Nucleic Acids Research, Vol. 20, No. 12 3139–3145

Identification of an internal *cis*-element essential for the human L1 transcription and a nuclear factor(s) binding to the element

Reiko Minakami[1], Kouichi Kurose[1], Kanako Etoh[1], Yoshiaki Furuhata[1], Masahira Hattori[2] and Yoshiyuki Sakaki[1,2]*

[1]Research Laboratory for Genetic Information, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812 and [2]Laboratory of Molecular Medicine, The Institute of Medical Science, The University of Tokyo, 4-6-1, Shiroganedai, Minato-ku, Tokyo 108, Japan

Received February 18, 1992; Revised and Accepted May 28, 1992 EMBL accession no. X52235

ABSTRACT

L1 (LINE-1) is a long interspersed repetitive sequence derived from a retrotransposon. Transfection studies using the CAT gene as a reporter demonstrated that the first 155bp in the human L1 sequence contains an element(s) responsible for the promoter activity in HeLa cells. The transcription was shown to initiate at the first nucleotide of the L1 sequence in the transgene. Three prominent nuclear protein binding sites were found in the 5′ region of the L1 sequence by DNaseI footprint analysis. One of the binding sites, designated as site A located at +3 to +26, was shown to be essential for the L1 transcription because the mutation at the site A caused almost complete loss of the promoter activity. A sequence AAGATGGCC at +11 to +19 in the site A was defined as a target core element for the protein binding. The site A–binding protein (designated TFL1-A) was found in various types of cells including an embryonic teratocarcinoma cell line. These results indicate that an internal short element located at the very 5′ terminal of L1 sequence and the nuclear factor binding to the element play a crucial role in the transcription of human L1.

INTRODUCTION

L1 (or LINE-1) is a highly repeated DNA sequence family interspersed in all mammalian genomes (see refs 1 and 2 as reviews). Sequences related to mammalian L1s are also found in other organisms including insects (3–7), protozoa (8, 9) and plant (10). The full-length unit of L1 is about 6kb, contains evolutionarily conserved regions which encode two open reading frames (ORF1 and ORF2) (1, 11, 12). The ORF2 product was shown to have a significant sequence homology with various reverse transcriptases (11, 12). These structural features strongly suggested that L1 is a family of retrotransposon-derived sequence and that L1 has dispersed on the genome through the RNA intermediate which is converted to cDNA by its own reverse transcriptase. The full-length or nearly full-length, polyadenylated L1 transcripts were actually identified in several cell lines including a human embryonic teratocarcinoma NTera2D1 (13, 14). Recently, the product of ORF1 was evidently detected in NTera2D1 and other cells (15), a human active L1 element was isolated (16) and the reverse transcriptase activity was verified (17). These findings strongly suggested that some genomic L1 sequences are actually transcribed to generate functionally active L1 mRNAs. However, the molecular mechanism of the L1 transcription, the first step of the L1 transposition, has been still obscure. Unlike retroviruses and known retrotransposons, L1s have no terminal repeats such as long terminal repeats (LTRs) which serve as a promoter. An obvious question is arisen how L1s are transcribed. Formally, two types of promoters for the L1 transcription can be considered; the promoter within the L1 unit or the one outside the L1 unit. Several investigators demonstrated the promoter activity in the 5′ terminal regions of L1s or L1-like elements including human L1 (18–21), strongly suggesting that L1 is transcribed using its own internal promoter. The presence of the promoter inside of the L1 sequence implies that the L1 promoter should have a unique machinery for starting the transcription at an upstream site, *i.e*, the first nucleotide of L1 to generate the full-length L1 RNAs. For characterizing this unique promoter, we attempted to find the *cis*-elements and *trans*-acting factors involved in L1 transcription.

We herein describe the identification of an internal short element essential for the L1 transcription and a nuclear protein binding to the element. We also show that the factor (designated TFL1-A) is widely distributed in various types of cells.

MATERIALS AND METHODS

Materials

Restriction enzymes and other enzymes were obtained from Takara Shuzo (Kyoto, Japan), Toyobo (Tokyo, Japan) and Nippon Gene (Toyama, Japan). Chemicals were obtained from Wako (Tokyo, Japan), Katayama (Osaka, Japan) and Sigma

* To whom correspondence should be addressed at: Laboratory of Molecular Medicine, The Institute of Medical Science, The University of Tokyo, 4-6-1, Shiroganedai, Minato-ku, Tokyo 108, Japan

(USA). Radioactive compounds were purchased from Amersham (UK) and New England Nuclear (USA). Oligonucleotides were synthesized with trityl-off mode by an Applied Biosystems automated DNA synthesizer (model 380A) and purified through NAP 10 column (Pharmacia). Plasmids pCAT-Basic and pCAT-Control were obtained from Promega.

