Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia

MICHAEL J. THIRMAN*, DENISE A. LEVITAN*, HIRO KOBAYASHI*, M. CELESTE SIMON*†‡, AND JANET D. ROWLEY*†

Departments of *Medicine, Section of Hematology/Oncology, and †Molecular Genetics and Cell Biology, ‡The Howard Hughes Medical Institute, University of Chicago Medical Center, Chicago, IL 60637

Contributed by Janet D. Rowley, August 24, 1994

ABSTRACT To characterize the functions of MLL fusion transcripts, we cloned the gene that fuses to MLL in the translocation t(11;19)(q23;p13.1). This translocation is distinct from another type of 11;19 translocation with a 19p13.3 breakpoint that results in the fusion of MLL to the ENL gene. By PCR screening of a cDNA library prepared from a patient’s leukemia cells with this translocation, we obtained a fusion transcript containing exon 7 of MLL and sequence of an unknown gene. The sequence of this gene was amplified and used as a probe to screen a fetal brain cDNA library. On Northern blot analysis, this cDNA detected a 4.4-kb transcript that was abundant in peripheral blood leukocytes, skeletal muscle, placenta, and testis and expressed at lower levels in spleen, thymus, heart, brain, lung, kidney, liver, and ovary. In addition, a 2.8-kb transcript was present in peripheral blood, testis, and placenta. On “zoo blots,” this gene was shown to be conserved in 10 mammalian species as well as in chicken, frog, and fish. We have named this gene ELL (for eleven–nineteen lysine-rich leukemia gene). A highly basic, lysine-rich motif of the predicted ELL protein is homologous to similar regions of several proteins, including the DNA-binding domain of poly(ADP-ribose) polymerase. The characterization of the normal functions of ELL as well as its altered function when fused to MLL will be critical to further our understanding of the mechanisms of leukemogenesis.

Aberrations in chromosome 11 band q23 are frequent cytogenetic abnormalities in hematologic malignancies, especially in acute leukemias. Previously, we identified a yeast artificial chromosome that contained the breakpoint region in leukemias with several common 11q23 translocations (1). Subsequently, we cloned a gene named MLL (for myeloid–lymphoid leukemia or mixed-lineage leukemia) that spans this breakpoint (2). Other groups have also cloned this gene and have called it Hrx, ALL-1, and HRX (3–5). A 0.74-kb BamHI fragment isolated from MLL detects gene rearrangements on Southern blot analysis in all patients with the five most common types of 11q23 translocations as well as a large number of other, less common 11q23 abnormalities (6). All breakpoints in MLL occur within an 8.3-kb breakpoint cluster region. The MLL gene encodes an A-T hook domain at the N terminus and several regions of homology to the Drosophila trithorax gene, including a series of zinc fingers in the region immediately 3' to the breakpoint region and 210 C-terminal amino acids that do not contain a known motif. Mapping of functional domains of MLL with yeast GAL4 hybrids in transient transfections has revealed that MLL contains a strong activation domain 3' of the breakpoint and a repression domain 5' of the breakpoint region (7).

Six genes at 11q23 partner chromosomal breakpoints have been cloned. These include AF4 in t(4;11)(q21;q23) (4, 8), ENL in t(11;19)(q23;p13.3) (5), AF7 in t(9;11)(p22;q23) (9), AF6 in t(6;11)(q27;q23) (10), AFF1p in t(1;11)(p32;q23) (11), and AFX in t(X;11)(q13;q23) (12). The functions of these fusion partner genes have not been determined. Translocations at 11q23 result in the formation of two derivative chromosomes that encode chimeric transcripts. The der(11) transcript contains 5' MLL sequences fused to 3' sequences of the gene located on the partner chromosome, whereas the other derivative chromosome contains the 5' sequence of the partner gene potentially fused to 3' sequence of MLL. However, in 25% of patients, translocations are associated with deletions of MLL sequence that is 3' to the breakpoint (13). Thus, in these cases, a fusion transcript from the other derivative chromosome cannot be formed. In addition, analysis of complex 11q23 translocations reveals that the der(11) junction is always conserved (14). These data indicate that the fusion transcript encoded by the der(11) must be critical to leukemogenesis.

