Human Molecular Genetics, 1999, Vol. 8, No. 8 **1557–1560**

Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells

Michelle L. Kimberland, Vladimir Divoky[1], Josef Prchal[1], Uwe Schwahn[2], Wolfgang Berger[2] and Haig H. Kazazian Jr[+]

Department of Genetics, University of Pennsylvania, 475 CRB, 415 Curie Boulevard, Philadelphia, PA 19104-6145, USA, [1]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA and [2]Max-Planck-Institute for Molecular Genetics, Berlin (Dahlem), Germany

Received May 7, 1999; Revised and Accepted May 19, 1999

DDBJ/EMBL/GenBank accession nos AF148856 and AF149422

Functional L1 elements are autonomous retrotransposons that can insert into human genes and cause disease. To date, 10 of 12 known L1 retrotranspositions into human genes have been found to be 5'-truncated and incapable of further retrotransposition. Here we report the nucleotide sequences of the two full-length L1 elements, $L1_{\beta\text{-thal}}$ and $L1_{RP}$, that have inserted into the β-globin and retinitis pigmentosa-2 (RP2) genes, respectively. $L1_{\beta\text{-thal}}$ is 99.4% identical to a consensus sequence of active human L1s, while $L1_{RP}$ is 99.9% identical. Both elements retain impressive capacity for high frequency retrotransposition in cultured HeLa cells. Indeed, $L1_{RP}$ is the most active L1 isolated to date. Our data indicate that not all L1 insertions into human genes are 'dead on arrival'. Our findings also lend further credence to the concept of *cis* preference, that the proteins encoded by a particular L1 preferentially act upon their encoding RNA as opposed to other L1 RNAs.

INTRODUCTION

LINE-1 or L1 elements are autonomous non-LTR retrotransposons that can be transcribed into RNA, reverse-transcribed into cDNA and integrated into new genomic locations. Human L1 elements consist of a 5'-UTR with an internal promoter (1), two open reading frames (ORF1 and ORF2), and a 3'-UTR and a poly(A) tail (2). ORF1 encodes a nucleic acid binding protein (3,4) and ORF2 encodes a protein with endonuclease activity (5), reverse transcriptase activity (6) and a C-terminal cysteine-rich motif (7). There are ~100 000 L1 elements in the human genome (8) but only ~30–60 are thought to be capable of retrotransposition (9). Greater than 95% of all human L1s are 5'-truncated while others contain termination codons, frameshifts or other deleterious mutations that abolish activity. To date, 12 L1 retrotranspositions into genes have been reported as disease-causing mutations (10–18; E. Bakker and G. van Ommen, personal communication). Eleven of the 12 insertions belong to the Ta subset (19; C. Meischl, personal communication), a group of transcriptionally active L1s that contains most retrotransposition-competent elements. Of the 12 insertions, 10 were 5'-truncated and therefore incapable of subsequent retrotransposition. Here we demonstrate that the two full-length L1 insertions are highly active in a cell culture-based retrotransposition assay and are likely capable of secondary retrotransposition *in vivo*.

RESULTS AND DISCUSSION

$L1_{\beta\text{-thal}}$ was discovered in a mother and daughter of Ukrainian descent with hematological indices typical of β-thalassemia trait (15). The L1 was inserted into the 3'-end of intron 2 of the β-globin gene in the antisense orientation. It has two ORFs, a 107 bp poly(A) tail and is flanked by a 9–13 bp target site duplication. $L1_{\beta\text{-thal}}$ is 99.4% identical to a consensus sequence of active human L1s and its encoded proteins contain 11 amino acid differences from consensus (Fig. 1). $L1_{RP}$ was found in intron 1 of the retinitis pigmentosa-2 (RP2) gene of a patient with X-linked retinitis pigmentosa and his carrier mother (17). $L1_{RP}$ has two ORFs, an unusual poly(A) tail in which the A tract is interrupted by 11 nt (GTTTTAAATTT) and it is flanked by a 14 bp target site duplication. The $L1_{RP}$ sequence is 99.9% identical to the active L1 consensus sequence and its encoded proteins contain only three amino acid changes from the consensus (Fig 1).