Cell culture

ITO-2 teratocarcinoma cells (22) were provided from Prof. J.Hata (Keio University) and JEG-3 choriocarcinoma were from Dr. Y.Misumi (Fukuoka University). HeLa epidermoid carcinoma cells were from our laboratory collection. All cells were maintained at 37°C in a humidified 5% CO_2 atmosphere. HeLa cells were grown in Dulbecco's Modified Eagle Medium (DMEM) and JEG-3 and ITO-2 cells were in RPMI-1640 medium. Both media were supplemented with 10% fetal bovine serum (GIBCO) and appropriate antibiotics.

Construction of plasmids

CGL1-1 and CGL1-3 (23) were employed for the construction of plasmids used in this study. Plasmid pCGL1-1SS was constructed by the insertion of *Ssp*I-*Ssp*I fragment from −102 to +287 of CGL1-1 into *Sma*I site of pUC12 and was used to prepare probes for DNaseI footprint analysis. Plasmid pCGL1-1/−24/+287 was constructed by insertion of the fragment from the *Dra*I site at −27 to the *Ssp*I site at +285 of CGL1-1 into the blunt-ended *Acc*I site of pCAT-Basic (Promega). The pCGL1-1/−24/+155 was constructed by the digestion of pCGL1-1/−24/+287 with *Bss*HII (at +155 in the L1 sequence) and *Xba*I (in the polylinker located at the downstream of the L1 sequence) followed by self-ligation of the blunt-ended sites. The pCGL1-1/−24/+40 was constructed in the same manner using *Acc*I (at +37 in the L1 sequence) and *Xba*I. The pCGL1-1/−24/+14 was constructed from a synthetic double-stranded oligonucleotide containing the sequence from −24 to +14 of CGL1-1 and pCAT-Basic. The pCGL1-1/+152/+287 was constructed by the digestion of pCGL1-1/−24/+287 with *Hin*dIII (in the polylinker located at the upstream of the L1 sequence) and *Bss*HII followed by self-ligation of the blunt-ended sites. The pCGL1-1/+39/+287 was from pCGL1-1/−24/+287 using *Hin*dIII and *Acc*I digestion in the same manner. A sequence from −8 to −3 in pCGL1-1/−24/+287 was replaced by *Eco*RV site by *in vitro* mutagenesis to construct the mutant plasmid pCGL1-1/−24/+287/*Eco*RV. The pCGL1-1/−24/−5 was constructed from the pCGL1-1/−24/+287/*Eco*RV by the digestion with *Eco*RV and *Xba*I followed by self-ligation of the blunt-ended sites. The pCGL1-1/−5/+287 was obtained from the pCGL1-1/−24/+287/*Eco*RV by the digestion with *Pst*I (in the polylinker located at the upstream of the L1 sequence) and *Eco*RV followed by the blunt-end self-ligation and the pCGL1-1/−5/+155 was constructed by subsequent digestion of the pCGL1-1/−5/+287 with *Bss*HII and *Xba*I followed by the blunt-end ligation. The pCGL1-1/−5/+155A and pCGL1-1/−5/+155B were constructed using pCGL1-1/−5/+155 and mutated oligonucleotides by *in vitro* mutagenesis, respectively. The pCGL1-1/+18/+155 was constructed from the digestion of pCGL1-1/−5/+155A with *Hin*d III and *Eco*R V followed by the blunt-end self-ligation. The orientation of the fragment in the constructs was determined by digestion with appropriate enzymes and the mutagenized sequences were

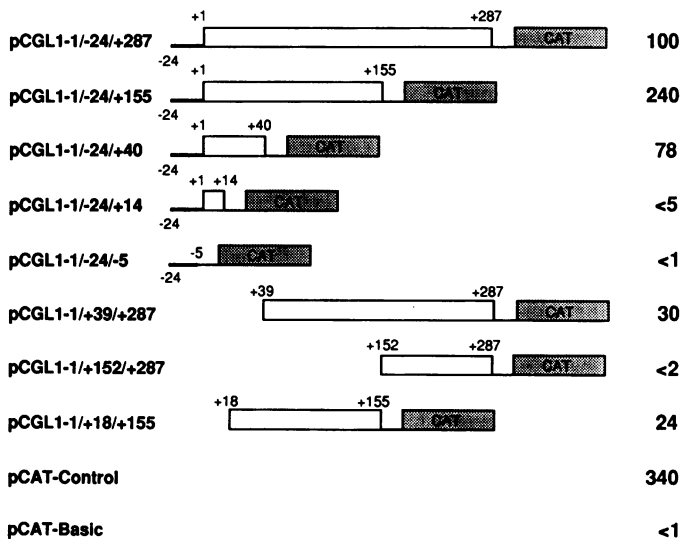

Fig. 1. CAT activity of various L1-CAT constructs in HeLa cells. Construction of L1-CAT fusion plasmids and the experimental conditions were described in Materials and Methods section. Open boxes show the L1 sequence and CAT-coding sequence shown by shadowed boxes. Flanking genomic sequences are depicted by bold lines. Numbers with plus indicate the nucleotides from the 5' end of CGL1-1 (23) and +1 denotes the first nucleotide of the L1 sequence. The CAT activity was normalized by the protein amounts and the luciferase activity of an internal control pRSVLUC (26). The activity of each construct is shown at the right as the value relative to that of pCGL1-1/−24/+287. The pCAT-Basic is a promoter-less CAT vector and pCAT-Control is the CAT construct in which the SV40 early promoter with the enhancers is fused to the CAT gene.

