assay (Transgenomic, Omaha, NE, USA), which cleaves DNA heteroduplex at mismatch sites. Generated fragments were resolved by 10% polyacrylamide electrophoresis as previously described (10, 12).
Sequence analysis of targeted PSIP1 gene in TZMbl cells. Genomic DNA from sorted cell clones was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Barcelona, Spain). Extracted DNA was used to amplify the LEDGF gene using Expand High Fidelity PCR System (Roche, Barcelona, Spain) and the following primers (Forward 5’- TTCAAGTCATGTGGA TTCTTTGA-3’ and Reverse 5’- TCTAGCTTTTGTGGCCC-3’). PCR products were cloned into the pGEM-T Easy Vector System (Promega, Madrid, Spain) according to the manufacturer’s instructions. Plasmid sequencing was carried out by the Macrogen Genomic Division, Seoul, Korea using ABI PRISM Big Dye TM Terminator Cycle Sequencing technology (Applied BioSystems), with the following internal primers of PSIP1 genomic sequence (Forward: 5’- TTGGAACGATCTTTAGAAACAGA-3’ and Reverse: 5’- CAGTGAAACTATGTATGAAAGCCATT-3’). When bigger deletions were observed, sequences were obtained directly from mRNA sequence (see above). Sequences were analyzed with the Sequencher 4.5 software (Gene Codes Corporation, Ann Arbor, MI USA). In addition, a bioinformatics algorithm (Sigma) was used to evaluate the putative off-target effects for ZFNLEDGF throughout the human genome. In silico prediction of mRNA and protein sequences of the edited cell clones was performed using the Expasy: SIB bioinformatics resource portal (37).

Quantitative RT-polymerase chain reaction (qRT-PCR) and mRNA expression assessment. Relative mRNA quantification of PSIP1 expression was assessed by qRT-PCR as previously described (38, 39). Briefly, RNA was extracted using the Qiagen RNeasy Mini Extraction kit (Qiagen) according to the manufacturer’s instructions, including the DNase I treatment step. Reverse transcriptase was performed using the High Capacity cDNA Reverse Transcription Kit (Life technologies, Madrid, Spain). To assess presence of the predicted PSIP1 mRNA truncated forms, its mRNA sequence was divided in two fragments of similar size, which were PCR amplified with the following primers for the 5’-containing fragment, Forward: 5’- GGCAAAACAAATAAAAAAGAAAAGG-3’ and Reverse: 5’- CTTGCTTTCGTTTTCGATCT-3’; and 3’-containing fragment, Forward: 5’- AAAAGGGGGGAGAATTCTTCA-3’ and Reverse: 5’- GCAGTCTATTCTAAATGAAAACCATT-3’. Relative gene expression of PSIP1 wild-type form was also measured by two-step quantitative RT-PCR and normalized to
GADPH gene expression using the delta-delta Ct method. Primers and DNA probes were purchased from Life Technologies.

Western blot. Treated cells were rinsed in ice-cold PBS, nuclear proteins were enriched using the CellLytic Nuclear Extraction Kit (Sigma) and extracts were prepared in lysis buffer (50 mM Tris HCl pH 7.5, 1 mM EDTA, 1 mM EGTA, 1 mM Na3VO4, 10 mM Na β-glycerophosphate, 50 mM NaF, 5 mM Na Pyrophosphate, 270 mM sucrose and 1% Triton X-100) supplemented with protease inhibitor cocktail (Roche) and 1 mM phenylmethylsulfonfluryl fluoride. Samples were run using NuPage 4-12% Bis-Tris gel (Novex, Lifetechnologies) and blotted onto nitrocellulose membranes. Blocked membranes were incubated overnight with monoclonal antibodies (mAbs) against human C-terminus LEDGF protein (C57G11, Cell Signalling Technologies), human N-terminus LEDGF protein (Clone 26, 611714, BD transduction laboratories) and β-actin (Sigma-Aldrich) at 4°C. After washing, the membranes were incubated with a secondary conjugated horseradish peroxidase antibody for 1 hour at room temperature and then revealed with SuperSignal West Pico Chemiluminescent Substrate (Pierce Chemical, Rockford, IL).

