Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 340 (2006) 553–559
BBRC
www.elsevier.com/locate/ybbrc

Identification of the proteins specifically binding to the rat LINE1 promoter

Anton V. Fedorov *, Dmitry V. Lukyanov, Olga I. Podgornaya

Institute of Cytology, Russian Academy of Sciences, Tikhoretsky pr. 4, 194064 St-Petersburg, Russia

Received 6 December 2005
Available online 19 December 2005

Abstract

The initial step of LINE1 retrotransposons dissemination requires transcription from species-specific promoter located within 5′-untranslated region of LINE1. Although the 5′-untranslated region of the rat LINE1 element shows promoter activity, no promoter-binding proteins have been discovered so far. Using an EMSA and Southwestern blotting methods, we identified Sp1 and Sp3 proteins, which specifically bind to the rat LINE1 promoter in vitro. The Sp1/Sp3-binding motif within rat LINE1 promoter is located downstream of the major predicted transcription initiation site.

Keywords: Retrotransposon; LINE1; Sp1; Sp3; Promoter-binding proteins

Long interspersed nuclear elements-1 (LINE-1 or L1 sequences) belong to the superfamily of autonomous retrotransposons and account for about 10–20% of mammalian genomes [1,2]. L1 elements are able to propagate in the genome by a reverse transcription mechanism via an RNA intermediate; the integrated L1 is generally flanked by variably sized target site duplications (TSDs) [3]. L1 elements encode two proteins: one is RNA-binding protein that specifically interacts with L1 transcripts to form ribonucleoprotein particles [4,5] and the second one possesses reverse transcriptase and endonuclease activities [6,7]. Both the proteins are assumed to be required for retrotransposition of LINEs [8]. L1 elements are thought to be transcribed by RNA polymerase II (Pol II). Transcription of full-length RNA, the first step in the L1 retrotransposition, provides a template for synthesis of both L1-encoded proteins and a DNA copy, which then becomes integrated into the genome [1,2]. Despite their prominence and their impact on the genome, mechanisms of the L1 transcription regulation remain largely unknown.

The 5′-untranslated regions (5′UTRs) of L1 elements are known to have promoter activity, these regions are GC-rich and lack conserved TATA-box [9–11]. The 5′UTRs of rat, mouse, and human L1 repeats are not homologous in contrast to the rest of the L1 elements [9,12,13]. Such variability could be explained if one supposes a repeated acquisition of novel 5′-sequences during species divergence [2]. In contrast to human, 5′UTRs of rat and mouse L1 elements contain a variable number of tandemly repeated units, called monomers, followed by a short non-monomeric region [9,10,12]. In mouse there were found several distinct L1 subfamilies that are defined by different types of monomers (A, F, Tf, and Gf) within their 5′UTRs [10,12,14]. Interaction of nuclear proteins with 5′UTRs of L1 elements could probably be one of the mechanisms regulating L1 expression. Species-specificity of L1 promoter sequences makes the range of proteins binding to them of particular interest. It was shown that the YY1 factor could bind 5′UTR of human L1 [15]. Mutations in the YY-1 binding site have only a minor effect on transcription activation of human L1; instead, these mutations

* Corresponding author. Fax: +7 812 2470341.
E-mail address: an_tn@mail.ru (A.V. Fedorov).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.12.040

disrupt the proper transcription initiation [16]. Two SRY-related transcription factor binding sites and a RUNX3 transcription factor binding site have been identified within human L1 5′UTR as well [17,18]. Factors involved in the regulation of murine L1 transcription have not been identified yet.

Rat L1 (L1Rn) 5′UTR contains tandemly repeated monomers of about 600 bp [9,19]. The promoter activity of L1Rn 5′UTR lies within the monomer sequence, which was named rat L1 promoter [20]. It was shown that L1Rn promoter could bind nuclear protein and the sequence motif responsible for such interaction was recognized [2]. Still, the proteins specific in binding to this sequence have not been identified.

The goal of the present work was to identify DNA-binding proteins which specifically recognize L1Rn monomer sequence. We herein identify Sp1 and Sp3 proteins as factors which interact in vitro with the rat L1 promoter.

Materials and methods

DNA and plasmids. The 408 bp fragment of L1Rn 5′UTR was amplified from rat genome DNA by PCR using the primers L1For 5′-AAGTCCCACACCCGCGATC-3′ and L1Rew1 5′-GTGCACCAGGG TTCCAGAT-3′; the PCR product was cloned into the pCR4 TOPO vector (Invitrogen). The insert was sequenced and termed PRLP (Part of Rat LINE1 Promoter) (GenBank Accession No. DQ279827), and the respective plasmid was termed pPRLP. The fragment PRLP corresponds to positions 47–457 of the rat L1 promoter monomer consensus [19].

Preparation of nuclear extracts. Nuclei from testis cells of 1-month-old rats were isolated as described [21]. Full-length L1 transcripts were found at precise stages of germ cell development [22]; thus, we used rat testis as the source of nuclear extract. Purified nuclei were extracted for 1 h at 4 °C in 10 mM Tris–HCl, pH 8.0, 0.1 mM $MgCl_2$, 0.2 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol, 5% glycerol, and 0.35 M NaCl, centrifuged for 10 min at 8000*g*, and the supernatant nuclear extract was collected.