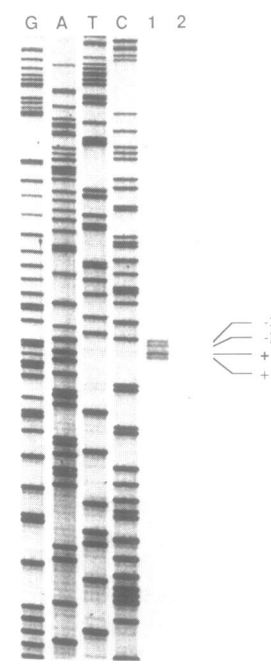

Fig. 2. Primer extension analysis of the L1 transgene transcripts in transfected cells. Left four lanes show a sequence ladder as a size marker. Extension products from HeLa cells transfected by pCGL1-1/−5/+155 are shown in lane 1 and the products from nontransfected cells in lane 2. Arrows with numbers at the right indicate the position of the extension products. The first nucleotide of CGL1-1 sequence is taken as +1.

confirmed by a denatured plasmid DNA sequencing (24). The plasmids used for the transfection experiment were purified by the CsCl-equilibrium ultracentrifugation.

DNA transfection, CAT assay and luciferase assay

Transfection was done by a calcium phosphate method (25) using 20 μg of test plasmid and 2 μg of pRSVLUC, a RSV promoter-firefly luciferase gene fusion plasmid (26). The pRSVLUC was used as an internal standard to monitor the transfection efficiency by measuring luciferase activity. About 10^6 cells were transfected and cultured for 4 hrs with the calcium phosphate-DNA precipitate. The cells were treated with 15% of glycerol for 30 sec and cultured for further 48 hrs. Cell extracts were prepared and assayed for CAT and luciferase activities according to the procedures described by Gorman *et al.* (27) and de Wet *et al.* (26). Transfection experiments were done at least three times for each construct and the CAT activity was normalized by the protein amounts and luciferase activities.

Primer extension

RNA was isolated from transfected and nontransfected HeLa cells by the guanidine-thiocyanate method (28). The primer extension was done using 50 μg of total RNA and 5×10^5 cpm of a ^{32}P-labeled primer. The reaction mixture was denatured at 85°C for 10 min, annealed at 50°C for 10 min and then incubated with AMV reverse transcriptase for further 60 min at 42°C. The products were analyzed by electrophoresis on a 6% polyacrylamide-urea gel. The complementary oligonucleotide (5′-TCCATTTTAGCTTCCTTAGCTCCTGAAAATC-3′) in the CAT gene was used as a primer. The size marker sequence ladder was obtained by sequencing of the pCGL1-1/−5/+155 with the primer same as that used in the primer extension experiment.

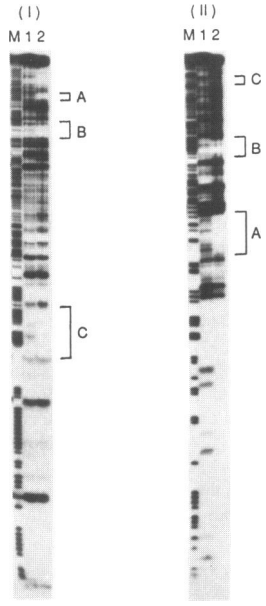

Fig. 3. DNaseI footprint analysis of 5′ region of the CGL1-1. The fragment containing the sequence from −102 to +287 of CGL1-1 was used as a probe for coding strand (I) and noncoding strand (II). M shows the Maxam-Gilbert G+A sequence products (35) as a size marker. Lanes 1 and 2 show the footprinting patterns in the absence and presence of nuclear extracts, respectively. Three protected regions are indicated by A, B and C at the right.

DNaseI footprint analysis

Nuclear extracts were prepared from HeLa, JEG-3 and ITO-2 cells according to the method of Shapiro *et al.* (29). The extracts equivalent to 17 μg of proteins were used in the binding reaction. Probes prepared from pCGL1-1SS contained the sequence from −102 to +287 of CGL1-1. The binding reaction was carried out in a 40 μl of the extract, ^{32}P-labeled probes (5×10^4 cpm), 10mM Hepes (pH 7.9), 50mM KCl, 0.1mM EDTA, 0.1mM EGTA, 1mM dithiothreitol, 5mM $MgCl_2$, 10% glycerol and 5 μg of poly(dI-dC) for 15 min on ice and then for 5 min at room temperature. Subsequent DNaseI digestion was done for 2 min at room temperature using appropriate amounts of enzyme. The footprinting pattern was analyzed by electrophoresis on a 6% polyacrylamide-urea gel.

