Gene 511 (2012) 46–53

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Sense transcripts originated from an internal part of the human retrotransposon LINE-1 5′ UTR

Elena A. Alexandrova [a], Ivan A. Olovnikov [b], Galina V. Malakhova [a], Anastasia A. Zabolotneva [a], Maria V. Suntsova [a], Sergey E. Dmitriev [b,*], Anton A. Buzdin [a,**]

[a] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya, Moscow 117997, Russia
[b] Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119234, Russia

ARTICLE INFO

Article history:
Accepted 4 September 2012
Available online 13 September 2012

Keywords:
Human transposable element
LINE1
Retrotransposon L1 5′ UTR
Promoter activity
Transcription start site mapping
Transcriptional enhancer

ABSTRACT

L1 (LINE-1) is one of the most abundant families of human transposable elements. Full-length human L1 has an ~900 bp long 5′ untranslated region (5′ UTR) which harbors an internal promoter for the RNA polymerase II. It is generally accepted that the first 100 bp of the 5′ UTR function as a "minimal promoter" which directs transcription of the entire LINE-1 unit from the extreme 5′ terminus. We re-investigated promoter activities of the different DNA fragments that cover the whole L1 5′ UTR in cultured human cells by using the luciferase reporter system. Analysis of both mRNA expression and luciferase activity levels indicated that the very important region for the effective transcription is located within the internal part of the L1 5′ UTR between nucleotide positions +390 and +526. 5′ RACE analysis revealed that in the context of the complete 5′ UTR, this part drives mRNA synthesis both from the canonical 5′-terminal transcription start site (TSS) and from within the internal region. In the absence of the first 100 bp, the L1 5′ UTR efficiently directed transcription from aberrant TSSs located within its 3′ proximal part or the ORF1. Finally, we analyzed transcripts originated from endogenous (genomic) L1 elements and identified two novel TSSs located at positions +525 and +570. We propose a model in which the internal part (390–526) of the L1 5′ UTR plays a key role for recruitment of transcription initiation complex, which then may be either positioned onto the 5′ terminally located "minimal promoter", or used proximately to direct 5′ truncated RNA copy. Intriguingly, this internal regulatory element substantially overlaps with the region of the L1 5′ UTR that is known to drive transcription in the opposite direction suggesting the existence of a common core for the bidirectional transcription.

1. Introduction

Long interspersed nuclear element-1 (LINE-1; L1) belongs to the class I transposable elements (retrotransposons) proliferating in the host genome through the mechanism termed retrotransposition (Ostertag and Kazazian, 2001). Retrotransposition includes a stage of the reverse transcription of an RNA-copy of a retrotransposon. The human genome harbors more than 500 000 L1 copies comprising ~17% of the human DNA (Lander et al., 2001). The majority of the L1s are represented by the 5′-truncated copies which are incapable for retrotransposition. Estimation of the retrotransposition-competent L1 copy number revealed that only 80–100 elements in the human genome may be able to retrotranspose (Brouha et al., 2003). Full-length human L1 is ~6-kb long and consists of 5′ untranslated region (UTR), two non-overlapping protein coding open reading frames ORF1 and ORF2 and a 3′ UTR (Babushok and Kazazian, 2007; Ostertag and Kazazian, 2001). When the genomic copy of retrotransposition-competent L1 is transcribed, its full-length mRNA is exported to the cytoplasm, where the ORF1 and ORF2 are translated (Alisch et al., 2006; Dmitriev et al., 2007; Hohjoh and Singer, 1996; Martin, 1991).

The ORF1p possesses nucleic acid chaperone activity (Martin, 2010), whereas the ORF2p has endonuclease (Feng et al., 1996) and reverse transcriptase (RT) (Mathias et al., 1991) activities. Both protein products bind preferentially to the self-encoding L1 mRNA (so called *cis*-preference effect (Wei et al., 2001)) followed by the formation of ribonucleoprotein complexes consisting of the L1-encoded RNA and protein products (Doucet et al., 2010). During the process termed "target site-primed reverse transcription" (TPRT), a new double stranded DNA-copy of L1 element appears in genomic DNA (Cost et al., 2002).