Virus production and infections. VSV-pseudotyped NL4-3 GFP-expressing virus (40) was produced as described previously (39). The HIV-1 viral strains BaL (R5-tropic) and NL4–3 (X4-tropic) were obtained from the MRC Centre for AIDS Reagents (London, UK). BaL and NL4-3 strains were grown in PBMCs or lymphoid MT-4 cell line, respectively. Both viral stocks were titrated for its use in TZM-bl cells. For infections, 1.5x10^4 TZM-bl cells were seeded in 96-well plates and infected with BaL and NL4-3 viral strains at a multiplicity of infection of 0.01. The CXCR4 antagonist AMD3100 (Sigma-Aldrich), the reverse transcriptase inhibitor 3-azido-3-deoxythymidine (zidovudine; AZT, Sigma-Aldrich) and the IN strand transfer inhibitor Raltegravir (Ral, Merck) were used as controls. For the β-galactosidase assays, cells were lysed 72 h after infection and kept frozen until the β-galactosidase determination. To assess ZFNLEDGF effect on viral integration, cells were infected with NL4-3 and after 8 h of infection, cells were lysed and DNA was harvested to measure viral DNA. To determine viral integration, cells were infected and 8 h after infection fresh growth medium supplemented with 10 μg/ml of the neutralizing anti-gp120 mAb IgGb12
(Polymun Scientific) was added and left in the incubator for 24 h, when DNA was harvested and stored at -20 ºC until the viral integrated DNA was determined.

**β-Galactosidase detection assay.** β-galactosidase activity in 30 μl cell extracts was quantified by a colorimetric assay as described elsewhere (39, 41). Absorbance (405-620 nm) of non-infected samples was subtracted from the rest of samples and values expressed as percentage of β-galactosidase activity relative to non-drug treated control.

**Viral and integrated DNA determination.** Viral DNA was extracted using a QiaAmp DNA extraction kit (QIAamp DNA Blood Mini kit; Qiagen). Total viral DNA was quantified by amplifying a Gag fragment as described elsewhere (primers, 5′CAAGCAGCCATGCAAATGTT-3′ [forward] and 5′-TGCACTGGATGCAATCTATCC-3′ [reverse]; probe, 5′-[FAM]AAAGAGACCATCAATGAGGAAGCTGCAGA[TAMRA]-3′) (39, 41). Integration was detected by Alu-long terminal repeat (LTR) preamplification (primers, 5′-GCCTCCCAAAGTGCTGGGATTACAG-3′ [forward, Alu] and 5′-AGGGTTCCCTTGCTTTGT-3′ [reverse, LTR] followed by gag qPCR.

**Statistical analyses.** Experimental data are presented as mean ± SD. Paired Student’s t test was used for comparison between two groups, using the GraphPad Prism software (GraphPad Software, San Diego, California, USA). A p-value of 0.05 was considered to be statistically significant.
RESULTS

Design and efficiency of the ZFN_{LEDGF}. ZFNs targeting PSIP1 were designed to specifically disrupt the C-terminal region where mutations in HIV patients have been identified, after the sequence coding for the IBD of LEDGF/p75 (Fig. 1 A) (30, 32). The endonuclease activity of the ZFN_{LEDGF} was determined in vitro in K562 cells and by assessment of heteroduplex formation in the surrounding region of the targeted DNA sequence within the PSIP1. Endonuclease activity in human K562 cells was estimated as 6.6% when transfecting plasmids encoding the ZFN, increasing to 12% when using ZFN mRNA. Applicability of ZFNs is hampered by the potential risk of off-site cutting events leading to undesired side-effects (42). Since off-target events have been reported for the previously described ZFNs targeting the CCR5 gene (43, 44), the possibility of off-site cutting events for the ZFN_{LEDGF} was in silico assessed. The in silico analysis predicted up to 19 potential off-target sites for ZFN_{LEDGF}, all with at least 4 mismatches compared to target sequence and corresponding to 9 coding regions distributed in 7 different chromosomes of the human genome (Suppl. Table 1).