To determine if the $L1_{\beta\text{-thal}}$ and $L1_{RP}$ insertions have retained retrotransposition capability, we tested them in a cultured cell assay shown to detect authentic retrotransposition events (20). For this assay, we cloned the elements into an episomal expression vector such that a *mneo1* indicator cassette was present in their 3'-UTRs in the antisense orientation. This cassette can be activated upon retrotransposition to confer G418 resistance to transfected cells (20). For each L1, we tested three different constructs (Table 1), one containing both the heterologous immediate early CMV promoter (pCMV) 5' to the L1 and an SV40 polyadenylation signal [SV40 poly(A)] 3' to the L1, the second containing only the SV40 poly(A) and the third containing neither pCMV nor the SV40 poly(A) signal (a native L1). As predicted by their high degree of similarity to the consensus sequence of active elements, both $L1_{\beta\text{-thal}}$ and $L1_{RP}$ retrotransposed at frequencies comparable with those reported for L1.3 (9), previously the most active element in this assay. Indeed, $L1_{RP}$ constructs retrotransposed at frequencies twice

[+]To whom correspondence should be addressed. Tel: +1 215 898 3582; Fax: +1 215 573 7760; Email: kazazian@mail.med.upenn.edu

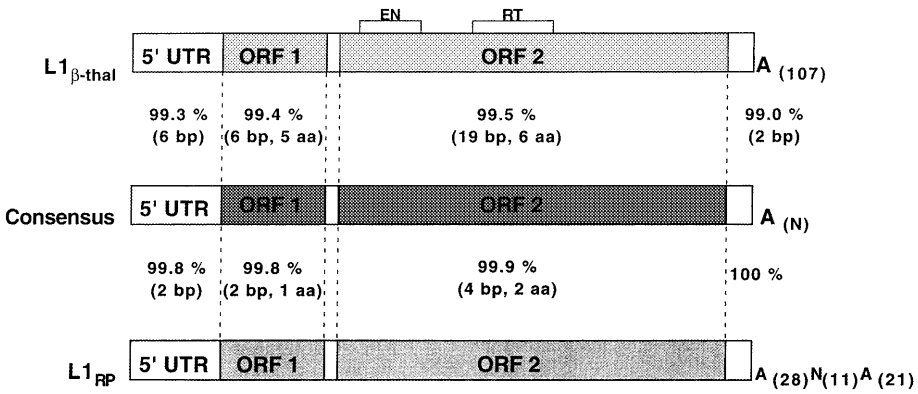

Figure 1. $L1_{\beta\text{-thal}}$ and $L1_{RP}$ are closely related to the consensus sequence of active L1 elements. This consensus sequence was generated by comparing the sequences of seven full-length elements (L1.2A, LRE2.1, L1.3, L1.4, L1.19, L1.20 and L1.39) capable of retrotransposition in a cultured cell assay (25). The approximate locations of the endonuclease and reverse transcriptase domains are indicated by EN and RT, and the full-length L1s are 6.0 kb.

Table 1. Retrotransposition frequencies of $L1_{\beta\text{-thal}}$ and $L1_{RP}$ are comparable with those of L1.3

Construct	*n*	Retrotransposition frequency/10^6 hygromycin^R cells (range)
pJM101/L1.3	7	4450 (2610–5690)
pJM101/L1.3 ΔCMV	9	16 140 (7480–30 430)
pJM101/L1.3 ΔCMV ΔSV40 poly(A)$^+$	9	8000 (3440–13 430)
pJM101/$L1_{\beta\text{-thal}}$	6	7770 (2810–14 300)
pJM101/$L1_{\beta\text{-thal}}$ ΔCMV	6	5310 (2400–9300)
pJM101/$L1_{\beta\text{-thal}}$ ΔCMV ΔSV40 poly(A)$^+$	6	2820 (1880–4400)
pJM101/$L1_{RP}$	6	8490 (6480–10 070)
pJM101/$L1_{RP}$ ΔCMV	6	30 750 (22 220–35 630)
pJM101/$L1_{RP}$ ΔCMV ΔSV40 poly(A)$^+$	6	17 910 (13 190–24 230)
pJM105	8	1.4 (0–4)
pJM102 ΔCMV	12	17 (0–58)

