Review

CellPress

Host factors for retroviral integration site selection

Zeger Debyser, Frauke Christ, Jan De Rijck, and Rik Gijsbers

Laboratory for Molecular Virology and Gene Therapy, KU Leuven, Leuven 3000, Flanders, Belgium

To achieve productive infection, retroviruses such as HIV stably integrate their reverse transcribed RNA genome into a host chromosome. Each retroviral family preferentially integrates near a unique subset of genomic features. HIV integrase (IN) is targeted to the body of active transcription units through interaction with lens epithelium-derived growth factor (LEDGF/p75). We describe the successful effort to develop inhibitors of the interaction between IN and LEDGF/p75, referred to as LEDGINs. Gammaretroviruses display a distinct integration pattern. Recently, BET (bromo- and extraterminal domain) proteins were identified as the LEDGF/p75 counterparts that target the integration of gammaretroviruses. The identification of the chromatin-readers LEDGF/p75 and BET as cellular cofactors that orchestrate lentiviral or gammaretroviral integration opens new avenues to developing safer viral vectors for gene therapy.

Introduction

Infection with HIV-1 remains a substantial global public health problem. Although combination antiretroviral therapy (cART) effectively suppresses HIV replication and profoundly increases survival of patients, a definitive cure is not achievable yet. Interruption of cART typically results in a rebound of virus replication. This is primarily due to the fact that HIV ingeniously escapes from the continuous immune surveillance in reservoirs of latently infected cells that are not susceptible to drug therapy. Moreover, a rapid replication rate, extensive genetic diversity, and suboptimal adherence to treatment sustain the emergence of drug-resistant viral strains resulting in treatment failure. Therefore, there is a continuous demand for novel and better antiretrovirals (ARVs) to control the HIV pandemic with the hope to eventually achieve permanent remission of the disease.

To achieve productive infection, retroviruses such as HIV stably integrate their reverse transcribed RNA genome into a host cell chromosome. Integration of retroviruses is not random, instead each retroviral family favors integration near a unique and specific subset of genomic features. Although the viral IN protein catalyzes the integration reaction, the virus depends on host cell proteins, the so-called cofactors, for guiding IN to the chromatin. HIV IN, as part of the pre-integration complex (PIC), is targeted to the body of active transcription units through interaction with LEDGF/p75, a host protein that binds to lentiviral IN and thereby tethers the HIV PIC to the chromatin [1–3]. In a previous review on HIV IN cofactors from 2006, LEDGF/p75 was singled out as a potential candidate target for antiviral therapy (Box 1) [4]. This prediction turned out to be correct [5].

In a first chapter we describe the further validation of LEDGF/p75 as an antiviral target and the efforts undertaken to develop small-molecule inhibitors of the interaction between IN and LEDGF/p75. The first inhibitors of the LEDGF/p75–IN interaction, the 2-(quinolin-3-yl)acetic acid derivatives were identified using structure-based drug design and potently inhibit HIV-1 replication in cell culture [5]. Multiple analogous inhibitors have been reported since (see [6] for review). Because all bind to the LEDGF/p75-binding pocket in IN, this class is referred to as LEDGINs [5,6]. Interestingly, LEDGINs have a multimodal mode of action: in addition to inhibiting the interaction with LEDGF/p75, they also modulate IN multimerization, thus interfering with integration and HIV assembly [7–13].

