REVIEW

The benefits of integration

A. Ciuffi

Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Abstract

Retroviruses, including the human immunodeficiency virus (HIV), are notorious for two essential steps of their viral replication: reverse transcription and integration. This latter property is considered to be essential for productive replication and ensures the stable long-term insertion of the viral genome sequence in the host chromatin, thereby leading to the life-long association of the virus with the infected cell. Using HIV as a prototypic example, the present review aims to provide an overview of how and where integration occurs, as well as presenting general consequences for both the virus and the infected host.

Keywords: Endogenous retrovirus, human immunodeficiency virus, integrase, integration, lens-epithelium derived growth factor/p75, nuclear import, nuclear pore proteins, persistence, retrovirus

Corresponding author: A. Ciuffi, Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Bugnon 48, 1011 Lausanne, Switzerland
E-mail: Angela.Ciuffi@chuv.ch

Introduction

Retroviruses are enveloped RNA viruses, containing two copies of single-stranded, non-segmented, positive RNA as genome. Like all retroviruses, the human immunodeficiency virus type 1 (HIV-1, abbreviated as HIV throughout the text) encodes three major open reading frames: (i) *gag*, coding for the internal structural proteins, which in the case of HIV are matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7) and p6; (ii) *pol*, coding for the virus enzymes, which are reverse transcriptase (RT, p66/p51), integrase (IN, p32) and protease (PR, p11); and (ii) *env*, coding for the envelope external structural proteins, which are the surface glycoprotein (SU, gp120) and the transmembrane glycoprotein (TM, gp41) for HIV (Fig. 1a). Moreover, HIV-1 encodes regulatory and accessory genes—*tat*, *rev*, *vif*, *vpr*, *vpu* and *nef*—and is therefore considered to be a complex retrovirus.

In order to establish a productive replication, HIV needs first to deliver its genome-containing viral core to the cytoplasm of the infected cell (Fig. 1b). Subsequently, the viral RNA genome is reverse transcribed in a double-stranded DNA linear copy, hence the name, retrovirus. The viral DNA genome is complexed with the viral integrase enzyme (also referred to as the intasome), as well as with additional viral and cellular proteins in a ribonucleoprotein complex called the pre-integration complex (PIC). The exact PIC composition is still controversial but additional proteins may include the viral proteins RT, MA, Vpr and CA, as well as the cellular proteins lens-epithelium derived growth factor (LEDGF/p75), barrier-to-autointegration factor (BAF) and high-mobility group AT-hook 1 (HMGA1) (reviewed in ref. [1]). Components of PIC are necessary for the successful nuclear import and final stable insertion of the viral genome in the host DNA. As discussed below, these two steps have been shown to impact integration efficiency and site location, and are essential to ensure the life-long persistence of the provirus in the infected cell.

How? Mechanism of Integration

The viral integrase enzyme is the only protein determinant required to successfully join and so insert a DNA fragment into a heterologous target DNA sequence (reviewed in refs [2,3]).

The prototypical HIV-1 IN is a 288-amino-acid protein, and is divided into three major domains: an N-terminal domain, a catalytic core domain and a C-terminal domain (Fig. 2a)

