ARTICLE

A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease

Christof Meischl[1], Martin de Boer[1], Anders Åhlin[2] and Dirk Roos[1]

[1]CLB, Sanquin Blood Supply Foundation and Laboratory for Experimental and Clinical Immunology, Academic Medical Center, University of Amsterdam, The Netherlands; [2]Sachs' Children's Hospital, Stockholm, Sweden

Long interspersed nuclear element-1 (LINE-1) or L1 elements are DNA elements present in the genome in high copy number and capable of active retrotransposition. Here we present a patient with severe chronic granulomatous disease (CGD) caused by insertion of an L1 sequence into intron 5 of the X-lined gene *CYBB*. Due to internal rearrangements, the insert introduced new splice sites into the intron. This resulted in a highly heterogeneous splicing pattern with introduction of two L1 fragments as new exons into the transcripts and concomitant skipping of exonic coding sequence. Because no wild-type cDNA was found, this mechanism is probably responsible for the patient's phenotype. The L1 fragment, which belongs to the Ta subset of transcriptionally active LINEs, illustrates a new mechanism by which these elements can modify the transcribed coding sequence of genes. *European Journal of Human Genetics* (2000) 8, 697–703.

Introduction

Mobile elements, which represent about one third of the human genome,[1] have been recognised over recent years as a very important factor influencing the plasticity and evolution of the genetic material of most, if not all, organisms.[2] The largest group of these elements are the long interspersed nuclear element-1 (LINE-1) or L1 elements, which comprise 15–20% of the human genome. Although more than 100 000 L1s are estimated to exist, more than 95% of them are 5' truncated to a variable extent, and 10% are internally rearranged. Most of the remaining 3000–4000 full-length elements have been rendered inactive by various sorts of point mutations.

A LINE-1 consensus sequence has been described[3,4] that defines the intact, active, full-length L1 element as a sequence of 6 kb, consisting of a 5' untranslated region (UTR) with an internal promoter activity, two open reading frames (ORF) separated by a 66-bp intergenic spacer, and a 205-bp 3' UTR ending with a poly-A tail. Typically, two short target duplication sites of variable length frame an L1 element.

ORF1 encodes a 40-kD RNA-binding protein with high affinity for the 5' end of ORF2 of the L1 transcript, whilst ORF2 has been shown to encode a reverse transcriptase with concurrent endonuclease activity.

Based on these features, a model for retrotransposition of L1 elements has been proposed[2,5] supported by a growing body of evidence. According to this model, the L1 transcript, together with its two gene products, forms a ribonucleoprotein complex that is subsequently transported to the nucleus. There, by means of the endonuclease component of the ribonucleoprotein complex, a nick is introduced at an arbitrary location of the genome, and the resultant lagging 3' end is used for target-primed reverse transcription. Subsequent second-strand synthesis will then result in a copy of the original LINE at a novel genomic location.

So far, LINE-1 sequences have been implicated in 13 cases of human disease, in most instances due to insertion into the coding sequences of the affected genes.[4,6–17]

Here, we describe a patient with chronic granulomatous disease (CGD), caused by L1 insertion into an intronic sequence of the X-linked *CYBB* gene. CGD is a severe congenital immunodeficiency syndrome caused by defects in the NADPH oxidase in phagocytic leukocytes. This enzyme is responsible for the generation of the reactive oxygen species

Correspondence: Dirk Roos PhD, CLB, Dept. IHE, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands. Tel: + 31 20 5123349; Fax: + 31 20 5123474; E-mail: d — roos@clb.nl

used by phagocytes for the killing of various pathogens. Four different genetic forms of the disease have been recognised to arise from defects in four different subunits of the oxidase.[18] By far the most frequent form is X-linked, caused by defects in the gp91-phox subunit, encoded by *CYBB*.[18]

Materials and methods

Clinical history of patient HA

A boy was born of healthy parents in 1976 and presented in the newborn period with an *S. aureus* omphalitis. Search for an immunodeficiency was started at the age of $4\frac{1}{2}$ years, after a history of repeated infections. Absence of a chemiluminescence reaction and a negative nitroblue tetrazolium reduction (NBT) test of the neutrophils disclosed chronic granulomatous disease (CGD). An aberrant band in a Southern blot with probes directed against *CYBB* confirmed this diagnosis and indicated the X-linked form of the disease. Prophylactic antibiotic therapy reduced the incidence of infections in the years following diagnosis but, at the age of 22 years, the patient died due to complications of a fungal pneumonia.