Although earlier cytogenetic studies did not discriminate between different breakpoints within band 19p13 in cases with a t(11;19)(q23;p13), it has become clear with more precise cytogenetic techniques that two breakpoints exist within band 19p13 (Fig. 1). These two breakpoints are easily distinguishable by fluorescence in situ hybridization (FISH) (15). The t(11;19)(q23;p13.3) breakpoint is common in acute lymphoblastic leukemia, primarily in infants and children. The translocation t(11;19)(q23;p13.1) is a recurring abnormality in both de novo and therapy-related acute myeloid leukemia and it has been observed in infants, children, and adults. The gene in t(11;19)(q23;p13.3) has already been cloned and named ENL (for eleven–nineteen leukemia).

To characterize the functions of MLL fusion transcripts further, we have cloned the gene that fuses to MLL in t(11;19)(q23;p13.1). The open reading frame of this gene codes for a hydrophilic, highly basic protein. We have named this gene ELL (for eleven–nineteen lysine-rich leukemia gene). The ELL gene has a broad tissue distribution and is conserved in all mammalian species as well as in chicken, fish, and frog.

MATERIALS AND METHODS

Cytogenetic Analysis and FISH. Cytogenetic analysis was performed with a trypsin–Giemsa banding technique. Chromosomal abnormalities are described according to the International System for Human Cytogenetic Nomenclature. FISH was performed with biotin-labeled probes (16).

Construction and Screening of cDNA Libraries. Poly(A)+ RNA was isolated with the FastTrack kit (Invitrogen) from cryopreserved bone marrow from a 41-year-old woman with

Abbreviations: FISH, fluorescence in situ hybridization; 5' RACE, rapid amplification of cDNA 5' ends.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
acute myeloid leukemia (FAB-M4) with a karyotype of 46,XX,t(11;19)(q23;p13.1)(100%). First-strand cDNA synthesis was catalyzed with Moloney murine leukemia virus reverse transcriptase using random hexamers (Pharmacia). The second strand was synthesized with RNase H and DNA polymerase I (Pharmacia). EcoRI/Not I adapters were blunt-end ligated to the cDNA with T4 DNA ligase (T4Saver cDNA library kit; Pharmacia) and then ligated into the EcoRI site of the λZAP II vector (Stratagene). The library was initially screened with the 0.74-kb BamHI cDNA fragment of MLL. Subsequently, the library was screened by PCR with nested primers from MLL exon 7 and the vector. PCR products were ligated into the pT7-Blue T vector (Novagen). A commercially prepared human fetal brain cDNA library (Stratagene) was screened with probes amplified from the MLL/ELL fusion transcript.

**Nucleotide Sequencing.** cDNA clones were sequenced entirely on both strands with Sequenase (United States Biochemical) using doubled-stranded templates. The DNA sequence and the predicted open reading frame were compared with GenBank data bases by using the BLAST and TFASTA programs.

**Southern and Northern Blot Analysis.** DNA was extracted and digested by standard techniques. RNA was isolated from cell lines with TriReagent (Molecular Research Center, Cincinnati, OH). The multiple-tissue Northern blots (Clontech), the somatic cell hybrid panel (Oncor), and the "zoo blots" (Bios, New Haven, CT) were used according to the manufacturers' suggestions.

**Rapid Amplification of cDNA 5' Ends (5' RACE).** First-strand cDNA was synthesized from 2 μg of placental poly(A)+ RNA with avian myeloblastosis virus reverse transcriptase (Clontech) and an antisense primer 150 bp downstream of the beginning of the 2.8-kb clone isolated from the fetal brain cDNA library. An anchor oligonucleotide was ligated to the 3' end of the cDNA by T4 RNA ligase (Clontech). After 30 cycles of amplification using the anchor primer (Clontech) and a primer 130 bp from the beginning of the 2.8-kb cDNA clone, a second round of 30 cycles of PCR was performed with a 17-mer oligonucleotide from the internal portion of the anchor primer and two different internal primers from the 2.8-kb clone (110 and 90 bp from the beginning of the clone).