Gel shift assay

Gel shift assay (30) was performed using nuclear extracts described above and 26bp of double-stranded oligonucleotides containing the 5′ region of CGL1-1. The sequences of the oligonucleotides are shown in Fig. 6B. The binding reaction was carried out for 20 min at room temperature in a 20 μl of the same buffer as that used for the DNaseI footprint analysis except 3 μg of poly(dI-dC), 6 μg of proteins and 5×10^4 cpm of ^{32}P-labeled probe and 3 pmol of competitor oligonucleotides (300-fold excess), if necessary. The samples were analyzed by electrophoresis on a 5% polyacrylamide gel at room temperature.

RESULTS

Promoter activity of the recently transposed L1 sequence

We previously isolated several full-length human L1 members which were inserted into the genome in a relatively recent past (23). We asked whether the 5′ terminal regions of these L1s designated CGL1s serve as a promoter. At first, we constructed a L1-CAT fusion plasmid in which a fragment from −24 to +287 of CGL1-1 (23) was fused with the bacterial chloramphenicol acetyltransferase (CAT) gene (27). The construct, pCGL1-1/−24/+287, was introduced into HeLa cells by a calcium phosphate method (25) and the promoter activity was monitored

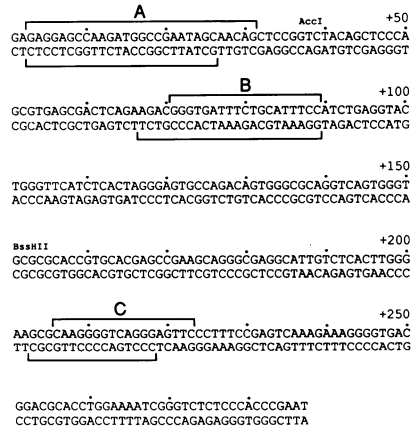

Fig. 4. Nucleotide sequences protected by nuclear factors. Numbers indicate the nucleotides from the 5′ terminal of CGL1-1 (23). Protected regions in coding and noncoding strands are shown by A, B and C. Boundary sites of each protected region are arbitrary because DNaseI did not cleave all the bases equally in the absence of proteins.

as the CAT activity in the transient expression system. A promoter-less vector pCAT-Basic and a positive control vector pCAT-Control having the SV40 early promoter and enhancer elements were used as references. Because of the substantial fluctuation of the transfection efficiency in each transfection experiment, we normalized the experimental data of each construct by monitoring the luciferase activity of the co-transfected RSV-firefly luciferase fusion plasmid, pRSVLUC (26). As shown in Fig. 1, the pCGL1-1/−24/+287 showed the significant CAT activity in HeLa cells. The activity was approximately 30% of that expressed by the pCAT-Control. On the other hand, when the same L1 fragment was fused with the CAT gene with the opposite orientation, no CAT activity was detected (data not shown). It is unlikely that the sequence flanking to CGL1-1 serves as a promoter because the region of −5 to −24 fused with the CAT gene had no promoter activity. These results indicate that the 5′ terminal region of CGL1-1 has a promoter activity. Similar results were obtained when the 5′ terminal region of CGL1-3 (23) was used as a promoter (data not shown).

A series of deletion plasmids were constructed to localize the regions responsible for the promoter activity. The results are summarized in Fig. 1. The 3′ deletion from +287 to +155 of CGL1-1 (pCGL1-1/−24/+155) enhanced the promoter activity more than two-fold. The further deletion to +40 (pCGL1-1/−24/+40) retained about 80% of the activity of that of pCGL1-1/−24/+287, but the deletion to +14 (pCGL1-1/−24/+14) diminished the activity down to 5% or less, and the further deletion to −5 caused the complete loss of the activity. On the other hand, the 5′ deletions from −24 to +39 (pCGL1-1/+39/+287) significantly reduced the promoter activity and the deletion to +152 almost completely eliminated the activity. It should be noted that the deletion of the first 17bp of the L1 sequence (pCGL1-1/+18/+155) reduced the activity to 10% of that of pCGL1-1/−24/+155. These results suggested that the first 155bp of L1, particularly the first 40bp, contains the elements essential for the promoter activity.

The transcription of L1 should start at the first nucleotide of L1 to produce the full-length L1 RNA. We asked whether the L1 promoter studied above starts the transcription properly at the position of +1 of CGL1-1. The transcription initiation site of the transgene was determined by a primer extension experiment using a ^{32}P-labeled primer complementary to the CAT gene sequence (Fig. 2). Four major bands were observed, which were proven to be generated from the transcripts initiated at −3, −2, +1 and +2 of the L1-CAT construct. The band at +1 was clearly identified as the extension product with the strongest intensity. These results suggested that the 5′ terminal region of L1 has an ability to start the transcription at or around the first nucleotide (G) of L1. No band was seen in the experiment using RNA from untransfected cells under the same conditions.