Preparation of RNA and DNA

Total RNA of the patient and a healthy control was purified from mononuclear leukocytes as described[19] and cDNA was synthesised. Genomic DNA was isolated from circulating blood leukocytes of the CGD patient, his parents, his sister and a healthy control with the Puregene kit (Gentra Systems, Minneapolis, MN, USA) according to the manufacturer's instructions.

PCR

The 13 exons with their adjacent intronic sequences (exon-1 + promoter) were amplified with the appropriate primer combinations (see Table 1). Genomic DNA (50–500 ng) was amplified by means of the Rapid Cycler (Idaho Technology, Idaho Falls, ID, USA) with 50 cycles of 95°C for 5 s, 60°C for 30 s and 72°C for 15 s and slope S9. The reaction volume of 15 µl contained 2 U of Taq DNA polymerase (Promega, Madison, WI, USA), 2 U of TaqStart antibody (Clontech

Laboratories, Palo Alto, CA, USA), 50 ng of each primer, 3 nmole of each of the dNTP's (Promega) and reaction buffer (50 mм KCl, 1.5 mм MgCl₂, 0.1% Triton X-100, 10 mм Tris and 8% DMSO, pH 9 at 25°C). The PCR reaction took place in 10 µl glass capillaries (Idaho Technology).

The cDNA of the coding region of *CYBB* was amplified with PCR in three overlapping fragments as described.[20]

XL PCR

XL-PCR products, spanning exon 5, intron 5 and exon 6, of the patient and a healthy control were obtained by means of the Expand High Fidelity PCR System from Boehringer (Mannheim, Germany) according to the manufacturer's recommendations. Briefly, the region of interest was amplified by XL PCR on an Omnigene thermocycler (Hybaid Limited, Teddington, Middlesex, UK), with 50–500 ng of genomic DNA as template in a total reaction volume of 100 µl. With the AmpliWax beads (Perkin Elmer Applied Biosystems, Warrington, Cheshire, UK) for the hot start technique, the PCR was run with the following settings. Stage 1: denaturation 95°C 3 min, annealing 60°C 1 min, extension 68°C 3 min 45 s, 1 cycle. Stage 2: denaturation 95°C 30 s, annealing 60°C 1 min, extension 68°C 3 min 45 s (time increment 3 s), 40 cycles. Stage 3: extension 68°C 5 min, 1 cycle. For subsequent cloning of the PCR products 1 µl of Taq DNA polymerase (Promega) was added to the reaction mixture immediately after amplification, and the mixture was incubated for another 10 min at 70°C.

Cloning

The PCR products were purified with the GFX PCR DNA Gel Band Purification kit (Amersham Pharmacia Biotech, Uppsala, Sweden) and cloned into the pGEM-T vector (Promega) according to the manufacturer's recommendations.

Sequencing and sequence analysis

The purified templates were cycle-sequenced by means of the Big Dye Terminator Cycle Sequencing Ready Reaction kit

Table 1 Primer combinations used for the amplification of the coding sequence of *CYBB* (see text for details)

