Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Immunology

# Integration of retroviral vectors

Richard Gabriel, Manfred Schmidt and Christof von Kalle

Retroviral vectors are unique in their ability to integrate their genome into the host genome of transduced cells. Several members of the retrovirus family show distinct pattern for preferential integration into the host genome. Despite many years of investigation, precise mechanisms of target site selection and the fundamental interplay of viral integrase and host cell proteins are still unknown. Improved methods to detect retroviral integrations genome-wide as well as recent advances on the retroviral integrase structure and integrase interacting proteins may lead to further uncover the process of retroviral target site selection. A better knowledge of these mechanisms and interactions will allow further improving safety of retroviral vectors for gene therapy by providing an opportunity to retarget retroviral integration into non-harmful genomic positions.

**Address**
Department of Translational Oncology, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany

Corresponding author: von Kalle, Christof (christof.kalle@NCT-Heidelberg.de)

**Current Opinion in Immunology** 2012, **24**:592–597

This review comes from a themed issue on **Immunogenetics and transplantation**

Edited by **Alain Fischer** and **Matthew Porteus**

For a complete overview see the Issue and the Editorial

Available online 14th September 2012

0952-7915/$ – see front matter, 

http://dx.doi.org/10.1016/j.coi.2012.08.006

## Introduction

The family of retroviruses consists of two subfamilies: *Orthoretrovirinae* with the genera of simple alpha-retroviruses, beta-retroviruses, gammaretroviruses and complex delta-viruses, epsilon-viruses and lentiviruses as well as *Spumaretrovirinae* with the genus of the foamy viruses. Integration of the viral genome is a hallmark for the retroviral life cycle. Because of the integration into the host genome the virus becomes an inheritable part of the host cell. This feature renders retrovirus-derived vectors attractive tools as gene transfer vehicles, as they stably introduce functional genes or new genetic information into the host cell which are then expressed from the integrated provirus.

The various genera exhibit different features regarding cell tropism and integration preference. One major difference is the ability to transduce resting or cycling cells. Gammaretroviral vectors are dependent on degeneration of the nuclear membrane during cell division to allow the virus to enter the nucleus. In contrast, lentiviral vectors actively enter the nucleus *via* the nuclear pore enabling them to transduce non-dividing cells efficiently. Despite many years of investigation, the exact interplay between viral integrase (IN) and host proteins and how cellular cofactors modulate target site selection of the different genera remain largely unknown.

Understanding the mechanisms underlying retroviral integration as well as identifying positions in the host genome where retroviral vectors integrate are crucial to further improve retroviral vectors for safe gene therapeutic applications. This review focuses on recent advances in understanding the integration mechanism of retroviral vectors, new strategies to indentify provirus locations in the host genome and the consequences of integrated provirus on the host cell.

## Detection of retroviral integration loci

Integration of retroviruses and derived vectors is mediated by the viral IN encoded by the pol gene. Retroviral integration has been extensively studied in the past (reviewed in [1]). IN assembles with viral DNA and cellular proteins to form the preintegration complex (PIC). IN possesses two catalytic activities: 3′ end processing, cleaving a dinucleotide from the 3′ end of each long terminal repeat (LTR) leaving an invariant CA dinucleotide and DNA strand transfer which uses the recessed 3′ termini to open the target DNA to join the linear viral genome to the host genome. Integration is completed by host DNA repair enzymes, resulting in species-specific 4–6 bp direct repeats flanking the provirus (reviewed in [2]).

Several PCR-based methods have been developed to efficiently amplify and identify retroviral integration sites (IS) from the host genome, and most prominent among those are LAM-PCR [3] and LM-PCR (including several modifications) [4,5]. All conventional IS analysis methods are dependent on restriction enzymes to digest the genomic DNA before ligation of a known linker sequence that allows subsequent amplification and sequencing of the vector genome junctions. Mapping of the genomic amplicon sequences enables us to precisely locate retroviral IS in the host genome. However, applying optimal restriction enzyme combinations is crucial to circumvent restriction and amplification biases [6].

Recently developed restriction enzyme independent methods uncover retroviral IS either by the ligation of single-strand oligonucleotides to amplified vector genome junctions [6,7•], by introduction of adaptor sequences on the basis of phage Mu transposition [8•] or by sonication of the genome [9•].

The recent implementation of high-throughput sequencing exponentially increased the number of available retroviral IS. However, as now millions of single sequences can be produced in a few days, sophisticated data analysis tools are indispensable. Recently, we and others published bioinformatical tools both for automatic processing of raw sequences [10,11] and downstream data analysis [12,13].