We calculated retrotransposition frequency as the number of G418^R foci per million hygromycin^R cells plated (20). *n* refers to the number of independent transfections. Retrotransposition frequencies for $L1_{\beta\text{-thal}}$ and $L1_{RP}$ were among the highest observed to date; the frequency for cells transfected with pJM101/$L1_{RP}$ ΔCMV reached one event in 30–35 hygromycin^R cells plated. A reverse transcriptase-defective allele of L1.2 (D702Y, pJM105) and an L1.2 construct lacking a presumptive promoter (no CMV promoter or L1 5'-UTR, pJM102 ΔCMV) generated G418^R foci at significantly lower frequencies (20). The G418^R foci generated by pJM105 are likely due to low level *trans*-complementation of the reverse transcriptase-defective allele by an endogenous L1 (20). The number of G418^R foci seen with pJM102 ΔCMV is greater than expected, but it should be noted that this construct contains an active L1. While the element lacks a promoter, transcripts arising from other promoters on the expression construct that fail to terminate prior to the L1 could generate a retrotransposition event.

those of L1.3, making $L1_{RP}$ the most active human L1 isolated to date. Deleting either pCMV or both pCMV and SV40 poly(A) increased the observed retrotransposition frequencies of L1.3 and $L1_{RP}$. Possible explanations for this increase include the loss of competition between neighboring promoters for polymerase and transcription factors or the relief of steric hindrance at polymerase binding sites. These results provide further evidence that L1 elements retrotranspose very efficiently in this assay (21).

Data from seven truncated human insertions have shown previously that newly inserted L1s likely arise from active, full-length progenitor elements (12,22). The protein coding regions in these previously characterized insertions were intact but truncated, suggesting that these elements might have retained activity had they not been truncated at some point during retrotransposition. Additionally, two newly retrotransposed full-length mouse L1 insertions retained retrotransposition activity in cultured cells and are likely capable of secondary retrotransposition in the mouse genome (23). $L1_{\beta\text{-thal}}$ and $L1_{RP}$ are the first recently retrotransposed, full-length human L1 insertions available for further characterization. That they have retained retrotransposition activity in this assay provides the first evidence that some disease-producing human L1 insertions are not inactivated upon retrotransposition. It should be noted that activity in the cultured cell assay does not prove that these elements will retrotranspose again from their new genomic locations. Other factors such as repressive effects of chromatin

structure and methylation may play a role in keeping retrotransposition in check (24).

It has been proposed that the protein products of a particular L1 bind primarily to their encoding RNA, i.e. *cis* preference (22). Although the vast majority of L1s are inactive, no characterized disease-producing L1 insertions contain nonsense or frameshift mutations. In addition, two full-length mouse L1 insertions and the precursors of two truncated human L1 insertions were retrotranspositionally active elements (20,23). Furthermore, an L1 mutated in either ORF1 or ORF2 is not efficiently complemented by co-transfection with an active L1 in the cultured cell assay (J.V. Moran and H.H. Kazazian, unpublished data). The observation that $L1_{\beta\text{-thal}}$ and $L1_{RP}$ are highly retrotransposition competent further strengthens the case for *cis* preference in L1 retrotransposition.