	Sense primer	*Antisense primer*
Exon 1	5′–GCT GGT TAG TTA AAA AGT TAT TTC ACT GTG–3′	5′–GAT AAC CCC AGA AGT CAG AG–3′
Exon 2	5′–CCA GTC TTG TGT GGA ATC TAC TGT GGA AAT GCG G–3′	5′–TGT GAC CAG CCA ATA TTG CAT G–3′
Exon 3	5′–CCT CAT GCT AAG AAC CTT GG–3′	5′–TTG ATG GCC TTT GAA AAT TAG AGG AAC TTA G–3′
Exon 4	5′–GTT AAC AAT TAC TAT TCC ATT CTT TCC CCC–3′	5′–CTA TGA ATA GAG GGA ACT CCC TGG TTC CAA G–3′
Exon 5	5′–GCT GTC CCA GAA ACC CAG CTT AC–3′	5′–TCT AAG TCA GAT AAT GAG AGG–3′
Exon 6	5′–GTG AAA ATA TCT ATT GTT CTA TAC ATA GGA CAC–3′	5′–GGA CAT GAA ATC CTT CAC TTC AG–3′
Exon 7	5′–TCC TAT TAC TAA ATG ATC TGG–3′	5′–TAC ATG TTT CTT AGA CAC AGG–3′
Exon 8	5′–CCT CTG AAT ATT TTG TTA TCT ATT ACC AC–3′	5′–TTA GAC ACT GAC CAC TAG T–3′
Exon 9	5′–CCA TAT GAC TAA AAA GGC AAG–3′	5′–CCT GAA CAC ATA TAC GTT GG–3′
Exon 10	5′–TGA AGA GCA AGA CAT CTC TGT AAC TA–3′	5′–ACT GCT CTA AGG CCC TCC GAT AAA T–3′
Exon 11	5′–GTG GAA TTC CAC ATG GTA ATG–3′	5′–GTC ACT ATG GAA GGA CCT GAG–3′
Exon 12	5′–CTG AAT TCA TGT CCT TTC CTG–3′	5′–GGC AGA TGC AAG CCT CAA AAA CAT C–3′
Exon 13	5′–TAC CTG CTT GTA GAC ATC TCA T–3′	5′–AGC ATT ATT TGA GCA TTG GCA GCA CAA CCC ACA–3′

(Perkin Elmer Applied Biosystems) and run on an ABI 377 XL Automated DNA Sequencer (Perkin Elmer Applied Biosystems). Sequence analysis was performed by means of the Sequence Analysis, Sequence Navigator and Auto Assembler software (all Perkin Elmer Applied Biosystems).

Northern blotting
Twenty μg of total RNA from mononuclear leukocytes were used for Northern blotting as described previously.[21]

Southern blotting
Southern blotting, with 8 μg genomic DNA, was performed as previously described.[21]

Results
Routine mutation analysis of *CYBB*, ie sequencing of all 13 exons and their adjacent intronic sequences, showed no abnormalities in these parts of the gene, and investigation of the promoter region up to nucleotide –290 also failed to indicate any difference with the wild-type sequence.

We then sequenced the whole coding sequence of *CYBB*, with cDNA as a template. Here, the region comprising exons 5 and 6 was unreadable due to what seemed to be multiple overlayed sequences. After cloning the PCR product into the pGEM-T vector and sequencing nine insert-containing clones, we found a very heterogenous splicing pattern, with complete absence of the wild-type splicing pattern. Seven clones containing the same PCR product, obtained from the cDNA of a healthy control, showed the expected normal splicing pattern.

The observed heterogeneity was due to insertion of a LINE-1 fragment, in two versions of different length, in most of the clones, and various combinations of exon skipping (Figure 1). The LINE-1 fragments were inserted between exons 5 and 6 and corresponded to nucleotides 5220–5334 and 5220–5372, respectively, of the consensus sequence of active LINE-1 elements.[4] The LINE-1 fragments both contained stop codons within the reading frames found in the sequenced clones, predicting heavily truncated translation products. In pattern 5, without LINE-1 sequences in the cDNA (see Figure 1), the skipping of exons 5 and 6 results in a frameshift and leads to a stop codon 50 nucleotides further downstream.

To elucidate the mechanisms responsible for the insertions in the coding sequence we analysed three cloned XL-PCR products from genomic DNA, spanning intron 5 of *CYBB*, and found a truncated and fragmented LINE-1 element of 836 bp (not including the poly-A tail) inserted between positions 1880 and 1881 of the intron (Figure 2).

The inserted sequence is framed by target duplication sites of 15 bp each and ends in a poly-A tail of about 70 bp. The first 31 bp of the element itself are inverted relative to the LINE-1-consensus sequence, whereas the following 11 nucleotides predicted by the consensus are missing. When

we correct for these rearrangements, the 836 nucleotides of the fragment still differ at five positions (0.6%) from the established consensus sequence, thereby introducing two amino-acid substitutions into the predicted protein consensus sequence. Furthermore, sequence analysis showed that the insertion is part of the transcriptionally active Ta subset of LINEs.[22] The Ta subset has been shown to provide nearly all the active L1 elements found so far.[6,23]

The rearrangements within the inserted element create a new branch site as well as a new variant 3' splice site[24] in close proximity of each other. This activates two cryptic variant 5' splice sites,[24] 115 bp and 153 bp downstream from the 3' splice site, the more 3' of which was created by one of the point mutations in the consensus sequence mentioned above. These splice-site combinations generate the two fragments we found inserted in the gp91-phox mRNA of patient HA.