L1s are thought to be the only active family of the human autonomous transposable elements. L1 elements influence the host genome

Abbreviations: LINE1, long interspersed nuclear element 1; UTR, untranslated region; RACE, rapid amplification of cDNA ends; TSS, transcription start site; ORF, open reading frame; TPRT, target site-primed reverse transcription; RLU, relative light unit; qRT-PCR, quantitative real-time reverse transcription-PCR; CMV, cytomegalovirus.

* Corresponding author. Tel.: +7 495 9394857.
** Correspondence to: A.A. Buzdin, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117997, Russia. Tel./fax: +7 495 727 38 63.
E-mail addresses: dmitriev_sergey@genebee.msu.ru (S.E. Dmitriev), bu3din@mail.ru (A.A. Buzdin).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.09.026

structure and function through a wide range of different mechanisms, including insertional mutagenesis, ectopic recombinations, 5′ and 3′ flanking sequence transduction, formation of pseudogenes and even of novel functional copies of genes, disruption of the pre-existing exon-intronic structures, altered epigenetic regulation of genes, and trans-mobilization of non-autonomous retrotransposons (Babushok and Kazazian, 2007; Bantysh and Buzdin, 2009; Gogvadze and Buzdin, 2009; Schumann et al., 2010).

Human L1 element transcription is driven by the quite unusual RNA polymerase II internal promoter located within their long 5′ UTR (Swergold, 1990). In general, the 5′-terminal parts of mammalian L1s are highly variable, so that the 5′ UTR sequences and the first thirds of the ORF1 gene share no sequence identity among the human, mouse and rat L1 elements (Furano, 2000). The presence of an internal promoter is of an extreme importance for the successful retrotransposition of the L1 elements. Most importantly, it allows transcription of the full-length L1 mRNA, which in turn can be converted to a complete active copy of the retrotransposon when inserted into a new location within the genome (Kimberland et al., 1999).

The detailed mechanism of the L1 internal promoter activity is largely unknown, but some of its structural and functional features have been decoded by now. The classical model of RNA-polymerase II eukaryotic promoter comprises several canonical regulatory sequences such as the TATA-box, located ~30 bp upstream of the initiator element containing the TSS. However, there are a growing number of experimental observations evidencing that the TATA-box dependent RNA pol II transcriptional initiation mechanism may be rather the exception, than the rule. Quantitative characterization of the mammalian promoters revealed that about 80–90% of them miss functional TATA box (Cooper et al., 2006; Gershenzon and Ioshikhes, 2005). Unique internal promoters of LINE retrotransposons belong to the group of TATA-less promoters, and a mechanism of their action is almost totally uncharacterized in molecular terms.

In the human L1, it was found that the most important part of the 5′ UTR for promoter activity is likely located within the first 668 bp-long region (Swergold, 1990). This region includes various transcription factor binding sites that may regulate L1 expression in different ways. According to the experiments carried out in 1990, deletion of the 5′ terminal ~100 bp long fragment led to a complete loss of the 5′ UTR promoter activity (Swergold, 1990). This finding suggested, therefore, that the strongest L1 promoter region is located within the first 100 bp of its 5′ UTR. More recently, Minakami et al. (1992) came to a similar conclusion, when studying promoter activity of the 287 bp long 5′-proximal region of the 5′ UTR. Finally, the discovery of the functional binding sites for the transcription factors YY1 and RUNX3 at positions +13…+21 (Becker et al., 1993) and +83…+101 (Yang et al., 2003), respectively, strengthened these observations and allowed to term the 5′ terminal segment of the human L1 5′ UTR the "minimal promoter".

However, some contradicting reports suggested that the 5′-terminal YY1 binding site *per se* is not necessary for efficient transcription (Athanikar et al., 2004) and that the remaining part of the 5′ UTR may also contribute significantly to the overall promoter activity (Muotri et al., 2005; Steinhoff and Schulz, 2003). Importantly, the internal portion of the human L1 5′ UTR (position +100…+668 bp) also has numerous binding sites for the transcription factors of the SRY-related (SOX) family (positions +472…+477 and +572…+577) (Tchenio et al., 2000), RUNX3 (+526…+508) (Yang et al., 2003), p53 (+452…+466) (Harris et al., 2009) and the Ets family (482–491) (Yang et al., 1998) that serve as transcriptional activators. The internal part of the human L1 5′ UTR has been shown to harbor also an antisense promoter, which drives transcription in the opposite orientation relatively to the conventional L1 mRNA (Speek, 2001). The functional role of the antisense promoter is still poorly understood and may represent a kind of coadaptation of the L1 with the host genome (Chen et al., 2012; Nigumann et al., 2002; Yang and Kazazian, 2006).