Electrophoretic mobility shift assays (EMSAs) and supershift assays. The 408 bp PRLP fragment was amplified from pPRLP and labeled with [α-^{32}P]dATP (Amersham, UK) by PCR, using the primers L1For and L1Rew1, and then purified by electrophoresis in 0.5% agarose [23]. The sequences of the oligonucleotides used in the DNA binding analysis were as follows: nodel 5′-TCAGGTGGGCACTCC-3′, shor 5′-GATCTC AGGTGGGATAC-3′, delb 5′-AGATCCACTCCATACGA -3′, Sp1wt 5′-ATTCGATCGGGGCGGGGCGAGC-3′, Sp1mut 5′-ATTCGATC GGTTCGGGGCGAGC-3′, YY1wt 5′-ATGTCAGCGCCATCTTGT AAG-3′, and Ebox-L 5′-AGATCCAGGTGATACGA-3′. Each listed oligonucleotide had an antiparallel partner for annealing. Nodel oligonucleotide and its antiparallel partner have 5′GATC overhangs for labeling. Specific consensus binding sequences for Sp1 [24] and YY1 [25] transcription factors are underlined, and the mutated nucleotides are given in boldface. Fragments of nodel oligonucleotide within sequences shor, delb, and Ebox-L are double underlined. Nodel oligonucleotide corresponds to positions 125–139 of the rat L1 promoter monomer consensus [19].

Complementary strands of oligonucleotides were annealed by heating to 95 °C for 5 min and then slowly cooling to room temperature. Nodel oligonucleotide was end-labeled with [α-^{32}P]dATP (Amersham, UK) using Klenow fragment (Sileks, Russia) and purified by electrophoresis in 12% polyacrylamide gel [23].

Binding reactions for electrophoretic mobility shift assays contained 10,000 cpm of probe DNA and 3 μg of nuclear extract in the total volume of 40 μl comprising: 10 mM Tris–HCl, pH 7.5, 2 mM $MgCl_2$, 1 mM dithiothreitol, and 4% (v/v) glycerol. The final concentration of NaCl did not exceed 0.15 M. Sonicated *Escherichia coli* DNA (the mean length ~800 bp) was included into each reaction as a non-specific competitor. The 437 bp part of L1Rn 3′UTR, termed 3UTR amplified from the rat genome DNA by PCR using the primers E21 5′-AACAGAGA CTGAAGGAA-3′ and E23 5′-AGTCTAGTTCACTGGGG-3′ [26], was used as a non-specific competitor. Unlabeled probes and oligonucleotides, at various molar ratios to ^{32}P-labeled probe, were added as specific competitors. The complete binding reactions were incubated for 30 min at room temperature and protein–DNA complexes were resolved by electrophoresis (20 mA, 3 h) on either 4% (for PRLP probes) or 6% (for nodel probes) polyacrylamide gels in 0.5× TAE buffer (20 mM Tris–acetate, pH 8.0, 1 mM EDTA) [23] and visualized by autoradiography. For the antibody supershift assay, 2–3 μg of specific rabbit polyclonal antibodies against Sp1, Sp3 or YY1 (all from Santa Cruz Biotechnology, USA) was added and incubated with the reaction mixture for an additional 50 min at room temperature.

Western blotting. Nuclear proteins and prestained protein molecular weight standard (Precision Plus Protein standard Dual Color, Bio-Rad) were separated in 10% SDS–polyacrylamide gels [27], electrotransferred onto a Hybond-P membrane (Amersham, UK), and immunoblotted by standard methods [28], using polyclonal antibodies against Sp1, Sp3 or YY1 (all from Santa Cruz Biotechnology, USA). Goat anti-rabbit antibodies, conjugated with alkaline phosphatase (Santa Cruz Biotechnology, USA), were used for the detection of primary antibodies. All antibodies were used in the dilutions recommended by the manufacturer.

Southwestern blotting. Nuclear proteins were separated and electrotransferred onto a Hybond-P membrane (Amersham, UK) as described above. The membrane was incubated in binding buffer B (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, 50 μM $ZnSO_4$, and 1 mM dithiothreitol) with 2.5% BSA for 5 h at room temperature and then in the same buffer with 0.25% BSA and either 120,000 cpm/ml ^{32}P-labeled PRLP probe with 50 μg/ml sonicated *E. coli* DNA or 120,000 cpm/ml ^{32}P-labeled nodel probe with 10 μg/ml sonicated *E. coli* DNA for additional 5 h. Unlabeled DNA probes were added as specific competitors. The membrane was finally washed three times with buffer B, dried, and analyzed by autoradiography.

Computer-based analyses. DNA sequence alignment procedures were carried out using BLAST (Basic Local Alignment Search Tool) [29]. The transcription factor sites were analyzed using TESS (Transcription Element Search System) (http://www.cbil.upenn.edu/tess). For the analysis of full-length L1Rn elements we used database L1Base [30].

Results

Analysis of L1Rn promoter monomer

L1Rn promoter is enriched in CpG dinucleotides and lacks a conserved TATA-box [9]. Some Pol II-dependent promoters do not have TATA sequences, but instead do contain initiator (Inr) elements that can independently direct transcription initiation [31]. Therefore, we searched for consensus Inr motifs (C/T)(C/T)A_{+1}N(T/A)(C/T)(C/T) (where A_{+1} is the common transcription start site) [32] within L1Rn promoter. We found six putative Inr elements in L1Rn promoter (Fig. 1). Transcription of *Drosophila melanogaster* Doc element, which belongs to LINE retrotransposons, is regulated by Inr motif in cooperation with downstream promoter element (DPE); DPE is located precisely at +28 to +32 relative to the central A_{+1} nucleotide of the Inr [33]. We failed to find consensus DPE motif (A/G)G(A/T)(C/T)GT [34] in the proper location to predicted Inr elements within rat L1 promoter.