To examine the effect of these rearrangements on mRNA stability, a Northern blot was performed that failed to detect any mRNA expression. Finally, a Southern blot confirmed the presence of an insertion in the gene *CYBB* of patient HA by demonstrating a band shifted from 7.2 kb to about 8.2 kb (Figure 3). This band, which represents the region between the 3' half of intron 3 and the 5' half of intron 6 of *CYBB*, was normal in the other family members (Figure 3).

Discussion
We have characterised an insertion of a truncated and internally rearranged L1 element as a cause of chronic granulomatous disease in a patient with the X-linked variant of this immunodeficiency syndrome. The molecular pathogenic mechanism described here differs from the previously reported cases of LINE-1 elements as determinants of phenotypic abnormalities.

Of the nine reported L1 insertions causative of human disease, seven are of exonic nature, thereby disrupting the gene structure, and only two are intronic.[11,14] In the case of the somatic insertion in intron 2 of the *myc* locus[11] no further investigation into the pathogenic mechanism has been described, but failure of RT PCR indicated a complete absence of the gene-specific transcript due to the insertion of a L1 element into intron 1 of the *RP2* gene.[14]

In mice, several exonic and intronic LINE-1 insertions have been shown to cause altered splicing patterns.[25–28] The murine *beige* phenotype is due to the intronic insertion of an L1 sequence in the *beige* gene, which results in the splicing of two L1 fragments of different length in between the neighboring exons.[29] This last case most closely resembles the mutational mechanism described in this article but differs in that no rearrangements of the inserted retrotransposon sequence, determining the aberrant splicing pattern, have been reported and in that no exonic sequence was lost during splicing.