To improve ZFN efficiency and speed up the selection of single cell clones edited by ZFN_{LEDGF}, the sequence encoding ZFN_{LEDGF} was cloned in plasmids carrying a ZsGreen reporter gene, whose expression was transient. After transfection, 15% of cells were ZsGreen positive, a population that was enriched to 95% of ZsGreen positive cells by cell sorting. Monoclonal cell populations of ZFN_{LEDGF}-treated cells were obtained by limiting dilution, resulting in 38 single clonal cell lines of which 6 presented slow growth kinetics and were discarded. Thus, further functional characterization was performed in 32 clonal cell lines.

The ZFN_{LEDGF} treated monoclonal cell lines were screened for its capacity to restrict single round HIV-1 infection by measuring the fluorescence induced by a VSV-pseudotyped NL4-3 GFP-expressing virus. Although different degrees of HIV infection were identified, 25 out of 32 cell lines blocked HIV replication over 50% compared to wild-type control cells, suggesting that PSIP1 gene has been effectively targeted and genomically edited (Table 1). Only in 7 cell lines residual or no effect on HIV replication was observed, representing an efficiency of 70% for the overall experimental procedure. Then, three of the cell lines were chosen for further validation, taking into account their different degrees of HIV infection impairment with a full replication virus (Fig. 1C). The deletions/insertions induced by ZFN_{LEDGF} were
additionally confirmed using the heteroduplex formation assay in the three selected cell lines (Fig. 1D).

**Molecular characterization of ZFN\textsubscript{LEDGF} edited cell lines.** To characterize the genomic defects introduced by ZFN\textsubscript{LEDGF}, the \textit{PSIP1} genomic sequence surrounding the target site of the selected monoclonal cell lines was amplified, cloned and sequenced. At least 16 different \textit{PSIP1} sequences were obtained from each selected cell line, sequences were aligned thus confirming that ZFN\textsubscript{LEDGF} successfully recognized and cleaved the target sequence. Two cell lines harbored large modifications in both alleles (LEDGF\textsuperscript{--} KO, cell lines 2 and 3), whereas another cell line, together with a large modification, presented a 3-bp deletion that led to a single aminoacid deletion at protein level, and thus resembled an heterozygote phenotype (hereafter LEDGF\textsuperscript{+/-}, cell line 1) (Fig. 2A). Sequencing analysis and alignment confirmed that ZFN\textsubscript{LEDGF} introduced genomic defects located in exon 13, in most cases after the IBD domain. The genomic defects detected were three deletions of 3-bp, 41-bp and 17-bp respectively, one insertion of 155 bp and two very large deletions that could not be well characterized at genomic level, but were characterized at mRNA level (Fig. 2A and data not shown). \textit{In silico} prediction of putative mRNA and protein sequences of LEDGF\textsuperscript{--} KO cell lines, suggested the presence of a truncated protein with conserved functional PWWP domain (both alleles of cell lines 1, 2 and 3) and IBD (cell line 1 – allele A and cell line 2 – allele B) but lacking the C-terminal part (Fig. 2B and 2C).