MATERIALS AND METHODS

Isolation and sequencing of $L1_{\beta\text{-thal}}$ and $L1_{RP}$

Using sequence derived from subcloned breakpoint fragments (15), we amplified the full-length L1 from the β-globin gene using the Expand Long Template PCR System (Boehringer Mannheim, Indianapolis, IN) and primers flanking the inserted L1 (GlobinL1-5', 5'-CCCAACCATAAAATAAAAGCAGAGAGG-3'; GlobinL1-3', 5'-GAATAACAGTGATAATTTCTGGGTTAAGGCAA-3'). PCR conditions were: one cycle of 94°C for 1 min; 10 cycles of 94°C for 10s, 63°C for 30s and 68°C for 4 min; 20 cycles of 94°C for 10 s, 63°C for 30s, 68°C for 4 min + 20 s auto-elongation/cycle; one cycle of 68°C for 7 min. Reactions contained 2.6 U of Expand Long Template enzyme mixture, 350 μM each dNTP and 300 μM each DNA primer in Expand PCR buffer 1. Because GlobinL1-5' contains 5 nt of 5' L1 sequence at the 3' end, this primer pair specifically amplified only the β-globin allele that contained the L1 insertion. We purified the 6 kb PCR product using the Wizard PCR Prep Purification System (Promega, Madison, WI) and cloned it into pPCR-script (Stratagene, La Jolla, CA) as per the manufacturer's protocol. We combined three clones in an equimolar ratio and sequenced them using ABI 377 or 373A stretch sequencers with Taq FS Dye Terminator chemistry to generate sequences for this element. We sequenced pPCR.L1Glo-62, the PCR clone used to generate the CEP4-based (Invitrogen, Carlsbad, CA) clones tested in the retrotransposition assay (20), separately to screen for PCR errors. Using this method, we determined that there were two errors, one in the 5'-UTR at nt 309 and a synonymous change in codon 651 of ORF2. Since no amino acids were mutated in pPCR.L1Glo-62, the errors were not corrected before additional cloning steps were performed.

We cloned $L1_{RP}$ in the same manner as $L1_{\beta\text{-thal}}$ using published flanking sequence and genomic DNA (17). The primers used for this amplification were: RP2L1-5', 5'-TGGAATGTGTATTAAGACTGTAAGGTGGG-3'; RP2L1-3', 5'-ACTGCCTGGTTTATGCCGCTTA-3'. PCR conditions were similar to those used to amplify $L1_{\beta\text{-thal}}$ but the annealing temperature was 68°C rather than 63°C. We generated the sequence for this element by sequencing four clones combined in an equimolar ratio. As with $L1_{\beta\text{-thal}}$, we sequenced the $L1_{RP}$ clone intended for use in further cloning steps, pPCR.L1RP-32, separately. There were seven errors in pPCR.L1RP-32, one in the 5'-UTR at nt 342, one in the inter-ORF region at nt 1952 and five in ORF2 (codons 192, 210, 236, 294 and 1168). Only one of the PCR errors in ORF2 resulted in an amino acid change (I1168T). Although residue 1168 is not thought to be within a functional domain, we corrected this error by replacing an *Spe*I–*Bst*Z17I fragment in pPCR.L1RP-32 with the same fragment from another $L1_{RP}$ clone with the correct sequence to create pPCR.L1RPS-5. We then used pPCR.L1RPS-5 to generate the pCEP4-based vectors for the retrotransposition assay. Genbank accession nos for $L1_{\beta\text{-thal}}$ and $L1_{RP}$ are AF149422 and AF148856, respectively.

Creation of constructs for the retrotransposition assay

The original retrotransposition construct, JM101 (20), consists of pCEP4 (Invitrogen), L1.2, and the *mneoI* cassette in the 3'-UTR of L1.2. Most L1 elements contain a *Bst*Z17I site in their 3'-UTR close to the 3' end of the element, while a *Not*I site is present in the cloning vector between the CMV promoter and the 5'-UTR of the L1. To create pJM101/ $L1_{\beta\text{-thal}}$ and its ΔCMV and ΔCMV ΔSV40 poly(A)$^+$ derivatives, we digested pJM101, pJM101ΔCMV and pJM101ΔCMV ΔSV40 poly(A)$^+$ with *Not*I and *Bst*Z17I to remove L1.2 (20). We then removed $L1_{\beta\text{-thal}}$ from pPCR.L1Glo-62 as a 6.0 kb *Not*I–*Bst*Z17I fragment and ligated it into the various pJM101 backbones. We generated pJM101/$L1_{RP}$ and its ΔCMV and ΔCMV ΔSV40 poly(A)$^+$ derivatives in the same manner using a 6.0 kb *Not*I–*Bst*Z17I fragment from pPCR.L1RPS-5. The retrotransposition assay was carried out as described previously (20).