A

```
   Wild type: CACTCTGCGATTCACACCATTGCACATCTATTTAATGTGGAATGGTGTGTGAATGCCCGAGTCAATAATTCTGATCCTTATTCA
Pattern 1(2x): CACTCTGCGATTCACACCATTGCACATCTATTTAATGTGGAATGGTGTGTGAATGCCCGAGTCAATAATTCTGATCCTTATTCA
Pattern 2(2x): CACTCTG-------------------------------------------------------------------------------
Pattern 3(2x): CACTCTG-------------------------------------------------------------------------------
Pattern 4(2x): CACTCTG-------------------------------------------------------------------------------
Pattern 5(1x): CACTCTG-------------------------------------------------------------------------------
               Exon 4 |◄─────────────────────────────────────────────────── Exon 5
```

```
   Wild type: GTAGCACTCTCTGAACTTGGAGACAGGCAAAATGAAAGTTATCTCAATTTTGCTCGAAAGAGAATAAAG--------------
Pattern 1(2x): GTAGCACTCTCTGAACTTGGAGACAGGCAAAATGAAAGTTATCTCAATTTTGCTCGAAAGAGAATAAAGACACTTCTCAAAAGA
Pattern 2(2x): -----------------------------------------------------------------------ACACTTCTCAAAAGA
Pattern 3(2x): -----------------------------------------------------------------------ACACTTCTCAAAAGA
Pattern 4(2x): -----------------------------------------------------------------------ACACTTCTCAAAAGA
Pattern 5(1x): ----------------------------------------------------------------------------------------
               Exon  5 ──────────────────────────────────────────────►|◄∼∼∼∼∼∼∼∼∼∼
```

```
   Wild type: ----------------------------------------------------------------------------------------
Pattern (2x)1: AGACATTTATGCAGCCAAAAAACACATGAAGAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAATCAAAACCACTATGAG
Pattern 2(2x): AGACATTTATGCAGCCAAAAAACACATGAAGAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAATCAAAACCACTATGAG
Pattern 3(2x): AGACATTTATGCAGCCAAAAAACACATGAAGAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAATCAAAACCACTATGAG
Pattern 4(2x): AGACATTTATGCAGCCAAAAAACACATGAAGAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAATCAAAACCACTATGAG
Pattern 5(1x): ----------------------------------------------------------------------------------------
               ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ LINE-1 Insert ∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼∼
```

```
   Wild type: ---------------|-----------------------------------------|AACCCTGAAGGAGGCCTGTACCTGGCTGTG
Pattern 1(2x): ATATCATCTCACACCGGTTAGAATGGCAATCATTAAGAAGTCAGGAAACAACAGAACCCTGAAGGAGGCCTGTACCTGGCTGTG
Pattern 2(2x): ATATCATCTCACACCGGTTAGAATGGCAATCATTAAGAAGTCAGGAAACAACAGAACCCTGAAGGAGGCCTGTACCTGGCTGTG
Pattern 3(2x): ATATCATCTCACACCG----------------------------------------AACCCTGAAGGAGGCCTGTACCTGGCTGTG
Pattern 4(2x): ATATCATCTCACACCG-----------------------------------------|-----------------------------
Pattern 5(1x): ---------------|-----------------------------------------|-----------------------------
               ∼∼∼∼∼∼∼∼∼∼►| ──────────────────────────────────────►|◄──────────
```

```
   Wild type: ACCCTGTTGGCAGGCATCACTGGAGTTGTCATCACGCTGTGCCTCATATTAATTATCACTTCCTCCACCAAAACCATCCGGAGG
Pattern 1(2x): ACCCTGTTGGCAGGCATCACTGGAGTTGTCATCACGCTGTGCCTCATATTAATTATCACTTCCTCCACCAAAACCATCCGGAGG
Pattern 2(2x): ACCCTGTTGGCAGGCATCACTGGAGTTGTCATCACGCTGTGCCTCATATTAATTATCACTTCCTCCACCAAAACCATCCGGAGG
Pattern 3(2x): ACCCTGTTGGCAGGCATCACTGGAGTTGTCATCACGCTGTGCCTCATATTAATTATCACTTCCTCCACCAAAACCATCCGGAGG
Pattern 4(2x): ----------------------------------------------------------------------------------------
Pattern 5(1x): ----------------------------------------------------------------------------------------
               ──────────────────────────────────── Exon 6 ─────────────────────────────────
```

```
   Wild type: TCTTACTTTGAAGTCTTTTGGTACACACATCATCTCTTTGTGATCTTCTTCATTGGCCTTGCCATCCATGGAGCTGAACGAAT
Pattern 1(2x): TCTTACTTTGAAGTCTTTTGGTACACACATCATCTCTTTGTGATCTTCTTCATTGGCCTTGCCATCCATGGAGCTGAACGAAT
Pattern 2(2x): TCTTACTTTGAAGTCTTTTGGTACACACATCATCTCTTTGTGATCTTCTTCATTGGCCTTGCCATCCATGGAGCTGAACGAAT
Pattern 3(2x): TCTTACTTTGAAGTCTTTTGGTACACACATCATCTCTTTGTGATCTTCTTCATTGGCCTTGCCATCCATGGAGCTGAACGAAT
Pattern 4(2x): -----------------------------------------------------------------------------ACGAAT
Pattern 5(1x): -----------------------------------------------------------------------------ACGAAT
               ──────────────────────────────────►| Exon 7
```
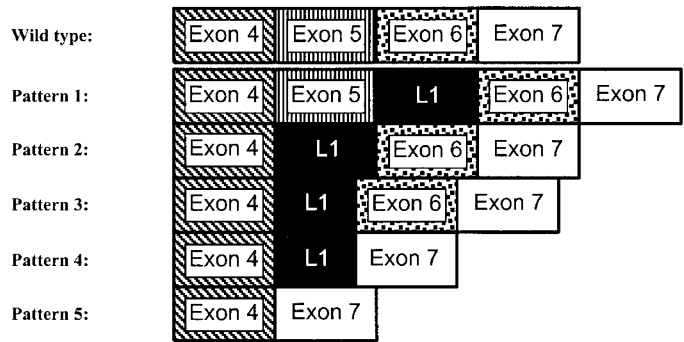
B

Figure 1 mRNA splicing pattern found in patient HA caused by the insertion of variably spliced LINE-1 sequences (solid and dashed arrows). The numbers in parentheses indicate the number of cDNA clones in which the corresponding sequence was found. The wild-type sequence is included here for comparison only but was not found in the patient. **A** Sequence pattern; **B** Graphical illustration of A.