Nuclear protein binding sites at the 5′ region of L1

In general, promoters and other regulatory elements are the binding sites for the nuclear proteins involved in the transcription regulation (31). We thus attempted to identify the nuclear factors which bind to the 5′ region of CGL1-1 by DNaseI footprint analysis. When the fragment containing the sequence from −102 to +287 of CGL1-1 was used as a probe, three regions were prominently protected by the nuclear proteins from HeLa cells (Fig. 3). Protection was observed in both coding and noncoding strands. These protected regions were designated as sites A (+3 to +26), B (+71 to +89) and C (+206 to +218), respectively. Similar results were obtained in the DNaseI footprint analysis using the sequence from CGL1-3 (data not shown). The nucleotide sequences of the protected regions are summarized in Fig. 4. These sequences are well conserved among CGL1s (23) and other CpG-rich L1s (16, 32, 33) (see below for discussion).

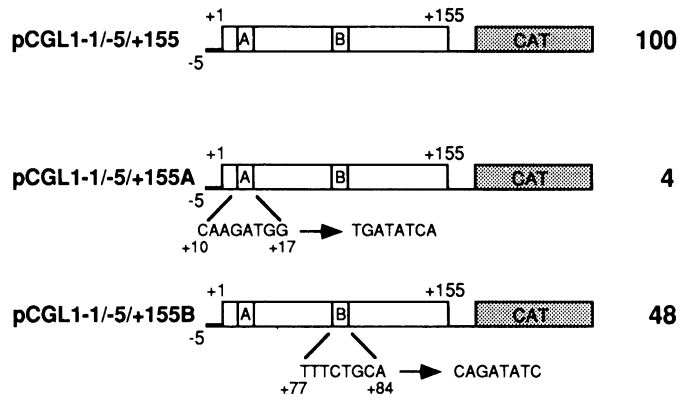

Fig. 5. CAT activity of the mutant constructs in HeLa cells. The experimental conditions are the same as those described in Fig. 1. The CAT activity for the pCGL1-1/−5/+155 was taken as 100% and the relative values were shown at the right for pCGL1-1/−5/+155A and pCGL1-1/−5/+155B. Wild and mutant sequences are shown in the protein binding sites A and B of each construct.

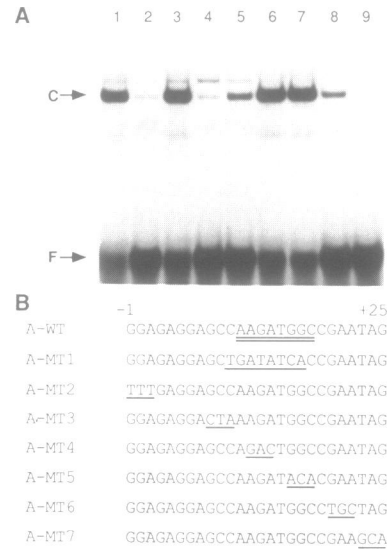

Fig. 6. Gel shift assay using ds-oligonucleotides containing the 5′ region sequence of CGL1-1. The experimental conditions are described in Materials and Methods. (A): Extracts of HeLa cells were incubated with the ^{32}P-labeled ds-oligonucleotide probe A-WT without competitor (lane 1) or with a competitor of A-WT (lane 2), A-MT1 (lane 3), A-MT2 (lane 4), A-MT3 (lane 5), A-MT4 (lane 6), A-MT5 (lane 7), A-MT6 (lane 8) or A-MT7 (lane 9). Three hundred fold excess of competitors was used in each competition assay. The position of specific DNA-protein complex (C) and free probes (F) are marked at the left side. (B): Coding strand sequences of ds-oligonucleotides used for probes and competitors are shown. The A-WT contains the sequence from −1 to +25 of CGL1-1. Substituted nucleotides are indicated by underlines in each competitor sequence. Double-underline in the A-WT sequence shows the sequences conserved between human and prosimian L1s (See text).

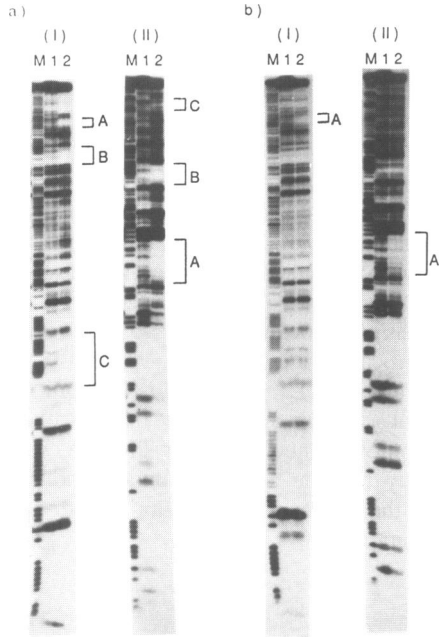

Fig. 7. DNaseI footprinting experiments using the nuclear extracts from different types of cell lines. The nuclear extracts from (a) JEG-3 (choriocarcinoma) and (b) ITO-2 (embryonic teratocarcinoma) were used. The experimental conditions and symbols in the Figure are the same as those for Fig. 3.