Recently, we reported that the human L1 5′ UTR lacking the 5′-terminal ~100 bp, is still capable to direct efficient reporter expression in transiently transfected human cells (Olovnikov et al., 2007). This data argue with the commonly accepted concept of the "minimal promoter", although may only be regarded as preliminary, since, first, only the protein level of the reporter was measured in that study, and second, a foreign regulatory element (downstream SV40 enhancer) in addition to L1 5′ UTR-derived sequences was used. It is well known that post-transcriptional events may significantly bias reporter expression: for example, Belancio (2011) recently reported that alternative splicing events may disrupt open reading frame for the reporter gene placed downstream of the L1. Thus, determining mRNA concentrations for the respective gene constructs would be a more adequate way of analyzing L1 promoter strength. It should be also important to exclude any potential impact of the foreign enhancer.

Nevertheless, the results of Olovnikov et al. raised questions on possible existence of alternative forward transcription start site (TSS) within the internal part of the human L1 5′ UTR, or some kind of internal enhancer located there. These intriguing possibilities required additional study. Here, we addressed these questions by investigating the L1 promoter activity in a context lacking any foreign regulatory sequences, and by measuring reporter gene expression at both transcriptional and protein levels. We performed a systematic screen of the entire 907 bp-long human L1 5′ UTR and its shorter segments for promoter activity in the transient transfection experiments. Our data confirm that the region of the "minimal promoter" is not crucial for efficient L1 transcription. However, its deletion resulted in the appearance of novel TSSs, located at the 3′-terminal part of the 5′ UTR and at the beginning of the ORF1. Moreover, we detected internally initiated transcripts also in the presence of the "minimal promoter". We also found similar transcripts in non-transfected human cells where they originated from endogenous L1 copies. These results allow us to conclude that the key role in transcription initiation activity of the human retrotransposon L1 promoter is played by the internal region of its 5′ UTR which forms a core for initial assembly of a transcriptional activation complex. We propose that this complex then may be either recruited onto the 5′ terminus by the "minimal promoter" to direct classical full-length L1 transcripts, or positioned onto the internally located alternative TSSs identified in this study.

2. Materials and methods

2.1. Plasmid constructs

Plasmids were engineered on the basis of the pGL3 vector (Promega). The original construct L1wtE was described (Olovnikov et al., 2007). It contained a fragment corresponding to positions +1…+952 of the human L1-Ta consensus sequence. The fragment was subcloned from p1LZ plasmid, a gift from Dr. G. Swergold (Columbia University, USA), and originated from a cDNA obtained by Skowronski et al. (1988) and used by Swergold (1990) in his study of L1 promoter. To delete region corresponding to a functional part of SV40 enhancer, L1wtE was treated with BamHI and AvaIII, blunt-ended and ligated. The resulting construct (hereafter referred to as L1wt) contains the first 952 nt of L1Hs corresponding to full-length 5′ UTR region and the first 15 codons of the L1 ORF1 (Fig. 1A). Constructs with the 5′ UTR deletions were prepared by L1wt plasmid digestion with PvuII-Acc651 (L1Δ(1–98)), PvuII-NheI (L1Δ(1–386)), PvuII-BglII (L1Δ(1–664)), Acc65I-NheI (L1Δ(103–386)), NheI-StuI (L1Δ(390–526)), and StuI-BglII (L1Δ(527–662)) after which plasmids were blunt-ended and ligated. Constructs L1Δ(667–887) and L1 Δ(133–887) were prepared by L1wt plasmid digestion with BglII-BstXI and PflMI-BstXI respectively, after which plasmids were ligated with the use of double-stranded DNA adapters prepared by annealing