2762–2769 Nucleic Acids Research, 1995, Vol. 23, No. 14

Alternative splicing and genomic structure of the *AML1* gene involved in acute myeloid leukemia

Hiroyuki Miyoshi, Miki Ohira, Kimiko Shimizu, Kinuko Mitani[1], Hisamaru Hirai[1], Takashi Imai[2], Kazushige Yokoyama[3], Eiichi Soeda[3] and Misao Ohki*

Radiobiology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan, [1]Third Department of Internal Medicine, Faculty of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan, [2]Division of Genetics, National Institute of Radiological Sciences, Anagawa 4-9-1, Chiba 260, Japan and [3]Tsukuba Life Science Center, RIKEN, 3-1-1 Koyodai, Tsukuba, Ibaraki 305, Japan

Received March 7, 1995; Revised and Accepted June 5, 1995

DDBJ accession nos D43967–D43969 (incl.)

ABSTRACT

We previously isolated the *AML1* gene, which is rearranged by the t(8;21) translocation in acute myeloid leukemia. The *AML1* gene is highly homologous to the *Drosophila* segmentation gene *runt* and the mouse transcription factor PEBP2 α subunit gene. This region of homology, called the Runt domain, is responsible for DNA-binding and protein–protein interaction. In this study, we isolated and characterized various forms of *AML1* cDNAs which reflect a complex pattern of mRNA species. Analysis of these cDNAs has led to the identification of two distinct AML1 proteins, designated AML1b (453 amino acids) and AML1c (480 amino acids), which differ markedly from the previously reported AML1a (250 amino acids) with regard to their C-terminal regions, although all three contain the Runt domain. The large C-terminal region common to AML1b and AML1c is suggested to be a transcriptional activation domain. AML1c differs from AML1b by only 32 amino acids in the N-terminal. Characterization of the genomic structure revealed that the *AML1* gene consists of nine exons and spans >150 kb of genomic DNA. Northern blot analysis demonstrated the presence of six major transcripts, encoding AML1b or AML1c, which can all be explained by the existence of two promoters, alternative splicing and differential usage of three polyadenylation sites. A minor transcript encoding AML1a which results from alternative splicing of a separate exon can be detected only by reverse transcription–polymerase chain reaction amplification. The distinct proteins encoded by the *AML1* gene may have different functions, which could contribute to regulating cell growth and/or differentiation through transcriptional regulation of a specific subset of target genes.

INTRODUCTION

The t(8;21)(q22;q22) translocation is one of the most frequent chromosome abnormalities in acute myeloid leukemia (AML). We previously isolated the *AML1* gene, located on chromosome 21, which is rearranged by the t(8;21) translocation (1). The t(8;21) breakpoints are consistently clustered within a specific intron of *AML1* (2–5). It has subsequently been shown that this translocation juxtaposes the *AML1* gene with the *MTG8(ETO)* gene on chromosome 8, and the resultant AML1–MTG8(ETO) fusion product has been suggested to be responsible for leukemogenesis (6–8). The *AML1* gene is also disrupted in the t(3;21) translocation in blast crisis of chronic myeloid leukemia (CML) and therapy-related AML, leading to the production of fusion genes (9–12).

The *AML1* gene product shares a 128 amino acid region of homology, called the Runt domain, with the *Drosophila* runt protein and the α subunit of polyomavirus enhancer binding protein 2 (PEBP2α) (6,13,14). The *runt* gene is a pair-rule gene and regulates the expression of other pair-rule genes (15). This gene is also involved in somatic sex determination and neural development (16,17). PEBP2, also called CBF (18), is a transcription factor consisting of a heterodimer with α and β subunits. PEBP2α is capable of binding to a core sequence of polyomavirus enhancer and its association with PEBP2β results in greater affinity to DNA. The consensus binding sequence, (Pu/T)ACCPuCA (19,20), is also found in the enhancer regions of murine leukemia virus, several T cell-specific genes and the myeloperoxidase gene (21,22). Interestingly, PEBP2β (18,23) is involved in the inversion of chromosome 16 in AML (24). Recently, the Runt domain has been shown to be essential for DNA-binding and heterodimerization properties (20,25). The Runt domain remains intact in the fusion proteins resulting from translocations, suggesting that it is an important functional domain of oncoproteins.

