Innate immunity against retroviral pathogens: from an ambiguous genetic self to novel therapeutic approaches

Sylvain Meylan, Didier Trono
School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Summary

Almost half of the human genome is derived from exogenous genetic invaders, most of them related to retroviruses. This is the consequence of longstanding interactions between retroelements and higher organisms, governed by a delicate equilibrium between virus-based evolutionary forces and control by host defense mechanisms. Insight into these longstanding genetic conflicts is suggesting leads for novel therapeutic strategies to fight HIV infection.

In 2001, the Human Genome Organisation (HUGO) and the Human Genome Project (HGP) consortia provided first drafts of the sequence of the human genome [1, 2]. Completed in 2004, this effort opened new avenues for understanding the genetic bases of both physiological and pathological process, hence to move towards a more global comprehension of health and disease. As insights into the content of the human genome were gained, it quickly became apparent that only a fraction, barely two percent, encodes for proteins, which had been traditionally considered as the main executive arms of the cell. More strikingly even, close to 50 percent of human DNA was found to derive from genetic invaders called transposons, most of them retroelements functionally related to the human immunodeficiency virus (HIV). Found in the genomes of all species, transposons are motors of evolution, yet their uncontrolled spread can be fatal to their host organisms. Their presence in higher species thus reflects a delicate equilibrium between occasional permissiveness and tight restriction, through innate immunity mechanisms also engaged in protection against their exogenous viral counterparts.

Key words: genome; retroelement; HIV; innate immunity; evolution

Internal mobility

Mobile genetic elements were discovered some sixty years ago by Barbara McClintock based on her observation of changing colour patterns in maize. Her results and their implications were long held in such scepticism that she even came to renounce publishing in the scientific literature. After progress in the understanding of gene regulation revealed the seminal importance of her discoveries, she was ultimately awarded the Nobel Prize in Physiology and Medicine, some thirty years later. And it was another two decades before the sequencing of the human genome revealed that close to half of its DNA comes from such mobile elements. These can be classified in two general categories [3, 4]. A small minority (about 5%) are DNA transposons that replicate by a “cut-and-paste” mechanism, encode for a recombinase that mediates their “hopping” along the cell genome. These elements do not get amplified during this process, and no currently active DNA transposon has been detected in the mouse or human genomes. This somewhat belittles their perceived impact on recent evolution, hence their relevance to the present discussion. By contrast, the vast majority (almost 95%) of mobile elements harboured by the genomes of higher species are retrotransposons, which replicate through a “copy-and-paste” mechanism leading to an amplification that explains their prevalence. The biology of these retrotransposable elements and the intricacy of their interactions with higher species is the topic of this review.
Biology of retrotransposons

The DNA of retrotransposons is transcribed by the cellular machinery into RNA molecules that are copied by the group-defining reverse transcriptase back into DNA, which can become integrated again into the genome of the cell. Because the RNA intermediates can be generated by multiple rounds of transcription, this process results in a marked amplification of retroelements once they invade the genome of a species, unless and until control mechanisms are in place. Furthermore, the reverse transcriptase enzyme being error-prone, retrotransposons and other retroelements can evolve rapidly when confronted with selective pressures.

Retrotransposons are further classified into Long Tandem Repeat (LTR)-containing retroelements (8.5% of the human genome, representing over a quarter million copies) and LTR-less or non-LTR retroelements (35% of the human genome, 2.4 million copies) (fig. 1). LTR-retrotransposons are endogenous retroviruses (ERV), that is, retroviruses that once infected the germ line of their host species or of one of its ancestors, and thus became a stable component of the genome of this lineage. The timing of such invasion can often be traced through phylogenetic studies (table 1) [5, 6]. Rare are the endogenous retroviruses (none in humans) that conserve the ability to perform a full extra-cellular cycle, but their genomes carry the hallmarks of retroviruses, with sequences (or remnants thereof) encoding the structural Gag components, a reverse transcriptase, an integrase and for some an envelope [7]. Many of these elements have become inactive over time through the accumulation of mutations, but they are also kept at bay by host defence mechanisms. Non-LTR retrotransposons lack several prototypic features of exogenous retroviruses, amongst which an envelope coding sequence. This group comprises autonomous and non-autonomous members, the prototypes of which are L1 elements and Alu repeats, respectively, found each at close to one million copies in the human genome.