ACKNOWLEDGEMENTS

We thank John Goodier, Eric Ostertag and Eline Luning Prak for helpful discussions, and the DNA Sequencing Core of the University of Pennsylvania School of Medicine for DNA sequencing. This work was supported by a grant from the NIH.

REFERENCES

1. Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.*, **10**, 6718–6729.
2. Kazazian, H.H.Jr and Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nature Genet.*, **19**, 19–24.
3. Hohjoh, H. and Singer, M.F. (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.*, **16**, 6034–6043.
4. Kolosha, V.O. and Martin, S.L. (1997) *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl Acad. Sci. USA*, **94**, 10155–10160.
5. Feng, Q., Moran, J.V., Kazazian, H.H. and Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
6. Mathias, S.L., Scott, A.F., Kazazian, H.H.Jr, Boeke, J.D. and Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.
7. Fanning, T. and Singer, M. (1987) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res.*, **15**, 2251–2260.
8. Hutchison, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) LINES and related retroposons: long interspersed sequences in the eucaryotic genome. In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA.* ASM Press, Washington, DC, pp. 593–617.
9. Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D. and Kazazian,

H.H.Jr (1997) Many human L1 elements are capable of retrotransposition. *Nature Genet.*, **16**, 37–43.
10. Kazazian, H.H.Jr, Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G. and Antonarakis, S.E. (1988) Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, **332**, 164–166.
11. Woods-Samuels, P., Wong, C., Mathias, S.L., Scott, A.F., Kazazian, H.Jr and Antonarakis, S.E. (1989) Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics*, **4**, 290–296.
12. Holmes, S.E., Dombrowski, B.A., Krebs, C.M., Boehm, C.D. and Kazazian, H.H.Jr (1994) A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nature Genet.*, **7**, 143–148.
13. Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H. and Matsuo, M. (1993) Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.*, **91**, 1862–1867.
14. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B. and Nakamura, Y. (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.*, **52**, 643–645.
15. Divoky, V., Indra, K., Mrug, M., Brabec, V., Huisman, T.H.J. and Prchal, J.T. (1996) A novel mechanism of β-thalassemia. The insertion of L1 retrotransposable element into β globin IVSII. *Blood*, **88**, 148a.
16. Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S. and Takeda, S. (1998) Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Hum. Mol. Genet.*, **7**, 1129–1132.
17. Schwahn, U., Lenzner, S., Dong, J., Feil, S., Hinzmann, B., van Duijnhoven, G., Kirschner, R., Hemberger, M., Bergen, A.A., Rosenberg, T., Pinckers, A.J., Fundele, R., Rosenthal, A., Cremers, F.P., Ropers, H.H. and Berger, W. (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nature Genet.*, **19**, 327–332.
18. Meischl, C., De Boer, M. and Roos, D. (1998) Chronic granulomatous disease caused by line-1 retrotransposons. *Eur. J. Haematol.*, **60**, 349–350.
19. Skowronski, J., Fanning, T.G. and Singer, M.F. (1988) Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.*, **8**, 1385–1397.
20. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
21. Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H.Jr (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530–1534.
22. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. and Kazazian, H.Jr (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.
23. Naas, T.P., DeBerardinis, R.J., Moran, J.V., Ostertag, E.M., Kingsmore, S.F., Seldin, M.F., Hayashizaki, Y., Martin, S.L. and Kazazian, H.H. (1998) An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.*, **17**, 590–597.
24. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
25. Sassaman, D.M. (1996) Characterization of five novel human L1 elements capable of retrotransposition. Doctoral dissertation, Johns Hopkins University, Baltimore, MD.