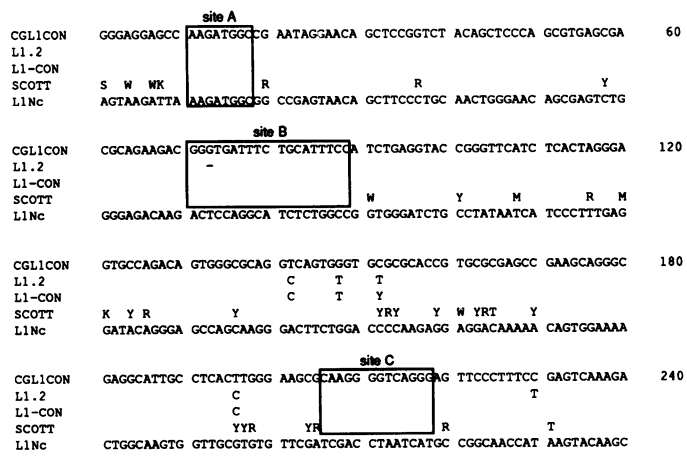

Fig. 8. Alignment and conservativeness of the 5′ terminal sequences of L1s. CGL1CON represents a consensus sequence of human CGL1s which were shown to be generated by recent transposition (23). L1.2 indicates a human L1 sequence which was proven to be active (16). L1-CON and SCOTT represent consensus sequences of selected human genomic L1s which are rich in G+C at the 5′ regions, respectively (32, 33). L1Nc shows a consensus sequence of prosimian L1s (11, 34). Only different nucleotides were indicated. The protein-binding sites A, B and C were parenthesized. Abbreviations for ambiguous bases are R=A/G; Y=T/C; S=C/G; K=G/T; W=A/T; M=A/C.

A protein binding site at the 5′ region is crucial for the expression of L1

As shown by the deletion analysis (Fig. 1), the first 155bp of L1 contains the elements essential for the L1 promoter activity. We asked whether the nuclear protein binding sites in the region are important for the expression of L1. Two mutant plasmids were constructed, pCGL1-1/−5/+155A in which a sequence from +10 to +17 at the center of the site A was replaced by an unrelated sequence TGATATCA and pCGL1-1/−5/+155B in which a sequence from +77 to +84 at the center of the site B was replaced by a sequence CAGATATC. The transfection experiments using these mutant plasmids showed that the mutation at the site A dramatically reduced the promoter activity, while the mutation of the site B retained a half of the activity (Fig. 5). From these results, we concluded that the site A is a crucial element for the promoter activity in HeLa cells.

Protein binding core element at the site A

The nuclear factor which binds to the site A was also detected by gel shift assay (Fig. 6). A 26bp of double-stranded (ds) oligonucleotide probe (A-WT) containing the sequence from −1 to +25 of CGL1-1 gave a band of DNA-protein complex in the presence of HeLa nuclear extracts (Fig. 6, lane 1). The complex formation was reduced in the presence of unlabeled A-WT as a competitor (Fig. 6, lane 2), but not in the presence of a ds-oligonucleotide (A-MT1) having an unrelated sequence at the center of the site A (Fig. 6, lane 3). To more precisely determine the core element for the protein binding, competition experiments were done using a series of mutant competitors. As shown in Fig. 6, the sequences mutated in A-MT4 and 5 are essential for the formation of the specific complex. The sequences mutated in A-MT3 and 6 are also involved in the complex formation to some extent. From these results, we concluded that among the protected bases in the site A, the ones from +8 to +22 are directly involved in the complex formation. Particularly, the sequence sandwiched between the mutated sequences of A-MT3 and A-MT6 seems essential, so that the sequence from +11 to +19 is designated the core element herein. We previously pointed out that some sequences at the 5′ end of L1s are evolutionarily conserved (34). It is interesting to note that the core element sequence is conserved in the 5′ terminal regions of prosimian L1s (indicated by double-underline in Fig. 6B)