Origin of retrotransposons

Phylogenetic analyses of the reverse transcriptase sequences of endogenous and exogenous retroviruses demonstrate a clear evolutionary relationship, pointing to a likely common ancestor hundreds of millions years ago [8, 9]. Similar studies also indicate a common origin for all classes of retrotransposons (LTR-retrotransposons, LINE — including L1- and SINE — including Alu-elements), and it is tempting by extension to relate the origin of retrotransposons to bacterial and mitochondrial genetic elements that encode a reverse transcriptase, or to the eukaryotic telomerase. However, the latter connections, albeit plausible, remain speculative. According to the sequence divergence between members of one class and their functional consensus sequence, and taking as a reference the average mutation rate of eukaryotic genomes (e.g. divergence between humans and old-world monkeys reflects 25 mio years of evolution), one can estimate the periods that have witnessed retrotransposition activity [1]. Interestingly, over the last 100 million years,

Table 1
Superfamilies of LTR-containing retroelements. The LTR-retrotransposons are classed in six major superfamilies and based on sequence analysis are related to the retroviral subfamilies, suggesting different infectious events. Reproduced from Bannett et al. [6].

<table>
<thead>
<tr>
<th>Element</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I ERV</td>
<td>Similar to type C or retroviruses</td>
</tr>
<tr>
<td>Class II ERV</td>
<td>Similar to type B or retroviruses</td>
</tr>
<tr>
<td>Class III ERV</td>
<td>Distantly related to spuma retroviruses</td>
</tr>
<tr>
<td>MER4</td>
<td>Non-autonomous class I-related ERVs</td>
</tr>
<tr>
<td>MST</td>
<td>Named for common MstII restriction site</td>
</tr>
<tr>
<td>MLT</td>
<td>Mammalian LTR retro-transposons</td>
</tr>
</tbody>
</table>
LTR-retrotransposons were very active up to 50 million years ago (mya), but since then have progressively lost most activity in primates, with the last new insertion dated back to 0.2 mya in the precursor of modern humans [1, 6]. Although also active 100 mya, non-LTR retrotransposons peaked later (approx 50 to 25 mya) in primates, but then declined to become almost completely controlled in modern humans (only 80 active L1 elements out of the million or so present in the human genome, with only one human birth in a 100 carrying a new insertion) [6, 10]. Interestingly, these most recent bursts of retrotransposon activity were synchronous with the development of mammals (approximately 100 mya) and the speciation of hominids (50 mya). Of note, retrotransposons have remained far more active in rodents, causing close to 10% of spontaneous mutations in inbred strains of mice.

### Impact of retrotransposons

Retrotransposons have shaped the genomes of higher organisms by exerting direct influences on their architecture, by modulating their expression, and by triggering the emergence of defence mechanisms.

1. **Retrotransposons as architects of the human genome:** Through their “copy-and-paste” replication mechanism, retroelements have filled genomes with hundreds of thousands to millions of near-identical DNA sequences. This constitutes a most favourable ground for rearrangements of several kinds: homologous recombination, duplications, deletions and translocations (fig. 2). As such, retrotransposons are formidable evolutionary forces. Comparing the human and chimpanzee genomes at orthologous loci reveals species-specific losses of up to several megabases, and insertions of several tens of thousands of base pairs [11], involving both coding sequences and regulatory elements. Duplications have promoted the expansion of gene classes. For instance, the some 600 human genes coding for zinc-finger transcription factors are often found in clusters enriched in SINE retrotransposons, and most likely derived from a common ancestor. Conversely, retrotransposon-induced deletions account for almost 50 human diseases [4], including familial hyperlipidaemia, due to the deletions in the LDL receptor gene, and acute mylogenous leukaemia, secondary to a deletion in the MLL gene [12].

2. **Retrotransposons as contributors to the coding potential of the host:** retrotransposons tend to integrate in gene-rich regions, and sometimes become part of the exons of existing genes or even contribute new genes. One of the most striking examples of the latter situation is that of synctin in mice and humans. This protein is essential for the formation of the placental syncytiotrophoblast, and perhaps crucial for maternofoetal immune tolerance [13]. Remarkably, its various forms are all encoded by the env gene of endogenous retroviruses that independently invaded the germ cells of mice and humans, and likely substituted for a resident gene until then responsible for this mammal-defining function. Also, retroelement-derived sequences are over-represented in mRNAs of rapidly changing genes such as those involved in immune or stress responses, suggesting that retrotransposons have favoured the rapid evolution of these genes [14]. As well, the inclusion of sequences adjacent to those of retroelements during their copy-and-paste replication (a process called transduction) can cause the fusion of gene fragments resulting in novel proteins, as for example the HIV1 restriction factor TRIMcyp of owl monkeys [15]. In fact, it is suggested that up of 0.1% of human protein coding regions contain transposable elements [16].