The effect of ZFN\textsubscript{LEDGF} on gene expression was assessed both at mRNA and protein LEDGF/p75 level by RT-qPCR and Western-blot, respectively. \textit{In silico} prediction of a truncated mRNA was confirmed by testing mRNA expression of 5’ and 3’ fragments of \textit{PSIP1}. Expression of \textit{PSIP1} 5’ mRNA fragment (from \textit{PSIP1} mRNA position +204 to +951) was detected in cDNA samples of all selected cell lines (Fig. 2D). Conversely, expression of \textit{PSIP1} 3’ mRNA (from \textit{PSIP1} mRNA position +864 to position *66 after the stop codon), which included the ZFN\textsubscript{LEDGF} cutting site, was not detected in the LEDGF\textsuperscript{--} KO cell lines with larger deletions (cell lines 2 and 3) but was present in cell line 1 resembling an heterozygote, compared to wild-type or mock transfected controls (Fig. 2D). Cell line 1 also presented a lower size mRNA fragment indicative of a large deletion, the boundaries of which cannot be characterized at genomic level. However, we identified an mRNA form lacking exons 12, 13 and 14 by
cloning and sequencing the cDNA products, indicating that the induced deletion expanded at least over the 700bp genomic region that includes all 3 exons (Fig. 2C).

To confirm mRNA truncation, quantitative analysis of gene expression corresponding to the C-terminal end of the PSIP1 was performed, confirming a significant decrease in C-terminal mRNA expression in ZFNLEDGF treated cells, compared to the wild-type and mock transfected cells (60% reduction in heterozygote cells, residual expression in homozygotes p<0.01, Fig. 2D). Similarly, when using an antibody specifically recognizing LEDGF/p75 C-terminus no protein expression was observed in cell lines 2 and 3 but was partly detected in cell line 1, further confirming that ZFNLEDGF induces the formation of a truncated sequence that leads to a potent abrogation of a complete LEDGF/p75 expression at mRNA and protein levels (Fig. 2D, 2E and 2F upper panel).

Deletions from the IBD C-terminus of LEDGF have been reported to affect protein stability and reduce solubility of the protein (15). Thus, to determine whether the predicted truncated proteins were effectively expressed, a monoclonal antibody recognizing the N-terminus region of LEDGF was used. Expression of full length p75 and p52 isoforms was detected in control cell lines (wild-type and mock transfected, Fig. 2F, lower panel, first two lanes), as well as in edited cell line 1 (forth lane), whereas only the p52 isoform was detected in cell lines 2 and 3 (third and last lanes), confirming the results obtained with the C-terminus antibody. However, a band of intermediate molecular weight (MW) appeared in cell lines 1 and 3 (red arrows, lower panel), which was in accordance with in silico predictions (predicted MW LEDGF/p75=60.1 KD; LEDGF/p52=37 KD and truncated proteins=47-48 KD, based on Expasy prediction tools (37)) and therefore suggested the expression of LEDGF truncated proteins. No clear protein expression was detected in cell line 2, suggesting that genomic modifications induced in this cell line compromised either protein stability, solubility and/or folding, as previously reported for other truncated LEDGF constructs (15).

HIV-1 infection. Functional assays to establish the effect of ZFNLEDGF were performed with the LEDGF+/− (cell line 1) and one homozygote LEDGF−/− KO (cell line 2). No differences in the growth kinetics of the cell line compared to control parental cells were observed (data not shown). LEDGF+/− and LEDGF−/− KO were challenged with either HIV-1 X4 tropic NL4-3 or R5 tropic BaL strains and infection was monitored by
β-gal production (Fig. 3). LEDGF +/- cell line slightly inhibited HIV-1 infection compared to wild-type and mock-transfected controls (24% and 45% inhibition for HIV-1 X4 NL4-3 and R5 BaL strain respectively, Fig. 3A and 3B). Conversely, HIV-1 infection in LEDGF +/- KO cells was strongly inhibited compared to wild-type and mock transfected controls and regardless of the viral strain used (up to a 70%, p<0.01, Fig. 3A and 3B). Nevertheless and in line with previous reports, infection was not completely abolished in the KO cell line and residual infection persisted, accounting for roughly 30% of infection for both R5 BaL and X4 NL4-3 viral strains (Fig. 3A and 3B). To evaluate the effect of antiviral compounds in the ZFN_{LEDGF} cell lines, antiviral activity of the X4 entry inhibitor AMB3100, the reverse transcriptase (RT) inhibitor AZT and the integrase inhibitor Raltegravir were determined (Fig. 3C and 3D). No significant differences were observed in the efficacy of the different compounds in any of the drugs tested. Interestingly, Raltegravir was able to inhibit residual replication observed in LEDGF +/- KO, consistent with previous reports (20) (21) and confirming the different mechanism of action of both strategies.