Northern blot analysis has demonstrated the presence of various sizes of *AML1* mRNA (1,7). In this study, we isolated and

* To whom correspondence should be addressed

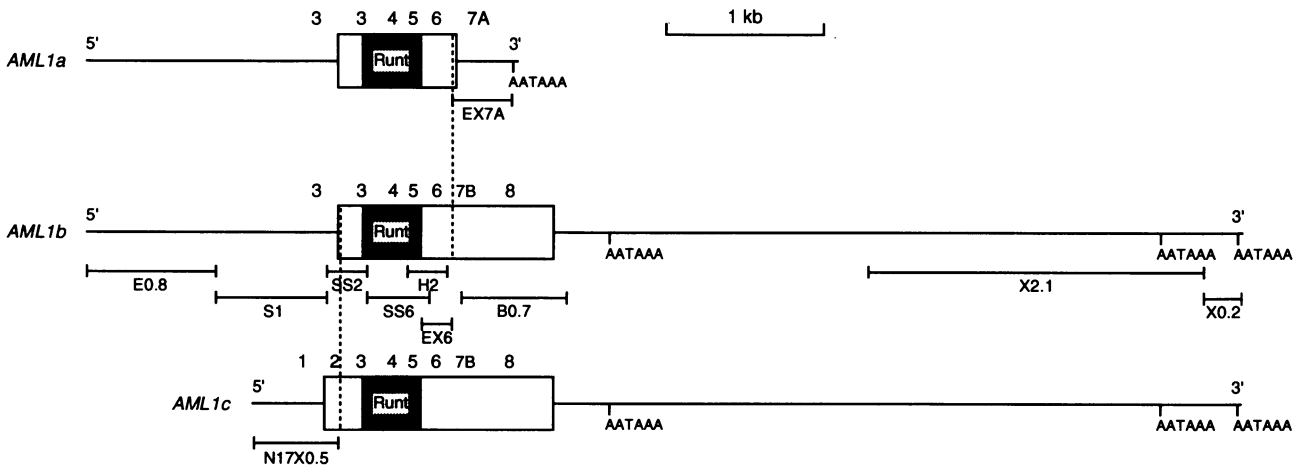

Figure 1. Schematic representation of the *AML1* cDNAs. Open boxes represent coding regions, and the Runt domain is shaded. Boundaries of exons are indicated by solid triangles. Positions of polyadenylation signals are shown by AATAAA. Probes used for Northern blot analysis are shown below.

characterized various forms of *AML1* cDNAs to determine the origin of these mRNAs. In addition, we determined the genomic structure, including the exon–intron boundaries. This information should be useful in further studies of the *AML1* gene and its involvement in leukemia.

MATERIALS AND METHODS

Cloning of *AML1* cDNAs

Poly(A)$^+$ RNA from Namalwa cells was converted to double-strand cDNA. A size-fractionated cDNA library (>2.0 kb) was constructed in the λ ZAPII vector (Stratagene) with *Eco*RI–*Not*I adaptors (Pharmacia). Human bone marrow and peripheral blood leukocyte cDNA libraries constructed in the λ gt10 vector were purchased from Clontech. The cDNA libraries prepared from Kasumi-1 and SKH1 cells, t(8;21)- and t(3;21)-carrying cell lines, respectively, have been described previously (7,10). The libraries were screened with a *AML1* cDNA probe (H2, identical to C6E6H2) (1) and multiple positive plaques were initially isolated. The cDNA inserts from each positive clone were subcloned into appropriate restriction enzyme sites of pBluescript II (KS+) (Stratagene) and characterized by restriction enzyme mapping and DNA sequencing.

