FEBS Letters 581 (2007) 2877–2882

Minireview

Chimeric retrogenes suggest a role for the nucleolus in LINE amplification

Anton Buzdin[a,*], Elena Gogvadze[a], Marc-Henri Lebrun[b]

[a] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow 117871, Russia*
[b] *UMR2847 CNRS/Bayer CropScience, 14-20 rue Pierre Baizet, 69263 Lyon, Cedex 09, France*

Received 22 March 2007; revised 3 May 2007; accepted 16 May 2007

Available online 25 May 2007

Edited by Lev Kisselev

Abstract **Chimeric retrogenes, found in mammalian and fungal genomes, are bipartite elements composed of DNA copies of cellular transcripts either directly fused to each other or fused to the 3′ part of a LINE retrotransposon. These cellular transcripts correspond to messenger RNAs, ribosomal RNAs, small nuclear RNAs and 7SL RNA. The chimeras are likely formed by RNA template switches during reverse transcription of LINE elements by their retrotranspositional machinery. The 5′ part of chimeras are copies of nucleolar RNAs, suggesting that the nucleolus plays a significant role in LINE retrotransposition. RNAs from the nucleolus might have protective function against retroelement invasion or, alternatively, the nucleolus may be required for retrotranspositional complex assembly and maturation. These hypotheses will be discussed in this review.**

Keywords: L1 LINE; Retrotransposition; Chimeric retroelements; Reverse transcription; Nucleolus; U6 small nuclear RNA

1. Introduction

Reverse transcription is one of the key processes that shape eukaryotic genomes. At least 40% of the mammalian genome is resulting from reverse transcription events [1,2]. This phenomenon was discovered by Temin and Baltimore while they purified and characterized the first retroviral RNA-dependant DNA polymerase (reverse transcriptase, RT), which catalyzes the synthesis of complementary DNA using a RNA template [3]. Since, RT sequences were found in diverse retroviruses, mitochondrial group II introns, bacterial retrointrons, plasmids and in genetic elements termed retroelements (REs). REs are transposable elements that proliferate through RNA intermediates using self-encoded or exogenous RT and insert the newformed DNA copy of the element into the host genome.

Autonomous retroelements that carry their own RT genes can be subdivided into two major classes: long terminal repeat (LTR) containing elements, and non LTR retrotransposons [4]. Autonomous non LTR REs are mostly long interspersed nuclear elements (LINEs) found in essentially all eukaryotic genomes [5,6]. LINEs are 3.5–8 kb long, encode a RT and few other proteins necessary for their transposition [7]. LINEs also provide their RT enzyme for the proliferation of non-autonomous REs such as short interspersed nuclear elements (SINEs) [8]. Novel insertions of LINEs in a genome can be easily identified by the presence of 10–20 bp long target site duplications flanking the REs, that are formed during the integration process. LINEs contain canonical polyadenylation signals, oligo(A) tails or, sometimes, other A-rich sequences at their 3′-termini [9]. LINEs are transcribed by the cellular RNA polymerase II from an internal promoter located in their 5′-untranslated region [10]. Another LINE distinguishing feature is their frequent 5′-truncation that likely result from an interruption of LINE RNA reverse transcription, as RT could frequently dissociate from its RNA template before having completed a full cDNA synthesis [11].

Most LINEs found in eukaryotic genomes are inactive 5′-truncated copies that are transpositionally deficient while only a small number of actively transposing full-sized elements are present [12]. However, LINEs have frequently expanded during genome evolution as observed for the human genome that contains 5×10^5 L1 elements representing 17% of the total human genomic DNA [1,2]. The presence of such a number of LINEs in genomes affects many cellular processes. Highly repetitive LINE sequences may serve as recombination hot spots, causing frequent host DNA rearrangements [13]. Moreover, LINEs may disrupt preexisting gene exon-intronic structures [14] and in different ways interfere with host gene expression [15–17]. Another interesting property of LINEs is their ability to transfer their 3′-flanking DNA to new genomic loci, termed L1 transduction [18,19]. Taken together, this transduced DNA makes up ~0.6–1% of the human genome.