To further characterize the functionality of ZFN_{LEDGF}, viral and integrated viral DNA were determined in the LEDGF +/- KO cell line. No differences were observed in HIV-1 viral DNA between mock-transfected cells and LEDGF +/- KO cells when challenged with HIV-1 X4 NL4-3 (Fig. 4A). Conversely, integrated viral DNA was inhibited up to 87% in LEDGF +/- KO cells compared to mock-transfected control (p<0.001, Fig. 4B). Residual viral integration was also observed in the LEDGF +/- KO cells (aprox. 12% of integration, p<0.001, Fig. 4B), which could be fully inhibited by the IN inhibitor Raltegravir (Fig. 4B).

Taken together, these results suggest that a complete LEDGF/p75 protein, including the C-terminal domain, is necessary to successfully tether HIV pre-integration complex into active transcriptional units. Consistent with previous reports (20) (19) (21), LEDGF +/- KO cells were able to support inefficient but detectable viral integration and produce new viral particles confirming the presence of alternative pathways for HIV-1 replication in the absence of LEDGF/p75.
DISCUSSION

Genome editing is an emerging strategy to study virus-host interactions and to combat and cure HIV-1 infection (7, 8). An ideal therapy for HIV, or other chronic viral infections that course with latent reservoirs, is believed to involve the generation of a source of long-lived, self-renewing and multilineage hematopoietic stem cells that would repopulate the host with genetically modified cells refractory to infection (45, 46). Since the unique and exceptional case of HIV-1 sterilizing cure of a patient due to bone marrow transplantation with a matched donor homozygote for the CCR5Δ32 mutation (47, 48), alternative strategies have aimed to reproduce the CCR5Δ32 phenotype using genome-editing tools. Indeed, ZFNs targeting the HIV coreceptors gene have been successfully developed to generate human CD4+ T cells and human embryonic cell precursors and induced pluripotent stem cells which were refractory HIV-1 infection in different mice models (9-13).

Here, we have evaluated the feasibility and efficacy of generating LEDGF/p75 knockout cell, a key factor for the integration of viral DNA into the host genome, using ZFN targeting at the C-terminal region of the LEDGF/p75 protein, outside the best described functional domains PWWP and IBD. LEDGF/p75 has already been validated as a candidate for gene therapy in a model where engraftment of lentiviral transduced CD4+ T cells overexpressing LEDGF325-530 induced a 3-log reduction in plasma viral load of HIV-1 infected mice (23). Overexpression of the deficient mutant LEDGF325-530 in primary CD4+ T cells impeded but did not completely block viral replication, due to minimal wild type LEDGF/p75 expression. Thus, the use of ZNFLEDGF might be advantageous in gene therapy settings as it confers a permanent disruption of the target gene avoiding the presence of residual levels of the wild-type LEDGF/p75 form that might be hijacked by the HIV-IN to successfully replicate.

Here, the use of the ZFNLEDGF generated TZM-bl cell lines carrying a truncated form of PSIP1 at the genomic region encoding for the protein C-terminal region that preserves the N-terminal functional domains. Although C-terminal region of LEDGF/p75 is functionally poor characterized, several genetic variants in HIV infected patients have been identified outside the known functional domains, variants which might be related to different susceptibility to HIV infection and disease outcome (28-30). Although identified LEDGF/p75 variants support efficient HIV-1 infection ex vivo, the involved C-terminal amino acid positions are well conserved throughout the evolution, suggesting an important role for protein functionality (31, 32). The induction
of large deletions in the sequence of PSIP1 gene coding for the C-terminal region of the LEDGF/p75 resulted in a LEDGF−/− KO phenotype and provided a genetic barrier to HIV-1 in vitro infection.