Gammaretroviruses display an integration pattern distinct from that of lentiviruses, preferentially integrating near strong enhancers, transcription start-sites, CpG islands, and DNase I hypersensitive regions [14,15]. The capacity to integrate into the host DNA makes retrovirus-derived vectors the most robust and reliable tools to stably insert therapeutic genes in human cells for gene therapy. However, uncontrolled clonal proliferation as a result of vector integration near proto-oncogene transcription start-sites (i.e., insertional mutagenesis) hampers broad clinical applications [16]. Following the identification of LEDGF/p75 as the exclusive tether of lentivirus integration, we and others launched a quest to identify the host proteins that direct gammaretroviral integration. Three groups independently identified BET (bromo- and extraterminal domain) proteins as the LEDGF/p75 equivalent targeting integration of gammaretroviruses, such as the murine leukemia virus (MLV), to transcription start-sites [17–19]. The identification and characterization of LEDGF/p75 and BET, both chromatin-readers, as cellular cofactors that orchestrate lentiviral and gammaretroviral integration, respectively, opens new avenues to develop safer viral vectors for gene therapy. In a second chapter we explain how BET proteins (BRD2, BRD3, BRD4) coordinate murine leukemia virus MLV integration into host cell chromatin, and compare what is known about their molecular mechanisms. In a last section

Corresponding author: Debyser, Z. (zeger.debyser@med.kuleuven.be).
Keywords: retrovirus; integration; integration site; LEDGF/p75; LEDGIN; BET proteins.

0968-0004/
 http://dx.doi.org/10.1016/j.tibs.2014.12.001

Box 1. Novel insights into the IN–LEDGF/p75 interaction

- The initial hypothesis that LEDGF/p75 is the dominant targeting factor of HIV IN has been confirmed [2,39,40]. Moreover, the counterpart for MLV IN, BET proteins, has been identified recently [17–19].
- The main success in the field is the discovery of non-toxic inhibitors of the IN–LEDGF/p75 interaction that block HIV replication [5]. The fact that such LEDGINs also inhibit IN multimerization and HIV assembly was not predicted.
- Hendrix *et al.* revealed how LEDGF/p75 displays a scan-and-lock mechanism when binding to chromatin in the presence of HIV IN [42].
- A role for LEDGF/p75 in the repair of double-stranded DNA breaks has been proposed [80]. Whether LEDGF/p75 plays a role as a survival factor during HIV replication has not been directly answered.
- Initial reports claiming lack of effect on HIV replication or viral vector transduction upon siRNA-mediated knockdown of LEDGF/p75 were based on false negative results [81,82]. Because LEDGF/p75 is a highly abundant protein, potent knockdown (>90%) is required to observe a phenotype [3,27].

we discuss the implications of the discovery of the LEDGF/p75 and BET proteins as retroviral targeting factors for the design of safer MLV-derived vectors for gene-therapeutic purposes [20,21].

Identification and validation of LEDGF/p75 as a host factor for lentiviral integration

LEDGF/p75, a transcriptional coactivator [22–24], was initially identified as an IN cofactor by coimmunoprecipitation from cells overexpressing HIV IN [1]. LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF) family composed of chromatin-associated proteins sharing particular structural features. The pivotal role of LEDGF/p75 in HIV replication was subsequently evidenced via mutagenesis, RNAi-mediated depletion, transdominant overexpression of the IN-binding domain (IBD) of LEDGF/p75, and cellular knockout studies [2,3,25–32].

LEDGF is encoded by the *PSIP1* (PC4- and SFRS-interacting protein 1) gene on human chromosome 9 and is expressed as two splice-variants, the LEDGF/p52 and LEDGF/p75 proteins. Both share a N-terminal region (aa 1–325) comprising the nuclear localization and chromatin-binding elements defined by the PWWP (Pro-Trp-Trp-Pro) domain, the NLS (nuclear localization signal), the A/T hook-like elements, and three charged regions (CR1, CR2, and CR3) [33,34] (Figure 1). The C-terminal region which is extended in the p75 splice variant (Figure 1) contains the IBD (aa 347–429). Although this domain was originally characterized by its association with HIV IN [35], later studies have shown that different cellular binding partners interact with LEDGF/p75 via this protein-binding domain [36–38]. Instead, LEDGF/p52 contains eight unique aa in its C-terminus and does not interact with HIV IN [25]. Colocalization studies mapped the LEDGF/p75 interaction site in IN to the catalytic core domain and to a lesser extent to the N-terminal domain [25]. Through its classical NLS, LEDGF/p75 predominantly locates to the nucleus where it tightly associates with chromatin via its PWWP domain. During HIV replication it thereby tethers IN associated with the viral genome to the host chromatin, facilitating the integration into