## Integration site selection by retroviral vectors

For many years structural insights into the complex of IN with the host DNA were limited. Recently, the structure of the prototype foamy virus (PFV) IN was resolved with high resolution by X-ray crystallography [14•]. The intasome — the complex of viral DNA with IN — appears to be a dimer-of-dimers of IN, where only one subunit of the dimer binds to the viral DNA end. During strand transfer reaction, severe bending of (chromosomal) target DNA allows the active sites of the intasome to access and cleave the target phosphodiester bonds [15••]. Consequently, PFV favors integration into genomic sequences harboring central, flexible pyrimidine-purine dinucleotides. As expected, base-specific interactions of IN with target DNA are few. The structure of PFV-IN allows modeling the structure of other retroviral IN, like HIV [16]. This modeling can serve as a starting point for the rational design of sequence specific retroviral IN [17]. Progress in structural biology of retroviral IN beyond this brief overview was reviewed in detail recently [18].

Retroviral IN must confer its action on two types of DNA, the viral LTR ends and the chromosomal target DNA. Suicidal autointegration is prevented by strong uracilation of the retroviral genome (>5%) [19]. In contrast to the sequence-specific recognition of viral DNA by IN [20], selectivity on the target DNA is more promiscuous, as strong sequence specificity would be disadvantageous by limiting the frequency of potential integration loci. However, weak palindromic consensus sequences are commonly found at retroviral IS [21,22]. Retroviruses show genus-specific preferences for integration. Since the first draft of the human genome sequence, numerous IS studies have revealed distinct patterns for different retroviruses. Murine leukemia viruses (MLV) and derived vectors from the gammaretrovirus genus show a strong preference for integration close to transcription start sites (TSS) and CpG islands [23–25]. In contrast, lentiviral vectors prefer integration inside transcription units of actively transcribed genes [26] that cluster in megabase-wide chromosomal regions [27••].

Alpha-retroviral vectors show a largely random and uniform integration pattern [25,28]. Similar, rather close to random IS distribution has been shown for PFV [29,30]. However, IS analysis uncovers only endpoint scenarios. Efficient IS analysis in freshly transduced cells is hampered by the presence of nonintegrated episomal vector forms. Thus, different IS patterns reflect integrations persisting until time of analysis, which is not necessarily congruent with the initial IS preference. For example, potential deleterious (and apoptotic) integration events can not be detected. Furthermore, a direct comparison of IS obtained from pretransplantation samples and later time points from patient samples in hematopoietic stem cell gene therapy is limited, as only a minor fraction of the initially transduced heterogenic CD34+ enriched cells will engraft and show self-renewal capacity.

The causal mechanisms underlying preferential integration into particular regions of the genome are still not dissolved. Substituting IN of a HIV-1 derived lentiviral vector by MLV IN changed the characteristic lentiviral integration pattern into a gammaretroviral-like integration pattern, indicating that the viral IN plays a major role in target site selection [24]. However, on the genome level host factors are thought to confer target site selection and interactions of the viral IN with cellular proteins (see below) as well as influence of transcriptional activity [23,31] and chromatin status [32] are likely further decisive. As such, cell type specific integration pattern have been reported for both gammaretro viral and lentiviral vector systems [33•,34,35]. For gammaretroviral vectors a preferential integration into genomic regions enriched with cell-type specific subsets of transcription factor binding sites has been observed. IN and LTR enhancer appeared to determine the tethering of retroviral preintegration complexes to transcriptionally active regulatory regions [31]. With regards to HIV-1 based vectors, lens epithelium-derived growth factor (Psip1/Ledgf/p75) is the best studied IN interacting protein tethering vector integration into actively transcribed genes. Downregulation of Psip1/Ledgf/p75 results in decreased preferential integration into transcription units [36,37]. Similarly, we recently showed close to random integration into genes in rodent postmitotic tissue, that express Psip1/Ledgf/p75 to much lower levels compared to dividing cells [33•]. Several other studies recently showed that lentiviral integration can be retargeted into heterochromatin by replacing the Psip1/Ledgf/p75 chromatin interaction domain [38–40]. Another study revealed that knockdown of nuclear pore proteins Transportin-3 and RanBP2 decreased targeting of HIV into gene dense regions, suggesting that the nuclear pore may be involved in trafficking HIV to preferred IS [41].

## Influence of proviral DNA on the host genome

Integration of a foreign DNA is *per se* a mutagenic event and can lead to malignant transformation of cells