Wide distribution of the nuclear protein binding to the site A

The data presented above indicated that the site A is essential for the L1 transcription, and it seems reasonable to consider that the nuclear protein binding to the site A plays an important role in L1 transcription. We designated the site A-binding factor as TFL1-A. Judging from its recognition sequence, TFL1-A seems to be a novel transcription factor. It is interesting to investigate the distribution of TFL1-A in various types of cells. We at first prepared the nuclear extracts from JEG-3 (a choriocarcinoma cell) and ITO-2 (an embryonic teratocarcinoma cell). JEG-3 cell is reported to synthesize the full-length L1 RNA (14) and also to produce the ORF1 protein (15). ITO-2 cell was established from a human embryonic teratocarcinoma (22) and is thought to be similar to NTera2D1 in which the full-length polyadenylated L1 transcripts are identified (13). The nuclear extracts from JEG-3 cells produced three prominent protection regions which were indistinguishable from those obtained by HeLa extracts (Fig. 7). In contrast, the nuclear extracts from ITO-2 cells protected only the site A. The significance of the lack of the site B- and the site C-binding proteins (designated TFL1-B and TFL1-C) remains to be elucidated. The distribution of TFL1-A was further analyzed by gel shift assay. The nuclear extracts obtained from F9 (a mouse teratocarcinoma cell), MEL (a mouse erythroleukemia cell) and RPMI 1788 (a human hematopoietic cell), gave the band of the site A oligonucleotide-protein complex (data not shown). Thus,

TFL1-A is present in a variety of cells and seems to have no cell-type specificity.

DISCUSSION

We examined the regulatory elements responsible for the transcription of the human L1 sequence. We mainly used a full-length L1 member CGL1-1 (23) for the experiments. CGL1-1 has a high number of CpG dinucleotide sequences at its 5′ region and transposition of CGL1-1 occurred in a relatively recent past (23), so that CGL1-1 is considered to retain the structural features of actively transcribed L1 sequences. HeLa cell was used for most of experiments. The protein encoded by ORF1 was shown to be produced in HeLa cells (15), so that HeLa has a potentiality for synthesizing some functional L1 mRNAs. Transfection experiments showed that the first 155bp sequence possesses the promoter activity and particularly the first 40bp are crucial. These results were consistent with those described by Swergold (25), who showed that the first 100bp of L1 sequence were essential for the full expression a L1-β-galactosidase fusion gene in NTera2D1 cells. It should be, however, noted that the promoter activity directed by the CGL1-1 sequence is comparable to that by SV40 promoter with the enhancer element in HeLa cells (Fig. 1). This is apparently inconsistent with the observation of Swergold (25) that the L1 transgenes were very poorly transcribed in this cell line. This discrepancy may be explained by the difference of the fusion gene constructs. Swergold fused the first 952bp of L1 with β-galactosidase reporter gene. On the other hand, we used the first 287bp of L1 as a promoter and CAT gene as a reporter. It is conceivable that the L1 sequence from +287 to +952 contains some elements for determining the cell specificity of L1 transcription. Despite of difference in quantity of L1 mRNA production between HeLa and NTera2D1 cells, the L1 transgenes were capable of being properly transcribed from +1 in both cell lines. This suggested that the transcription initiation may be regulated in the same manner in both cells and that the first 155bp sequence includes signals specifying the transcription initiated at +1. It has been reported that the full-length L1 transcripts are rarely detected in HeLa cells (13). This is also apparently inconsistent with our observation that L1 showed the high promoter activity. This might be explained by the difference of the physical/chemical status between the L1 transgenes and endogenous L1 sequences. For example, methylation of endogenous L1s (33) may repress their expression.

The DNaseI footprint analysis showed that three regions, designated as A, B and C sites, in the first 287bp were prominently protected by nuclear factors of HeLa cells (Figs. 3 and 4). The functional analysis of these protected regions by the transfection experiments indicated that the site A located at +3 to +26 is responsible for the promoter activity of L1 in HeLa cells. Furthermore, the gel shift assay indicated that an element AAGATGGCC at +11 to +19 is essential for the efficient formation of the DNA-protein complex *in vitro*. It seems reasonable to think that the binding of the factor to the sequence AAGATGGCC represents a crucial step for the transcription of the human L1 sequence. Since the site A-binding protein (TFL1-A) is present in various types of cells, TFL1-A seems to be involved in the basal transcription machinery for L1. The binding proteins for the sites B and C (TFL1-B and TFL1-C, respectively) were, on the other hand, found to be absent in ITO-2 cells. These proteins might be participated in a tissue-specific expression of the human L1. The role of TFL1-B and TFL1-C in L1 transcription remains to be an open question.

The binding sites A, B and C in the CGL1 sequence were found to be well conserved in other L1 sequences which seem to have recently transposed (Fig. 8). Interestingly, the octanucleotide AAGATGGC in the protein binding core element of the site A was perfectly conserved at the similar position of a prosimian L1 sequence (34). Prosimian is a primate evolutionarily most distant from man and about 1kb of the 5′ region has almost no sequence homology between human and prosimian L1s, except some short sequences in the first 33bp sequence (34). Evolutional conservativeness of the octanucleotide sequence may support its biological significance.

A unique feature of L1 promoter is its ability to start the transcription at the upstream site (the position of +1). Characterization of the nuclear factor TFL1-A, which binds to the site A, may provide a clue to understand the molecular mechanism for the upstream transcription initiation, the first step of L1 transposition.