The L1 insertion described here leads to insertion of retrotransposon sequences into the gene-specific transcript and concomitant skipping of upstream and/or downstream exons. The rearrangements of the LINE-1 sequence, which

A

TACCCCCTTA<u>GAAATAACTATTTAA</u>CCCACT**TTTTTGAT**GGGGTTGTTTG*TTTTTTT* *AACAGA*CACTTCTCAAAAGA

AGACATTTATGCAGCCAAAAAACACATGAAGAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAATCAAAACCA

CTATGAGATATCATCTCACA**CCgGTTAGA**AATGGCAATCATTAAg AAGTCAGGAAACAA**CAGGTGCTG**GAGAGGATGC

GGAGAAATAGGAACACTTTTACACTGTTGGTGGGACTGTAAACTAGTTCAACCATTGTGGAAGTCAGTGTGGCGATT

CCTCAGGGATCTAGAACTAGAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAAATGAGTATAAAT

CATGCTGCTATAAAGACACATGCACACGTATGTTTATTGCGGCACTATTCACAATAGCAAAGACTTGGAACCAACCC

AAATGTCCAACAATGATAGACTGGAT t AAGAAAATGTGGCACATATACACCATGGAATACTATGCAGCCATAAAAAA

TGATGAGTTCATATCCTTTGTAGGGACATGGATGAAATTGGAAACCATCATTCTCAGTAAACTAT c GCAAGAACAAA

AAACCAAACACCGCATATTCTCACTCATAGGTGGGAATTGAACAATGAGATCACATGGACACAGGAAGGGGAATATC

ACACTCTGGGGACTGTGGTGGGGTCGGGGGAGGGGGGAGGGATAGCATTGGGAGATATACCTAATGCTAGATGACAC

ATTAGTGGGTGCAGCGCACa AGCATGGCACATGTATACATATGTAACTAACCTGCACAATGTGCACATGTACCCTAA

AACTTAGAGTATAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AAAAAAAAG<u>AAATAACTATTTAA</u>GTACATCAGT

B
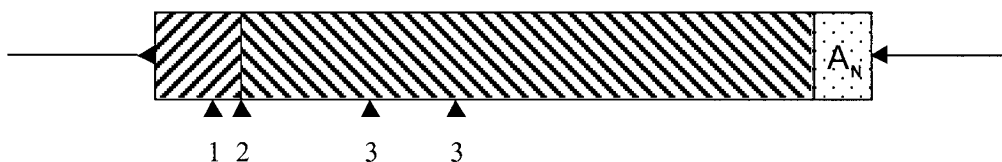
Figure 2 LINE-1 fragment inserted in intron 5 of CYBB of patient HA. **A** Detailed view of the sequence, the first and last 10 nucleotides being wild-type intron sequence. Underlined nucleotides, target duplication sites; framed sequence, sequence inverted with respect to the L1 consensus sequence; nucleotides underlaid in black, branch site; nucleotides printed in boldface italic, 3' splice site; nucleotides printed bold, underlined, 5' splice sites. Lower case letters indicate nucleotides divergent from the consensus sequence of active L1 elements.[4] **B** Schematic diagram of A –––, intron 5 of CYBB; ◄, target duplication sites; ▨, sequence inverted with respect to the L1 consensus sequence; ▨, L1 sequence; ▨, poly-A tail; **1**, branch site; **2**, 3' splice site; **3**, 5' splice sites.

most probably occurred during the process of retrotransposition, have created a new 3' splice site as well as a new branch site 11 bp further upstream (see Figure 2). Both of these sites, as well as the two newly activated cryptic 5' splice sites 115 bp and 153 bp downstream, are only weakly homologous with the respective consensus sites.[24] The homology is probably too weak to unambiguously define the exon/intron boundaries of the new 'exon', which would explain the observed highly heterogenous splicing pattern. Why, in eight out of nine cases (see Figure 1), this results in retention of the weakly defined new exon and skipping of the adjacent, well defined ones remains to be explained. The variant splicing patterns strongly reduce the stability of the transcript of *CYBB*, since Northern blotting proved unable to detect any mRNA.