**RESULTS**

Cloning of t(11;19)(q23;p13.1) Fusion Transcript. The 0.74-kb BamHI cDNA fragment from the MLL gene detected two rearranged bands in the BamHI digest of the DNA from the patient with the translocation t(11;19)(q23;p13.1) (Fig. 2A). The cDNA library prepared from the patient's leukemia cells was initially screened with the 0.74-kb BamHI cDNA fragment of MLL used to detect gene rearrangements. Three clones were isolated from the library but none contained a fusion transcript. To facilitate the cloning of this fusion transcript, we devised a PCR screening method using nested MLL-specific and vector primers. Detailed mapping of the genomic breakpoint within MLL in this patient's leukemia revealed a break between exons 7 and 9. Thus the MLL-specific primers were chosen approximately 50 and 30 bp upstream of the end of exon 7. After two rounds of amplification, several PCR products were obtained. Sequence analysis revealed normal MLL transcripts as well as a clone with a 270-bp insert that contained 30 bp of exon 7 and 240 bp that diverged from MLL sequences. The 240 bp that diverged from MLL were amplified by PCR and subsequently used as a probe to screen a fetal brain cDNA library. Three clones were isolated and sequenced. The largest of these clones (2.8 kb) contained the entire 240 bp found in the fusion transcript. The 2.2-kb clone contained a similar sequence (16 of 17 bp) to part of the 240-bp probe but was otherwise completely different from the rest of the probe or from any sequences in the 2.8-kb fragment. The 1.9-kb clone contained 46 bp that were present in the 2.8-kb clone and in the 240-bp fragment but diverged both 5' and 3' to this sequence. Preliminary sequence analysis revealed Alu sequences within this clone as well as consensus splice acceptor and splice donor sites flanking this 48-bp sequence, suggesting that this 1.9-kb clone represents a partially processed transcript or a genomic clone. The 1.9-kb and 2.2-kb clones were not characterized further.

**5' RACE.** An additional 23 nt that are upstream of the beginning of the 2.8-kb cDNA fragment were obtained with the 5' RACE strategy. These sequences contained a potential start codon flanked by an adenine in the −3 position and a guanine in the +4 position, thus following Kozak's consensus sequence for an initiation codon (17). Twelve base pairs of 5' untranslated sequence was obtained from the 5' RACE, but we cannot rule out the possibility that the 5' untranslated region may be longer.

**Chromosomal Localization.** The 240-bp fragment that diverged from MLL in the fusion transcript was found to hybridize to chromosome 19 in a somatic cell hybrid panel (data not shown). The entire 2.8-kb clone was labeled for FISH and it hybridized exclusively to chromosome 19, band p13.1 (Fig. 3).

**Northern Blot Analysis.** Multiple-tissue Northern blots were hybridized with a probe containing the first 500 bp of the

**RESULTS**

Cloning of t(11;19)(q23;p13.1) Fusion Transcript. The 0.74-kb BamHI cDNA fragment from the MLL gene detected two rearranged bands in the BamHI digest of the DNA from the patient with the translocation t(11;19)(q23;p13.1) (Fig. 2A). The cDNA library prepared from the patient's leukemia cells was initially screened with the 0.74-kb BamHI cDNA fragment of MLL used to detect gene rearrangements. Three clones were isolated from the library but none contained a fusion transcript. To facilitate the cloning of this fusion transcript, we devised a PCR screening method using nested MLL-specific and vector primers. Detailed mapping of the genomic breakpoint within MLL in this patient's leukemia revealed a break between exons 7 and 9. Thus the MLL-specific primers were chosen approximately 50 and 30 bp upstream of the end of exon 7. After two rounds of amplification, several PCR products were obtained. Sequence analysis revealed normal MLL transcripts as well as a clone with a 270-bp insert that contained 30 bp of exon 7 and 240 bp that diverged from MLL sequences. The 240 bp that diverged from MLL were amplified by PCR and subsequently used as a probe to screen a fetal brain cDNA library. Three clones were isolated and sequenced. The largest of these clones (2.8 kb) contained the entire 240 bp found in the fusion transcript. The 2.2-kb clone contained a similar sequence (16 of 17 bp) to part of the 240-bp probe but was otherwise completely different from the rest of the probe or from any sequences in the 2.8-kb fragment. The 1.9-kb clone contained 46 bp that were present in the 2.8-kb clone and in the 240-bp fragment but diverged both 5' and 3' to this sequence. Preliminary sequence analysis revealed Alu sequences within this clone as well as consensus splice acceptor and splice donor sites flanking this 48-bp sequence, suggesting that this 1.9-kb clone represents a partially processed transcript or a genomic clone. The 1.9-kb and 2.2-kb clones were not characterized further.