The full-sized LINE (+) RNA has a dual role as transpositional RNA intermediate and template for protein synthesis [20]. LINE transposition is known to proceed in several steps including RNA Pol II transcription of an active element, reverse transcription of the RNA formed with a LINE-encoded RT, and integration of the cDNA into a new position within the genome using an endonuclease [21] (Fig. 1A). A typical LINE element encodes two proteins: ORF1p that is a RNA binding protein which likely helps reverse transcription as a nucleic acid chaperone [22], and ORF2p, the reverse transcriptase and the endonuclease [21]. Due to a '*cis*-preference', the enzymatic machinery of a retrotransposition-competent LINE predominantly transposes its own copies [23] (Fig. 1A). However, LINEs are able to mediate the transposition of other sequences, mostly non-autonomous elements termed SINEs, but

*Corresponding author. Fax: +7 495 3306538.
E-mail address: anton@humgen.siobc.ras.ru (A. Buzdin).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.05.034

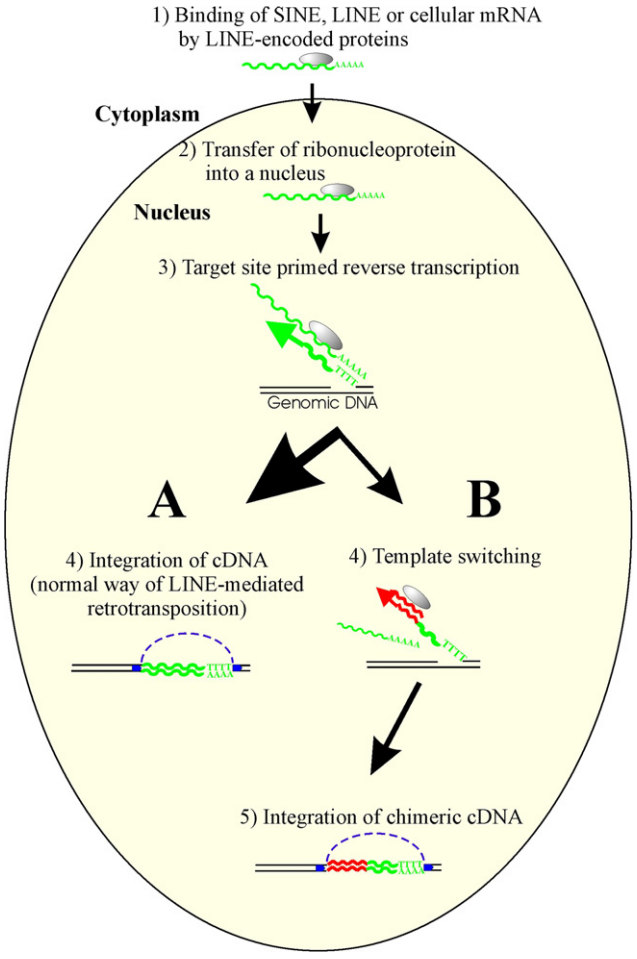

Fig. 1. Mechanism for the chimeras' formation by LINE enzymatic machinery. (Step 1) LINE pre-integration complex binds LINE, SINE or host mRNAs in the cytoplasm. (Step 2) The resulting ribonucleoprotein is transferred to the nucleus. (Step 3) Reverse transcription of the bound mRNA primed by a genomic DNA single-stranded break within the TTTTAA sequence (target site primed reverse transcription). (Step 4A) Successful integration of the reverse transcribed LINE cDNA copy into the genomic DNA. (Step 4B) Switch of templates to another RNA during the reverse transcription. (Step 5A) Integration of the chimera formed after RNA template switch into genomic DNA. This event leads to the formation of a bipartite chimeric retrogene carrying a poly(A) sequence at the 3′ terminus and flanked by short direct repeats. The normal LINE integration pathway is: Steps (1)–(2)–(3)–(4A).

also cDNAs originating from different cellular RNAs, leading to the formation of processed pseudogenes [24]. Recently, we have shown that LINEs are involved in the formation of bipartite chimeric retrogenes during reverse transcription in many genomes including human and fungi [25–29].