3. **Retrotransposons as modulators of transcriptional activity:** the spread of retroelements litters the genome with cis-acting regulatory sequences, comprising promoters, enhancers, splice-acceptor/donor and poly-adenylation signals as well as other transcriptional or posttranscriptional regulatory elements. Recent work suggests that LINE1 may have a role in chromosome X-inactivation (reviewed in [17]). However, if
some of these irruptions of regulatory elements can be beneficial (SINE retrotransposons can act as physiological enhancers as was demonstrated for the pro-opiomelanocortin gene [18]), one can easily imagine the deleterious effects of random insertions of transcriptional modifiers in a well-oiled transcriptional landscape (table 2). Progress in the understanding of epigenetics (changes in phenotype/gene expression that occur due to changes other than in the genomic sequence) reveals that retroelements can be associated with modifications of chromatin architecture [3] that influence neighbouring genes. In the murine genome, some members of the B1 SINE family can recruit transcriptional repressors [19], and recent work suggests that SINE partake in homeostatic responses such as heat shock by acting as transcriptional repressors upon cellular stress [20]. Moreover, retroelements can shuttle genes out of an unfavourable environment; X-autosomal retrotransposition is not uncommon and is proposed as a mechanism to dodge the silencing of house-keeping genes by X chromosome inactivation in somatic cells [21].

### Table 2

<table>
<thead>
<tr>
<th>Element</th>
<th>Gene</th>
<th>Transcriptional modification</th>
<th>Functional role</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV-E</td>
<td>Endothelin-B receptor</td>
<td>Promoter regulation</td>
<td>Placenta</td>
</tr>
<tr>
<td>HERV-L</td>
<td>β-Galactosyltransferase</td>
<td>Promoter regulation</td>
<td>Colon, mammary gland</td>
</tr>
<tr>
<td>LTR + LINE2</td>
<td>Chaperonin</td>
<td>Promoter dysregulation</td>
<td>McKusick-Kaufman syndrome</td>
</tr>
<tr>
<td>LINE1</td>
<td>Factor VII</td>
<td>Premature Poly-A truncation</td>
<td>Haemophilia</td>
</tr>
<tr>
<td>Alu</td>
<td>Pro-opiomelanocortin</td>
<td>Enhancer positive regulation</td>
<td>Endocrine regulation</td>
</tr>
<tr>
<td>Alu</td>
<td>NFI</td>
<td>Alternative splicing signal</td>
<td>Neurofibromatosis</td>
</tr>
</tbody>
</table>

### Controlling retroelements

As useful as retroelements can be, their uncontrolled spread would have rapidly lethal consequences. Correspondingly, all species have evolved sophisticated mechanisms to control this process. In primates, the emergence of several of these retroelements’ restriction factors coincided with the drop in retroelement activity some 50 mya.

The exogenous retroviral replication cycle provides a convenient background to describe the various mechanisms at play in the control of retroelements (fig. 3). The TRIM5α restriction factor interferes with the un-coating of the viral genome after entry, thus preventing reverse transcription and integration [22–24]. TRIM5α is an important barrier to the cross-species transmission of retroviruses. Human TRIM5α protects from infection with Murine Leukaemia Virus (MLV) and Equine Immune Anaemia Virus (EIAV), whereas the African green monkey aigmTRIM5α restricts not only these two pathogens but also HIV1, HIV2 and SIVmac. As a corollary, retroviruses have evolved to evade the TRIM5α orthologue encoded by their cognate host: HIV1 and HIV2, for instance, are not blocked by human TRIM5α, although the mechanism of their escape is incompletely understood. In two primate species (rhesus macaque and squirrel monkey), a retroelement-driven transduction event led to a fusion between TRIM5α and cyclophilin A (CypA), resulting in a fusion protein that can capture the capsid protein of HIV1, which has affinity for CypA. Thus, a retroelement-restricting factor arose through the action of another retroelement.