Southern blotting indicated that the insertion must have occurred either in the mother's germline or during the
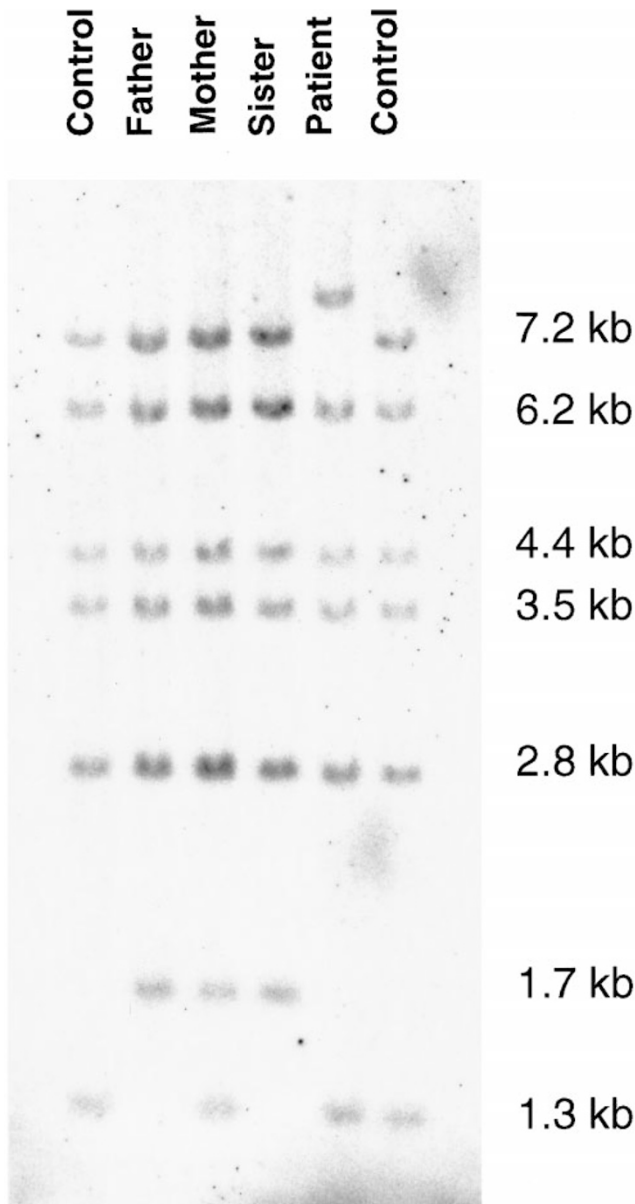
Figure 3 Southern blot of *Nsi I*-digested genomic DNA of two controls, the patient and the patient's father, mother and sister, showing an abnormal band at about 8.2 kb in the patient and a normal restriction pattern in the rest of the family. The bands at 1.3 kb and 1.7 kb correspond to a polymorphism at this locus.

patient's early embryogenesis, since his mother and sister showed a normal restriction pattern (see Figure 3).

The LINE-1 element itself largely confirms the evolving line of evidence concerning the active subgroup of these retrotransposons. So far, nine out of ten L1 insertions in humans have been found to belong to the transcriptionally-active Ta subset of LINEs (HH Kazazian, personal communication, 1999), and the present element also contains all of the four nucleotides that define this subset.[22] Furthermore,

while the new L1 insertion described here differs in two residues from the consensus amino-acid sequence for ORF2 of active L1s (HH Kazazian, personal communication, 1999), comparison with the nucleotide sequences from the L1 elements used for the construction of this consensus sequence shows that the responsible nucleic-acid substitutions can be found in some of the other active elements as well. This is relevant in two respects. First, it confirms and adds to the recently established consensus sequence. Second, it shows that the present element possibly also originates from an active predecessor – since all amino-acid substitutions have been found in other active elements as well – and this would further support the idea that the proteins encoded by LINE-1s preferably act in *cis*, by binding to their own mRNA.[4,6,30]

Recently the sequence specificity of the L1 endonuclease has been described,[31] characterising the sequence motif 5'-TTTT-NN-AAAA-3' as optimal, with nicking to take place between the two Ns. The corresponding sequence in the present case, 5'-TTTA-AG-TACA-3', although suboptimal, is in agreement with this description. Taken together, our results, while confirming the present-day model of retrotransposition, describe a novel way by which LINE-1 elements are able to act on the functioning of the genome.