ACKNOWLEDGMENTS

We thank Drs. J.Hata and Y.Misumi for kindly providing ITO-2 and JEG-3 cells, and Dr. H.Hohjoh for critical reading of the manuscript and for his helpful discussion. We also thank Miss. H.Ohgusu for preparation of synthetic oligonucleotides. This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. Fanning,T.G. and Singer,M.F. (1987) *Biochim. Biophys. Acta.*, **910**, 203–212.
2. DiNocera,P.P. and Sakaki,Y. (1990) *Trends Genet.*, **6**, 29–30.
3. Fawcett,D.H., Lister,C.K., Kellett,E. and Finnegan,D.J. (1986) *Cell*, **47**, 1007–1015.
4. DiNocera,P.P and Casari,G. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5843–5847.
5. Priimägi,A.F., Mizrokhi,L.J. and Ilyin,Y.V. (1988) *Gene*, **70**, 253–262.
6. Xiong,Y. and Eickbush,T.H. (1988) *Cell*, **55**, 235–246.
7. Besansky,N.J. (1990) *Mol. Cell. Biol.*, **10**, 863–871.
8. Kimmel,B.E., Ole-Moiyoi,O.K. and Young,J.R. (1987) *Mol. Cell. Biol.*, **7**, 1465–1475.
9. Aksoy,S., Williams,S., Chang,S. and Richards,F.F. (1990) *Nucleic Acids Res.*, **18**, 785–792.
10. Schwarz-Sommer,Z., Leclercq,L., GÖbel,E. and Saedler,H. (1987) *EMBO J.*, **6**, 3873–3880.
11. Hattori,M., Kuhara,S., Takenaka,O. and Sakaki,Y. (1986) *Nature*, **321**, 625–628.
12. Loeb,D.D., Padgett,R.W., Hardies,S.C., Shehee,W.R., Comer,M.B., Edgell,M.H. and Hutchison III,C.A. (1986) *Mol. Cell. Biol.*, **6**, 168–182.
13. Skowronski,J. and Singer,M.F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6050–6054.
14. Skowronski,J., Fanning,T.G. and Singer,M.F. (1988) *Mol. Cell. Biol.*, **8**, 1385–1397.
15. Leibold,D.M., Swergold,G.D., Singer,M.F., Thayer,R.E., Dombroski, B.A. and Fanning, T. G. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6990–6994.
16. Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. and Kazazian, Jr., H. H. (1991) *Science*, **254**, 1805–1808.
17. Mathias, S. L., Scott, A. F., Kazazian, Jr., H. H., Boeke, J. D. and Gabriel, A. (1991) *Science*, **254**, 1808–1810.
18. Mizrokhi,L.J., Georgieva,S.G. and Ilyin,Y.V. (1988) *Cell*, **54**, 685–691.
19. Nur,I.E. and Furano,A.V. (1988) *Nucleic Acids Res.*, **16**, 9233–9251.
20. Swergold,G.D. (1990) *Mol. Cell. Biol.*, **10**, 6718–6729.
21. Severynse, D. M., Hutchison III, C. A., and Edgell, M. H. (1992) *Mammalian Genome* **2**, 41–50.

22. Sekiya,S., Kawata,M., Iwasawa,H., Inaba,N., Sugita,M., Suzuki,N., Motoyama,T., Yamamoto,T. and Takamizawa,H. (1985) *Differentiation*, **29**, 259–267.
23. Hohjoh,H., Minakami,R. and Sakaki,Y. (1990) *Nucleic Acids Res.*, **18**, 4099–4104.
24. Hattori,M. and Sakaki,Y. (1986b) *Analytical Biochem.*, **152**, 232–238.
25. Wigler,M., Pellicer,A., Silverstein,S. and Axel,R. (1978) *Cell*, **14**, 725–731.
26. de Wet,J.R., Wood,K.V., Deluca,M., Helinski,D.R. and Subramini,S. (1987) *Mol. Cell. Biol.*, **7**, 725–737.
27. Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.
28. Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,A.E. (1979) *Biochemistry*, **18**, 5294–5299.
29. Shapiro,D.J., Sharp,P.A., Wahli,W.W. and Keller,M.J. (1988) *DNA*, **7**, 47–55.
30. Garner,M.M. and Revzin,A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.
31. Maniatis,T., Goodboum,S. and Fischer,J.A. (1987) *Science*, **236**, 1237–1244.
32. Scott,A.F., Schmeckpeper,B.J., Abdelrazik,M., Comey,C.T., O'Hara,B., Rossiter,J.P., Cooley,T., Heath,P., Smith,K.D. and Margolet,L. (1987) *Genomics*, **1**, 113–125.
33. Crowther,J. M., Doherty,J.P., Linsenmeyer,M.E., Williamson,M.R. and Woodcock,D.M. (1990) *Nucleic Acids Res.*, **19**, 2395–2401.
34. Sakaki,Y., Hattori,M., Fujita,A., Yoshioka,K., Kuhara,S. and Takenaka,O. (1986) *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 465–469.
35. Maxam,A.W. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.