2. Template switching generates bipartite chimeric retrogenes

Bipartite chimeric retrogenes with an unusual structure were recently identified in three mammalian and in one fungal genomes (Fig. 2): a total of 82, 116, 66 and 31 elements were found in human, mouse, rat and rice blast fungus *Magnaporthe grisea* DNAs, respectively [25,27–29]. These elements are composed of DNA copies from cellular transcripts either directly fused

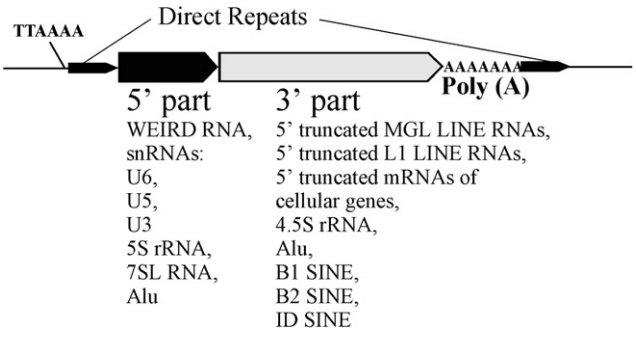

Fig. 2. Schematic representation of the bipartite chimeric retrogenes identified in eukaryotic genomes. The insertions in mammalian genomes are located downstream of the TTAAAA hexanucleotide motif. Insertions in mammalian and fungal genomes harbour poly(A) sequence and are flanked by 10–25 bp long genomic direct repeats.

to each other or more frequently fused to the 3′ part of a LINE retroposon. The various cellular transcripts found in these chimeras correspond to messenger RNAs, ribosomal RNAs, small nuclear RNAs, and 7SL RNA.

The chimeras have the following common features: (i) 5′-parts are full-length copies of cellular RNAs; (ii) 3′-parts are 5′-truncated copies of the corresponding RNAs (mostly LINEs); (iii) sites of these truncations occur at random in the corresponding RNA; (iv) both parts are directly joined with the same transcriptional orientation; (v) chimeras have a poly(A) tail at their 3′ end, and (vi) chimeras are flanked by short direct repeats.

The last structural feature demonstrate that these elements were transposed as bipartite DNA copies. Indeed, mammalian chimeras carried at their 5′ ends a T_2A_4 hexanucleotide or its variants [25,27,28] that correspond to the T_2A_4 genomic site used by LINEs to initiate reverse transcription on oligo(A) motifs and separate newly inserted DNA by short tandem repeats [30]. The simultaneous integration of both parts of these chimeras was further supported by the data came from PCR-based evolutionary insertion polymorphism assay [25,27].

The number of mouse and rat chimeras is likely underestimated as their 3′-terminal parts are often missing because they correspond to gaps in the genome sequence (23 and 33 cases for mouse and rat genomes, respectively). This significant sampling suggests that these bipartite elements are generated by a specific active mechanism. This mechanism frequently combines functional cellular transcripts that have nothing in common with transposable elements. Many of the chimeras can be considered as new genes, as they were shown to be transcribed, some of them in a tissue-specific manner [25,28,31] Gogvadze, 2007, unpublished for *M. grisea*.

We further hypothesised that these chimeric retrogenes were generated through a mechanism involving RNA recombination during the reverse transcription of cellular RNAs (Fig. 1B). This model includes a switch from the nascent cDNA serving as template for the reverse transcription of the 3′ part of the chimera to another RNA template corresponding to the 5′ part, followed by the chimera integration into the host genome [6]. Although RT main enzymatic activity is the continuous synthesis of the cDNA on RNA template, RT is able to switch templates during reverse transcription. For example, in retroviruses, RT jumps from one site of the RNA template to another site, are necessary for the synthesis