**5' RACE.** An additional 23 nt that are upstream of the beginning of the 2.8-kb cDNA fragment were obtained with the 5' RACE strategy. These sequences contained a potential start codon flanked by an adenine in the −3 position and a guanine in the +4 position, thus following Kozak's consensus sequence for an initiation codon (17). Twelve base pairs of 5' untranslated sequence was obtained from the 5' RACE, but we cannot rule out the possibility that the 5' untranslated region may be longer.

**Chromosomal Localization.** The 240-bp fragment that diverged from MLL in the fusion transcript was found to hybridize to chromosome 19 in a somatic cell hybrid panel (data not shown). The entire 2.8-kb clone was labeled for FISH and it hybridized exclusively to chromosome 19, band p13.1 (Fig. 3).

**Northern Blot Analysis.** Multiple-tissue Northern blots were hybridized with a probe containing the first 500 bp of the
2.8-kb cDNA (Fig. 4 A and B). A 4.4-kb transcript appeared to be expressed ubiquitously, with the highest level of expression in skeletal muscle, placenta, testis, and peripheral blood leukocytes. A 2.8-kb transcript was also present in placenta, peripheral blood, and testis. On prolonged exposures of the autoradiographs, this 2.8-kb transcript was also visualized in other tissues. Hybridization of this same probe to a Northern blot of hematopoietic cell lines revealed expression of both transcripts in T-cell, B-cell, and myeloid lineages (Fig. 4C).

**Evolutionary Conservation.** To determine whether this gene was evolutionarily conserved, blots containing DNA from multiple mammalian and other species were hybridized with a probe containing the first 500 bp of the 2.8-kb cDNA (Fig. 5). The equivalent intensity of the bands in all mammalian species suggests a high degree of conservation. In addition to mammalian species, this gene appears to be present in chicken, fish, and frog.

**Southern Blot Analysis.** To determine whether DNA rearrangements could also be observed in patient samples by probes from the partner gene on 19p13.1, the blot that was used to detect MLL rearrangements was rehybridized with a probe containing the first 300 bp of the 2.8-kb cDNA (Fig. 2B). This probe contains the sequences that are immediately 5' and 3' to the breakpoint within the ELL cDNA and could thus potentially detect both the der(11) and the der(19). However, no gene rearrangements could be identified, suggesting that the exons of ELL contained within the probe may be far from the genomic translocation breakpoint junction on chromosome 19.

**Sequence Analysis.** The entire 2.8-kb cDNA fragment was sequenced on both strands. The sequence that diverged from MLL in the fusion transcript begins at nt 148 of the 2.8-kb cDNA (Fig. 6A). The chimeric transcript generated by this translocation consists of 5' MLL through exon 7 fused in frame to 2747 nt of ELL (Fig. 6B). The ELL sequence in the fusion transcript contains a predicted open reading frame of 576 aa followed by 930 nt of 3' untranslated sequence. In addition to the sequence present in the fusion transcript, the 2.8-kb cDNA contains 124 nt 5' of the breakpoint on chromosome 19. When combined with the 23 nt obtained from 5' RACE, the sequence 5' of the breakpoint codes for 45 aa. The open reading frame of the unarranged ELL gene codes for a predicted protein of 621 aa.

**Homology Searches.** There are no significant nucleotide or amino acid homologies to the other MLL partner genes that have been cloned. However, ELL contains several regions that are rich in basic amino acids as has been noted in AF4, ENL, and AF9. The highest area of homology exists in a highly basic, lysine-rich region beginning at amino acid 442 of ELL. This region is homologous to the basic region of the DNA-binding domain of poly(ADP-ribose) polymerase (Fig. 7) (18–20). In addition, this region is also homologous to a lysine-rich region of arginine-rich nuclear protein (21). Data base searches also revealed a long stretch of limited homology to the Engrailed gene family, primarily upstream of the homeobox domain (22). A stretch of amino acids in this lysine-rich region is identical to a repeated element in the knob-associated histidine-rich protein of *Plasmodium falciparum* (23, 24). However, the adjacent amino acids are not highly conserved. In addition, the bovine leukemia virus receptor has a similar basic region near its N terminus (25).