Previous work reported a decrease in protein stability, expression levels, and solubility of recombinant LEDGF/p75 mutants lacking IBD and/or C-terminus (15). In accordance, stable expression of truncated proteins was detected in only 2 of the 5 edited alleles harbouring large modifications, pointing towards the need of LEDGF/p75 C-terminus domain for warranting protein stability and/or correct folding. Our data reinforce the relevance of the C-terminal end region for the LEDGF function and HIV infection outcome even if N-terminal functional domains and IBD of the LEDGF/p75 remain intact. The inhibition of the HIV-1 infection in our ZFN_{LEDGF} clones might seem less apparent comparing to previously described LEDGF/p75 knockout models using mouse embryonic fibroblasts (MEFs) (19) and/or LEDGF KO in human somatic Nalm-6 cells (pre-B acute lymphoblastic leukemia cells line) (20).

However, LEGDF/p75 KO were generated by the genetically modification at the exons 2 and 3 of the PSIP1 gene in the MEF model (19) and in human somatic Nalm-6 cells were obtained by homologous recombination of exons 11 to 14 containing the IBD (20). Contrary to previously models, our ZFN_{LEDGF} was not designed against the functional domains of the LEDGF protein, strengthening the relevance of the C-terminal end of the protein.

The modulation of LEDGF/p75 expression by different approaches involving RNA interference (49), short-hairpin RNA knockdown (50) and knockout models (19, 20), provided strong evidence of how LEDGF/p75 interacts with the HIV-IN and tethers HIV provirus into the host chromatin but also indicated that, in accordance to our results, residual HIV infection occurred even in absence of LEDGF/p75. Indeed, HRP2, which shares similar structural features with LEDGF/p75, has been postulated as an alternative factor triggering HIV integration (21, 51). Nevertheless, double PSIP1/HRP2 KO mice cells are still able to support HIV-1 integration (21), thus other alternative pathways cannot be excluded which might allow virus to overcome LEDGF/p75 deficiency. In that sense, and given that the TZM-bl cell line is a widely accepted model of HIV infection research, our recently developed TZM-bl LEDGF/p75−/− KO becomes a helpful tool to elucidate host factors involved in HIV integration.
In summary, we describe the generation of LEDGF/p75 knockout cells using a ZFN that successfully recognises and disrupts the sequence of the *PSIP1* gene coding for the C-terminal end of the LEDGF/p75 protein. The truncation of the C-terminal end of the LEDGF/p75 results in a reduced protein stability that lead to the generation of KO cells with an impaired HIV-1 replication independent of genetic modification concerning the N-terminal functional domains or the IBD of LEGDF protein. Further studies must be carried out to elucidate the functional role of genetic variants in the coding regions of the *PSIP1* gene *in vivo*. Our results confirm previous data indicating that other pathways rather than LEDGF/p75 might allow HIV integration. Finally, the ZFN$_{LEDGF}$ provides a new cellular model to study host factors involved in the HIV-1 integration process.
ACKNOWLEDGEMENTS

Cells, drugs and viruses were received from the EU Programme EVA Centralised Facility for AIDS Reagents, NIBSC, UK or the National Institutes of Health (AIDS Research and Reference Reagent Program). This work was supported by the Spanish MINECO projects BFU2012-06958, FIS PI01083 and Gala Contra la SIDA.
REFERENCES


factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase.

J. Biol. Chem. 279:48883-48892.


of polymorphisms in the LEDGF/p75 gene (PSIP1) with susceptibility to HIV-1 infection and disease progression. AIDS 25:1711-1719.