DNA sequencing

The inserts of cDNA clones were digested with appropriate restriction enzymes and subcloned into the pBluescript II (KS+) vector (Stratagene). A nested series of deletions was generated using the Exo/Mung Kit (Stratagene). Both strands of cDNA clones were sequenced by the dideoxy nucleotide chain termination method using an A.L.F. DNA sequencer (Pharmacia). Cycle sequencing for genomic clones was performed with synthetic oligonucleotide primers corresponding to the cDNAs using the AutoCycle sequencing kit (Pharmacia).

Isolation and characterization of genomic clones

Three yeast artificial chromosome (YAC) clones (154, 155 and 158) were isolated from the YAC library described previously (26) using *AML1*-specific primers for polymerase chain reaction (PCR) amplification. These clones were characterized by pulsed-field gel electrophoresis (PFGE) and Southern blot hybridization, and all gave similar or overlapping restriction maps. A cosmid library was prepared from *Sau*3AI partial digests of yeast DNA containing the YAC clone 154 with a 350 kb genomic insert and cloned into the *Bam*HI site of the pWE15 cosmid vector (Stratagene). The cosmid library was screened initially with the YAC clone 154 DNA as probe and subsequently with the *AML1b* cDNA probe. A cosmid contig was constructed by comparing restriction enzyme digestion patterns using *Eco*RI and *Bam*HI and by hybridization of the cosmids with RNA probes derived from the end of each cosmid using T7 and T3 RNA polymerases. Three other YAC clones (316E12, 790C02 and 806E05) were isolated from CEPH YAC libraries using an *AML1c*-specific probe (N17X0.5). P1 clones (T674 and S687) were also isolated from a chromosome 21-specific P1 library (Tanahashi *et al.*, manuscript in preparation) with the N17X0.5 probe.

Determination of the exon–intron boundaries

The location of exons was determined by restriction mapping and Southern blot analysis of genomic clones using various parts of *AML1* cDNAs as probes. The exon–intron boundaries, except for exon 2, were determined by cycle sequencing of cosmid and P1 clones using oligonucleotide primers corresponding to the *AML1* cDNAs. The ligation-mediated PCR method was used to determine the exon–intron boundaries of exon 2. Agarose plugs containing DNA from YAC clones (316E12, 790C02 and 806E05) were digested with *Alu*I, *Hae*III or *Rsa*I and ligated with annealed linker (27) as described previously (28). The ligated mixture was amplified by PCR using the linker primer (5′-GCGGTGACCCGGGAGATCTGAATTC-3′) and primer CF5 (5′-AATGCATACTTGGAATGAATCCTTCTAGAGACG TC-3′, nucleotides 503–537 of *AML1c*) or primer CR4 (5′-CGTGGACGTCTCTAGAAGGATTCATTCCAAGTATG- 3′, nucleotides 507–541 antisense of *AML1c*). PCR amplification was carried out for 35 cycles (denaturation at 94°C for 1 min, annealing at 60°C for 1 min, and extension at 72°C for 2 min) in a GeneAmp PCR System 9600 (Perkin Elmer). A second PCR