Polynucleotide cytidine deaminases of the APOBEC family are another group of restriction factors that target the early steps of retroviral replication, in their case reverse transcription [25, 26]. Various members of the family act on distinct subsets of retroelements, including exogenous and endogenous retroviruses, non-LTR retrotransposons and hepatitis B virus [27–29]. Again, viruses have evolved mechanisms to surmount cytidine deaminases-mediated restriction in their cognate host. The Vif protein of primate
lentiviruses, for instance, is the “antidote” against the APOBEC3F and 3G proteins present in these species.

Hosts have also developed mechanisms to limit the impact of retroelements once these are integrated in the genome. As a general rule, the DNA of LTR and non-LTR retrotransposons is methylated during the early embryonic period, which results in its irreversible silencing. Similarly, murine leukaemia virus is silenced in embryonic cells through the recruitment of a transcriptional repressor complex that acts by remodelling its chromatin environment [30]. This complex comprises a member of a large family of DNA-binding repressors known as KRAB-ZFPs. As KRAB-ZFPs differing in their DNA-binding specificity are encoded in the hundreds by the genomes of higher species, it is tempting to speculate that, by analogy, other members of this family are involved in controlling endogenous retroelements related to MLV. The Zap restriction factor is another example of antiviral effectors, which acts at the posttranscriptional level by inducing the rapid degradation of the MLV RNA [31] in rats. Finally, Tetherin is a cellular protein that can block the release of retroviral particles at the cell membrane [32].

Genetic conflicts

Reverse transcriptase being devoid of proof-reading capacity, the replication of retroelements is error-prone, which promotes its escape from the control of restriction factors. In return, genes encoding restriction factors show the marks of intense positive selection, with accumulation of non-synonymous mutations at positions encoding for charged amino acids, predicted to reside at protein-protein or protein-nucleic acid interfaces. The TRIM5α and APOBEC3G genes, for instance, have accumulated such mutations at far greater rates than housekeeping genes during the last 30 millions years of primate evolution [33, 34]. TRIM5α predictably exhibits species-specific differences over its retroviral capsid-binding domain, reflecting its selection towards the control of distinct sets of retroviruses. For the cytidine deaminases APOBEC3G, differences between orthologues are noted throughout its sequence, perhaps consistent with a much broader spectrum of restriction (MLV, HIV, HBV, LINE1, Alu).

Of note, phylogenetic analyses indicate that all known anti-retroviral restriction factors emerged long before the emergence of lentiviruses, which appeared only 1 mya [35]. Factors now at play against HIV and other members of the lentiviral family thus arose as barriers against other genetic invaders, a process in which endogenous retroelements most likely played a crucial role. Whether the current HIV pandemic is exerting an additional pressure on restriction factors is hard to say. Studies on cohorts of HIV-infected and -uninfected individuals have failed to identify alleles of APOBEC3G or TRIM5α predictive of a more lenient clinical course [36]. However, from an evolutionary perspective, the few decades since the start of the HIV epidemics are far too short a time for a noticeable impact. Furthermore, larger scale studies dissecting patient populations according to HIV strains and genetic background may be necessary to see significant patterns emerge.

Leads for novel therapeutic strategies

Our growing understanding of the molecular events that have governed hundreds of millions of years of interplay between retroelements and their hosts is suggesting novel therapeutic approaches to counter current infections, notably by HIV. Efforts to identify chemical substances capable of disrupting the APOBEC3G-Vif interaction, hence of exposing HIV to the antiviral action of the cytidine deaminase, have made significant progress [37]. As well, a fusion protein between human TRIM5α and CypA was recently demonstrated as an effective blocker of HIV infection in a humanised mouse model, suggesting that it could serve to treat HIV1 infection by gene therapy [38].

More promising even, since unlikely to be evaded by the virus extraordinary mutational ability, are strategies aimed at the host cell mechanisms leading to viral latency, the major obstacle to purging HIV from the infected host with currently available therapies. While latency probably reflects a partial control of HIV by endogenous defence mechanisms similar to those engaged against endogenous retroelements, it precludes the long-term success of HAART (highly active antiretroviral therapy) by providing a permanent source of virus that can reactivate infection in the face of fading pharmacological or immune control. Understanding what mechanisms lead to HIV latency and coming up with methods to force its reactivation are considered by many as imperative milestones for HIV research [39]. The continuing study of mechanisms by which endogenous retroviruses are kept at bay in the genome of higher organisms is likely to provide clues towards meeting these goals.
Correspondence:
Professor Didier Trono
EPFL – FSV
Station 15
CH-1015 Lausanne
E-Mail: didier.trono@epfl.ch

References