Figure 1. Experimental design and specificity of the ZFN\textsubscript{LEDGF} targeting \textit{PSIP1} gene. A. Schematic representation of LEDGF/p75 protein highlighting the position of the ZFN\textsubscript{LEDGF} targeted region. Cutting site of the ZFN\textsubscript{LEDGF} is located nearby the sequence coding for the integrase binding domain of the LEDGF/p75 (IBD). B. Experimental design used to generate and evaluate LEDGF/p75 knockout cells. C. Flow cytometry plots of wild-type, mock-treated cells and cells transfected with ZFN\textsubscript{LEDGF} plasmids and challenged with VSV-pseudotyped NL4.3 GFP-expressing virus. D. Gene editing by ZFN\textsubscript{LEDGF} induces heteroduplex formation determined by the surveyor mutation assay (Cel-I). After genomic DNA extraction, heteroduplex formation due to generation of insertions or deletions was assessed by the surveyor mutation assay. DNA fragments were resolved in a 10\% TBE-PAGE gel. Lower migrating products (arrows) are a direct measure of ZFN-mediated gene disruption.

Figure 2. Generation and phenotypic characterization of LEDGF/p75 knockout cells. A. Sequence analysis of the insertions and deletions identified in the three cell lines selected after ZFN\textsubscript{LEDGF} treatment. At least 16 different sequences from each selected cell line were sequenced and aligned. Consensus sequence of the modifications identified in each of the two alleles of the selected cell lines is depicted as allele A and allele B of cell line 1-3. B and C. Protein alignments of \textit{in silico} predicted sequences based on the sequencing data obtained from the two alleles of the ZFN\textsubscript{LEDGF} cell lines that introduce a premature stop codon (B) or of cell line 1 - allele B that harbours an in-frame deletion of 3 exons (C). The IBD region is highlighted in a red box. D. Gene expression of \textit{PSIP1} mRNA corresponding to the 5\' and 3\’ regions. Agarose gel where presence or absence of 5\’ and 3\’ fragments of \textit{PSIP1} mRNA was identified. In the case of the cell line 1, the full length and a truncated form of 3\’ mRNA fragment could be identified. E. Quantification of gene expression (mRNA) corresponding to the 3’ region of \textit{PSIP1} in the selected cell lines. Expression of LEDGF mRNA was completely inhibited in LEDGF\textsuperscript{c-} KO cell lines tested compared to untreated or mock-transfected cell. LEDGF\textsuperscript{c-} (cell line 1) showed a 50\% decrease of the LEDGF mRNA compared to control cells. Mean±SD of three independent determinations is shown. F. Assessment of protein levels of LEDGF/p75 determined by Western-blot in the selected LEDGF/p75\textsuperscript{c+} and LEDGF/p75\textsuperscript{c-} KO cell lines compared to control cells. Upper panel; antibody recognizing the C-terminus of LEDGF/p75 protein. Lower panel; antibody recognizing the N-terminus of LEDGF/p75 and LEDGF/p52 proteins. Molecular weight (MW) markers are depicted. Red arrows indicated putative truncated proteins.
Figure 3. Infectivity of HIV in ZNF\textsuperscript{LEDGF} treated cells and susceptibility to antiretroviral compounds. A and B. Relative infection of ZNF\textsuperscript{LEDGF} treated cells compared to wild-type (WT) and mock transfected (Mock transf) controls for NL4-3 (A) and Bal (B) viral strains. Mean±SD of three independent experiments is shown. *** p<0.001; ** p<0.01; * p<0.05.

C. Percentage of HIV X4 using strain NL4-3 replication in LEDGF\textsuperscript{+/-} and LEDGF\textsuperscript{-/-} KO mutants relative to wild-type and mock-treated controls, treated or not with the RT inhibitor AZT, the CXCR4 antagonist AMD3100 and the IN strand transfer inhibitor Ral. Mean±SD of three independent experiments is shown.