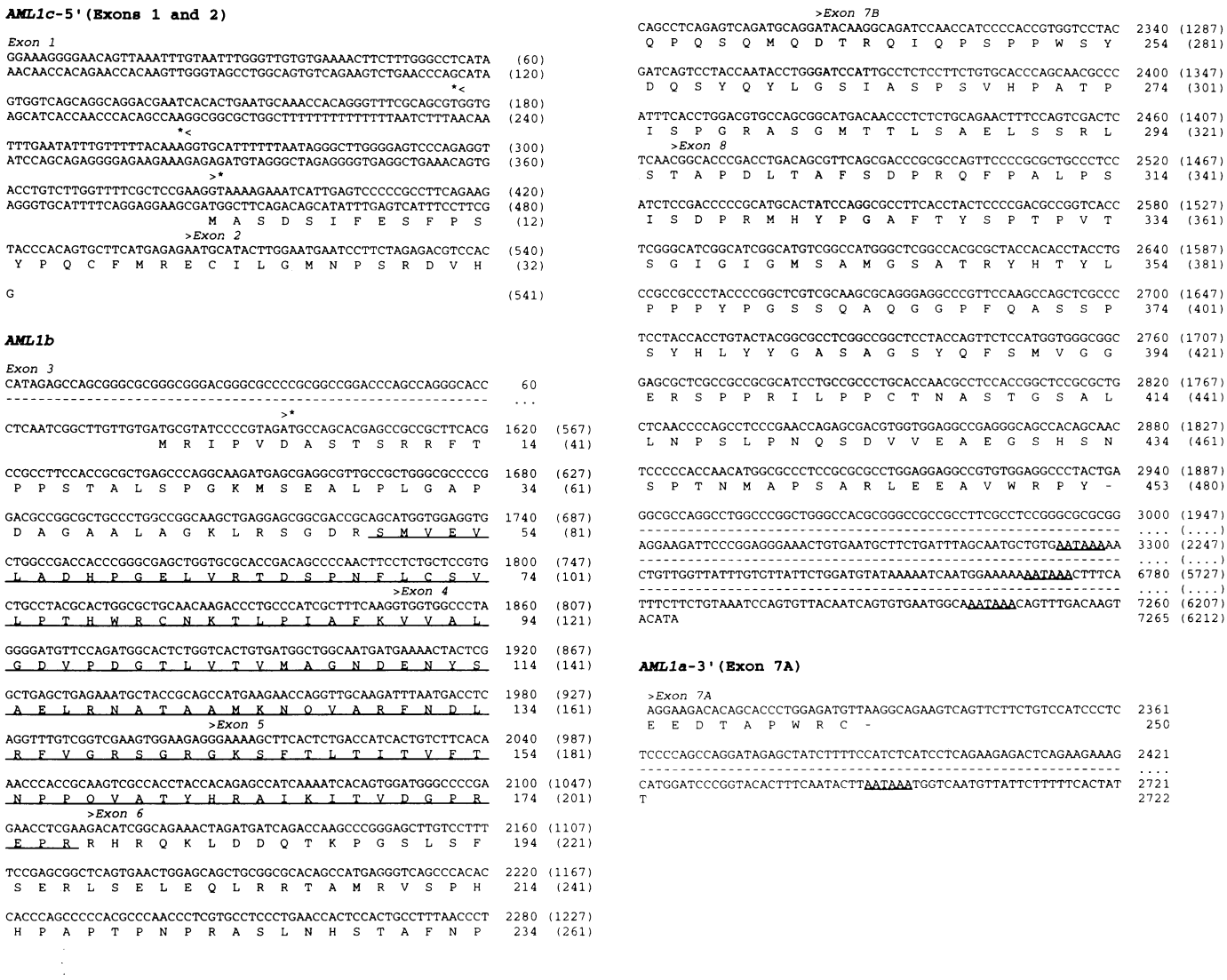

Figure 2. Nucleotide and deduced amino acid sequences of *AML1* cDNAs. The nucleotide and amino acid numbers are based on the sequence of the *AML1b* cDNA. The sequences of *AML1a* and *AML1c* cDNAs are indicated relative to the sequence of *AML1b*. The nucleotide and amino acid numbers of *AML1c* are shown in parentheses. The polyadenylation signals are underlined. Boundaries of exons are indicated by arrowheads above nucleotide sequences, at the beginning of each exon. Alternative splice sites are also indicated with an asterisk. The Runt domain is underlined. The sequence data reported here have been deposited in the GSDB, DDBJ, EMBL and NCBI sequence databases under the accession numbers D43967, D43968 and D43969.

amplification was carried out under identical conditions. The predominant PCR products were cloned into the pCR II vector (Invitrogen) and sequenced. Successful results were obtained using the PCR products from *Rsa*I digest.