http://dx.doi.org/10.1016/j.cmi.2016.02.013

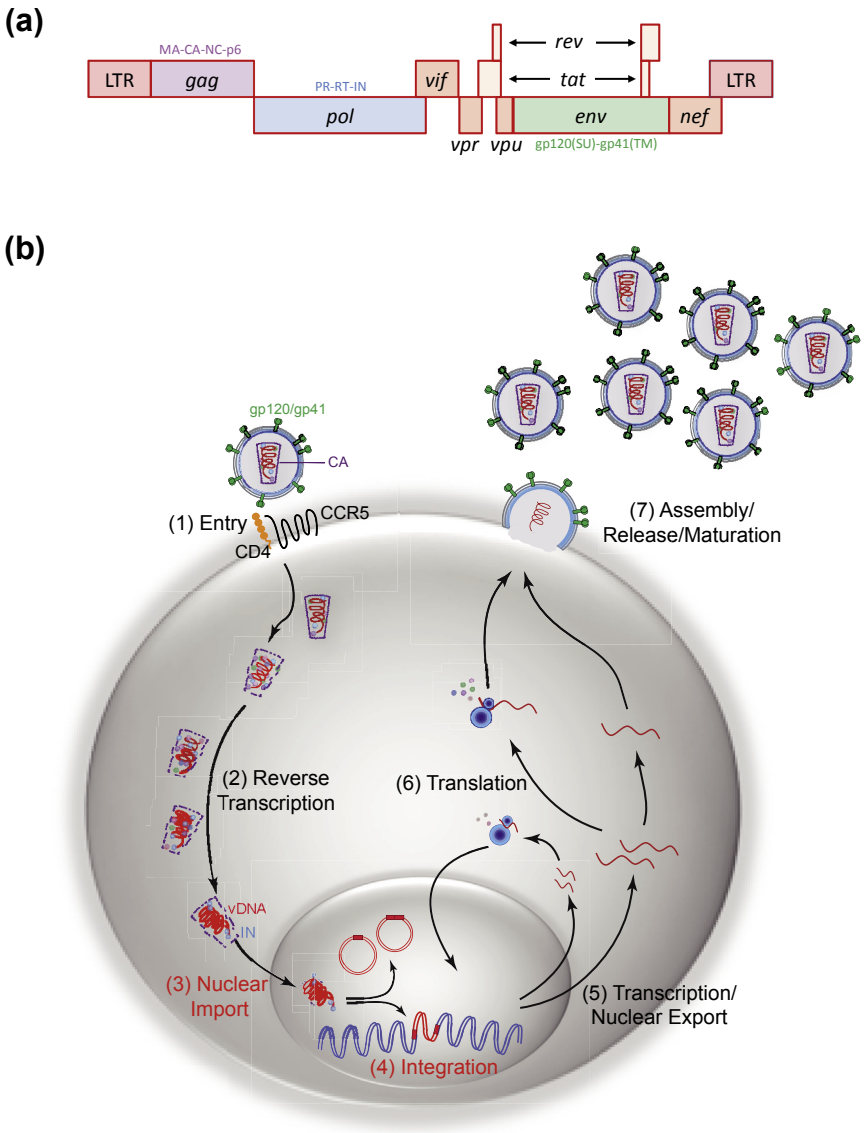

FIG. 1. Overview of the human immunodeficiency virus type 1 (HIV-1) genome organization and replication cycle. (a) HIV genome contains nine open reading frames (ORF), coding for 15 proteins. The proviral genome is flanked by direct long-terminal repeats (LTR) that contain all transcriptional regulatory sequences. The *pol* ORF encodes the three viral enzymes necessary for replication that are protease (PR), reverse transcriptase (RT) and integrase (IN). (b) The HIV replication cycle can be divided into seven steps: (1) HIV binds to its target cell through the interaction of gp120 to the cell CD4 molecule, which is mostly expressed at the surface of lymphoid and myeloid cells. This first interaction allows the subsequent binding of gp120 with a chemokine receptor, CCR5 or CXCR4, followed by the fusion between the viral membrane and the cellular membrane triggered by gp41. This ensures the release of the viral core in the cytoplasm of the host cell. (2) The viral core disassembles (uncoating process) and the viral RNA genome is reverse transcribed in a linear double-stranded DNA copy through the action of the viral reverse transcriptase enzyme, giving rise to the pre-integration complex (PIC). (3) The PIC, minimally containing the viral DNA (vDNA) genome and the viral IN enzyme, is translocated to the nucleus through the nuclear pore. This nuclear import step requires multiple interactions between viral and cellular proteins, including capsid (CA) binding to nuclear pore proteins (NUPs). (4) Once in the nucleus, the viral IN catalyses the stable insertion of the viral DNA genome into the host chromatin, tethered mainly by the cellular lens-epithelium derived growth factor (LEDGF)/p75 protein. LTR circles, 1-LTR and 2-LTR circles, are considered as dead-end by-products produced by the cellular DNA repair machinery, the homology repair (HR) or non-homologous end-joining (NHEJ) pathway, respectively. (5) Once integrated, the provirus is transcribed by the cellular RNA polymerase II machinery as most cellular coding genes. Viral transcripts (with different levels of splicing) are exported from the nucleus to the cytoplasm where they are translated (6). (7) Two copies of full-length (unspliced) viral RNA and viral proteins assemble, thereby producing new particles that are released from the cellular membrane. Finally, the viral protease cleaves viral polyproteins leading to mature and infectious viral particles.