**DISCUSSION**

Although MLL encodes DNA-binding motifs suggestive of a transcription factor, the normal functions of MLL have not
The precise breakpoints and regions of mRNA juxtaposition that lead to the formation of the chimeric mRNA and its roles in the leukemogenic function of the fusion transcript have been identified in previous studies. These breakpoints and regions are likely to include the 3' terminal portions of the MLL genes and the 5' terminal portions of the ELL gene. The fusion mRNA is transcribed from the MLL gene and is highly conserved, suggesting that it plays a critical role in the leukemogenic process.

**Figure 6.** (A) Nucleotide sequence and predicted open reading frame of the 2.8-kb ELL fragment. The breakpoint is indicated by an arrow. The termination codon is indicated by an asterisk. (B) Sequence of the fusion transcript.

Yet another breakpoint has been identified in the MLL gene, which could potentially contribute to the leukemogenic process. This breakpoint is located in the 3' terminal portions of the MLL genes and is highly conserved, suggesting that it plays a critical role in the leukemogenic function of the fusion transcript. The fusion transcript is transcribed from the MLL gene and is highly conserved, suggesting that it plays a critical role in the leukemogenic process.

**Homologies to other genes with highly basic, lysine-rich motifs have been identified.** Arginine-rich nuclear protein was cloned from an expression library screened with an antibody to nuclear proteins. The precise function of arginine-rich nuclear protein has not been determined, but it contains a region of homology to the U1 protein of small nuclear ribonucleoprotein particles and is thus thought to be involved with RNA processing. The homology to the engrafted family of genes spans a region in the E8-4 domain and the beginning of the homeodomain. Taken together, these homologies suggest that the ELL may also be a nuclear protein that potentially has DNA binding properties. A similar lysine-rich motif is present in the bovine leukemia virus receptor and the known-associated histidine-rich protein of *Plasmodium falciparum*. The significance of the homologies to these two genes is not clear.
None of the six MLL fusion partner genes cloned to date had previously been identified. The potential normal functions of these genes have not been characterized and thus comparisons of these genes can be based only on their nucleotide sequence and predicted proteins. The proteins encoded by AF4, AF9, and ENL contain nuclear targeting sequences and regions rich in serine and proline residues, suggesting that these proteins may act as transcription factors. These three proteins also have segments that are rich in basic amino acids, but none of these proteins contains a known DNA-binding domain. AF6 and AF1p do not appear to be similar to AF4, AF9, and ENL. AF6 encodes short stretches rich in prolines, charged amino acids, serines, and glutamates and does not have a nuclear localization signal. AF6 also encodes a GLGF motif found in proteins thought to be involved in signal transduction. AF1p encodes three acidic domains, a glutamine-rich region, and 15 copies of a DPF tripeptide repeat of unknown significance. AF1p is highly homologous to the murine eps 15 gene that encodes a cytoplasmic phosphoprotein. Only partial sequence of the AFX gene has been published. ELL is not homologous to any of the partner genes that have been cloned. Like AF4, AF9, and ENL, ELL encodes segments that are rich in basic amino acids, although this is a more prominent feature of ELL than of the other translocation partners. In addition, ELL also encodes a potential nuclear localization sequence.

Further work will be necessary to elucidate the normal functions of ELL as well as the contribution of its 3' sequences to leukemogenesis. The homology of its highly basic, lysine-rich region to a DNA-binding domain suggests that ELL may also be a nuclear protein with DNA-binding properties. One potential consequence of the formation of the MLL/ELL fusion transcript may be an altered DNA-binding specificity that results from the substitution of the DNA-binding domain of ELL for the loss of the zinc fingers encoded by 5' MLL. The characterization of ELL should provide insights into its function and into the pathogenesis of acute leukemia.

We thank Lisa Gottschalk for the preparation of figures. This research was supported in part by grants from the National Institutes of Health (CA42557 to J.D.R. and HL52904 to M.C.S.), from the Department of Energy (DE-FG02-86ER62048 to J.D.R.), and from the Spastic Paralysis Research Foundation, Illinois–Eastern Iowa District of Kiwanis International (to J.D.R.). M.J.T. is a recipient of a Postdoctoral Research Fellowship for Physicians from the Howard Hughes Medical Institute.