D. Percentage of HIV R5 using strain Bal replication in LEDGF\textsuperscript{+/-} and LEDGF\textsuperscript{-/-} KO mutants relative to wild-type and mock-treated controls, treated or not with the RT inhibitor AZT, the CXCR4 antagonist AMD3100 and the IN strand transfer inhibitor Ral. Mean±SD of three independent experiments is shown. WT, wild-type, Mock tr, mock transfected, ND, no drug.

Figure 4. Integration of HIV NL43 is impaired in LEDGF\textsuperscript{-/-} KO cells. A. Viral DNA in LEDGF\textsuperscript{-/-} KO cells (white bars) related to Mock-treated cells (black bars) with or without the RT inhibitor AZT (4 \textmu M) and the IN inhibitor Raltegravir (RAL, 2 \textmu M). Values are expressed relative to mock-transfected cells. Mean±SD of three independent experiments is shown.

B. Integrated viral DNA in LEDGF\textsuperscript{-/-} KO cells (white bars) related to Mock-treated cells (black bars) with or without the RT inhibitor AZT (4 \textmu M) and increasing concentrations of the IN inhibitor Raltegravir. Values are expressed relative to mock-transfected cells. Mean±SD of three independent experiments is shown.
Table 1. Inhibition of HIV replication of all monoclonal cell lines obtained. 

High efficacy (roughly 70% effective impairment of HIV replication) is observed as a consequence of ZNF\textsubscript{LEDGF} treatment.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Mean GFP (%)</th>
<th>Rel replication (%)</th>
<th>Rel inhibition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>2</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>3.9</td>
<td>93</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>12</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>7</td>
<td>1.7</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>0.6</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>26</td>
<td>74</td>
</tr>
<tr>
<td>10</td>
<td>1.8</td>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td>11</td>
<td>5.2</td>
<td>124</td>
<td>-24</td>
</tr>
<tr>
<td>12</td>
<td>5.1</td>
<td>121</td>
<td>-21</td>
</tr>
<tr>
<td>13</td>
<td>0.7</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>14</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>15</td>
<td>2.7</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>16</td>
<td>1.7</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>17</td>
<td>0.8</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>18</td>
<td>5.5</td>
<td>131</td>
<td>-31</td>
</tr>
<tr>
<td>19</td>
<td>0.6</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>20</td>
<td>0.8</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>21</td>
<td>3.8</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>23</td>
<td>0.8</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>24</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>25</td>
<td>4.1</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>27</td>
<td>0.9</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>28</td>
<td>1.1</td>
<td>26</td>
<td>74</td>
</tr>
<tr>
<td>29</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>31</td>
<td>1.1</td>
<td>26</td>
<td>74</td>
</tr>
<tr>
<td>32</td>
<td>1.1</td>
<td>26</td>
<td>74</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>34</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>35</td>
<td>0.6</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>36</td>
<td>1.4</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>37</td>
<td>0.7</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>38</td>
<td>0.6</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>Control</td>
<td>4.2</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

Values represent the mean of two independent determinations performed in duplicate. Rel replication; % replication relative to control cell line, Rel inhibition; % inhibition relative to control cell line, nd; not determined.
Figure 1

a. 

b. ZFN design and activity testing in vitro by Cel I assay → Plasmid transfection into TZM-bl cells → Enrichment of pLVX-ZsGreen ZFN-LEDGF positive cells by cell sorting → Single cell cloning of sorted cells by limiting dilution → Functional assessment of LEDGF negative clone selected: HIV-1 infection → Cell clonal population analysis:
1. Prescreening of refractory clones to VSV-NL4-3-GFP
2. Heteroduplex formation Cel I Assay
3. LEDGF sequence (plasmid clones)
4. LEDGF expression (mRNA, protein)

C.

D. WT, WT+AZT, MOCKtr: 

Clone 1, Clone 2, Clone 3:
Figure 4

(a) Relative Prov DNA

(b) Rel. Integrated HIV

Legend:
- MOCKtr.
- LEDGF⁻/⁻ KO