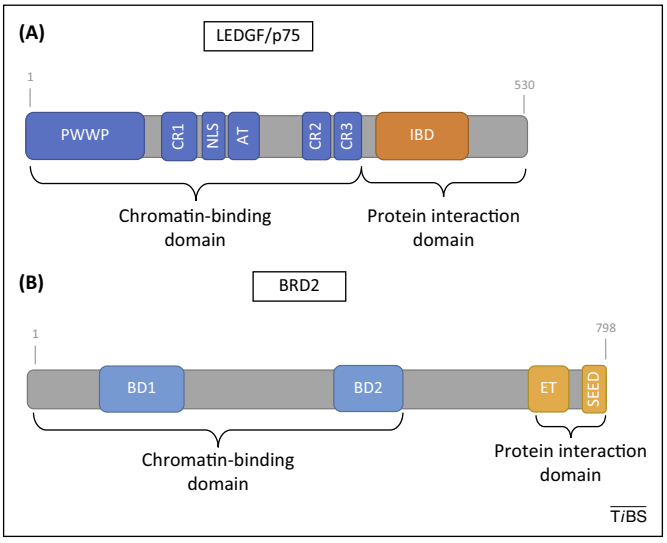

Figure 1. Structure of LEDGF/p75 and BET proteins. **(A)** Schematic representation of LEDGF/p75. LEDGF/p75 reads and binds methylated histones on the chromatin through its N-terminal region (dark blue), while HIV IN interacts with the C-terminal integrase-binding domain (IBD, orange). This interaction orchestrates the lentiviral integration profile. **(B)** Schematic representation of BRD2, one of the BET proteins. BET proteins read and bind to acetylated chromatin on promoters and enhancers in the chromatin through the bromodomains (BD) in their N-terminal regions (light blue), while the ET domain in the C-terminal part (yellow) interacts with MLV IN. Numbers represent amino acid positions. Abbreviations: AT, AT hook; BD, bromodomain; BET, bromodomain and extraterminal domain protein; CR1–3, charged region 1–3; ET, extraterminal domain; IN, integrase; LEDGF/p75, lens epithelium-derived growth factor; NLS, nuclear localization signal; PWWP, Pro-Trp-Trp-Pro domain; SEED, Ser-Glu-Glu-Asp domain.

HIV-preferred sites [2,25,31,39,40]. In addition to its tethering function, LEDGF/p75 protects IN from proteolytic degradation and stimulates the catalytic activity of IN *in vitro* as well as in *in vivo* [1,34,41]. Integration site analysis in human cells depleted of LEDGF/p75 by RNAi, in embryonic fibroblasts derived from LEDGF knockout mice, or in human preB cell lines with a specific LEDGF/p75 knockout corroborated a role for LEDGF/p75 as a tethering and targeting factor for HIV integration [2,31,32,39,40]. In LEDGF/p75-depleted cells the related HRP-2, that also contains a PWWP and IBD domain, partially rescues the replication deficit [32]. Biophysical studies evidenced a scanning mechanism of LEDGF/p75 on the chromatin whereby the LEDGF/p75 complex is locked on the chromatin in the presence of IN [42].

In search of small molecules inhibiting IN–LEDGF/p75 interaction

In 2005 the solution structure of IBD in complex with a dimer of the IN catalytic core domain was reported (PDB code: 2B4J) [43], identifying the amino acid residues of LEDGF/p75 mediating the interaction with IN (Ile365, Asp366, Phe406, Val408). The IBD structure is composed of a right-handed compact bundle of five α-helices. The amino acid residues contacting HIV IN are located in the interhelical loop regions of the structure. Two regions of the IN catalytic core domain are in direct contact with LEDGF/p75, namely the region around Trp131 and Trp132 and the region extending from Ile161 to Glu170 [26,30,43]. The interface is located in a pocket formed by the two subunits of the IN-core dimer (α1 and α3 of one