References

1 Smit AF: The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 1996; **6**: 743–748.
2 Kazazian HH Jr, Moran JV: The impact of L1 retrotransposons on the human genome. *Nat Genet* 1998; **19**: 19–24.
3 Scott AF, Schmeckpeper BJ, Abdelrazik M *et al*: Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* 1987; **1**: 113–125.
4 Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH Jr: Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet* 1999; **8**: 1557–1560.
5 Luan DD, Korman MH, Jakubczak JL, Eickbush TH: Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993; **72**: 595–605.
6 Kazazian HH Jr: Mobile elements and disease. *Curr Opin Genet Dev* 1998; **8**: 343–350.
7 Miki Y: Retrotransposal integration of mobile genetic elements in human diseases. *J Hum Genet* 1998; **43**: 77–84.
8 Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE: Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 1988; **322**: 164–166.
9 Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet* 1994; **7**: 143–148.

10 Narita N, Nishio H, Kitoh Y *et al*: 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* 1993; **91**: 1862–1867.

11 Morse B, Rothberg PG, South VJ, Spandorfer JM, Astrin SM: Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 1988; **333**: 87–90.

12 Miki Y, Nishisho I, Horii A *et al*: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* 1992; **52**: 643–645.

13 Divoky V, Indrak K, Mrug M, Brabec V, Huisman THJ, Prchal J: A novel mechanism of beta thalassemia: the insertion of L1 retrotransposable element into beta globin IVS II. *Blood* 1996; **88**: 148a.

14 Schwahn U, Lenzner S, Dong J *et al*: Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet* 1998; **19**: 327–332.

15 Segal Y, Peissel B, Renieri A *et al*: LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. *Am J Hum Genet* 1999; **64**: 62–69.

16 Van de Water N, Williams R, Ockelford P, Browett P: A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. *Thromb Haemost* 1998; **79**: 938–942.

17 Kumatori A, Faizunnessa NN, Suzuki S, Moriuchi T, Kurozumi H, Nakamura M: Nonhomologous recombination between the cytochrome b558 heavy chain gene (CYBB) and LINE-1 causes an X-linked chronic granulomatous disease. *Genomics* 1998; **53**: 123–128.

18 Roos D, de Boer M, Kuribayashi F *et al*: Mutations in the X-linked and autosomal recessive forms of chronic granulomatous disease. *Blood* 1996; **87**: 1663–1681.

19 Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 1979; **18**: 5294–5299.

20 Bolscher BG, de Boer M, de Klein A, Weening RS, Roos D: Point mutations in the beta-subunit of cytochrome b558 leading to X-linked chronic granulomatous disease. *Blood* 1991; **77**: 2482–2487.

21 Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Press: NY, 1989.

22 Skowronski J, Singer MF: The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. *Cold Spring Harbour Symposium on Quantitative Biology* 1986; **51** Pt 1: 457–464.

23 Sassaman DM, Dombroski BA, Moran JV *et al*: Many human L1 elements are capable of retrotransposition. *Nat Genet* 1997; **16**: 37–43.

24 Jackson IJ: A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 1991; **19**: 3795–3798.

25 Kohrman DC, Harris JB, Meisler MH: Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron. *J Biol Chem* 1996; **271**: 17576–17581.

26 Takahara T, Ohsumi T, Kuromitsu J *et al*: Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. *Hum Mol Genet* 1996; **5**: 989–993.

27 Kingsmore SF, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF: Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nat Genet* 1994; **7**: 136–141.

28 Mulhardt C, Fischer M, Gass P *et al*: The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. *Neuron* 1994; **13**: 1003–1015.

29 Perou CM, Pryor RJ, Naas TP, Kaplan J: The bg allele mutation is due to a LINE1 element retrotransposition. *Genomics* 1997; **42**: 366–368.

30 Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr: Isolation of an active human transposable element. *Science* 1991; **254**: 1805–1808.

31 Cost GJ, Boeke JD: Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* 1998; **37**: 18081–18093.