Northern blot analysis

Total RNA was isolated from Namalwa cells. Poly(A)$^+$ RNA was selected using Oligotex-dT30 (Roche). Samples of poly(A)$^+$ RNA (1.5 μg) were separated by electrophoresis on 1% agarose–formaldehyde gels. The cDNA probes were radiolabeled with ^{32}P using the Multiprime DNA labeling system (Amersham). Blotting, hybridization, washing and autoradiography were performed as previously described (1). Human multiple tissue Northern blots were purchased from Clontech. Probes used for Northern blots were: EX7A, nucleotides 2303–2678 of *AML1a*; E0.8, nucleotides 1–805 of *AML1b*; S1, nucleotides 806–1494 of *AML1b*; SS2, nucleotides 1495–1754 of *AML1b*; SS6, nucleotides 1755–2145 of *AML1b*; H2, nucleotides 2011–2243 of *AML1b*; EX6, nucleotides 2111–2302 of *AML1b*; B0.7, nucleotides 2363–3027 of *AML1b*; X2.1, nucleotides 4931–7034 of *AML1b*; X0.2, nucleotides 7035–7265 of *AML1b*; N17X0.5, nucleotides 1–526 of *AML1c*.

RT–PCR analysis

Poly(A)$^+$ RNA (0.5 μg) from Namalwa cells was converted to first-strand cDNA primed with random hexamer using the Superscript preamplification system (Gibco-BRL) under the conditions recommended by the manufacturer. A total volume of 25 μl of the PCR mixture contained one-fourth of the cDNA, 200 nM of each primer, 250 μM of each dNTP, 1× PCR buffer (Takara) and 2 U *Taq* DNA polymerase (Boehringer Mannheim). PCR amplification was carried out for 30 cycles (denaturation at 94°C for 1 min, annealing at 62°C for 1 min, and extension at 72°C for 1 min), which were preceded by denaturation at 94°C for 2 min and followed by extension at 72°C for 5 min. One-half

Table 1. Exon–intron boundaries of the *AML1* gene

Splice acceptor		Splice donor
	Exon 1 (502 bp)...	TCATGAGAG gtgagtacatgctggtcttg
cctactttctgttatttcag AATGCATAC...	Exon 2 (39 bp)....	ACGTCCACG gtatgttaaattccaagtta
TGTGATGCGTATCCCCGTAG ATGCCAGCA...	Exon 3 (1848 bp)..	GCTTTCAAG gtactggcccggagagggtg
gtcctttgactggtgtttag GTGGTGGCC...	Exon 4 (157 bp)...	GTGGAAGAG gtacgttatctgtcaaaact
tctctatcgtgtccccacag GGAAAAGCT...	Exon 5 (105 bp)...	AACCTCGAA gtaagtgcatccacttgggg
ctccctgctccccacaatag GACATCGGC...	Exon 6 (192 bp)...	AGATGCAGG gtaagtaccagatggagcca
gtcacacatggctccctcag AGGAAGACA...	Exon 7A(420 bp)	
acctttctgattctcttcag ATACAAGGC...	Exon 7B(162 bp)...	GACTCTCAA gtaagccacttgaaaacaca
ttctcttgcccgccctgcag CGGCACCCG...	Exon 8 (4801 bp)	

Exon and intron sequences are shown in uppercase and lowercase letters, respectively. The alternative splice site in exon 3 is also shown.

of the PCR product was separated by electrophoresis on 2% agarose gel stained with ethidium bromide. The primers used were: C, 5′-GAGGGAAAAGCTTCACTCTGA-3′ (nucleotides 2003–2023 of *AML1b*); 3A, 5′-GTGTACCGGGATCCATGC-TA-3′ (nucleotides 2659–2678 antisense of *AML1a*); 2B, 5′-GTTGAGAGTCGACTGGAAAG-3′ (nucleotides 2446–2465 antisense of *AML1b*); CF2, 5′-CACAGAACCACAAGTTGGGT-AGCC-3′ (nucleotides 67–90 of *AML1c*); and CR1, 5′-GACTCA-AATATGCTGTCTGAAGCC-3′ (nucleotides 447–470 antisense of *AML1c*).

RESULTS

Isolation and characterization of various forms of *AML1* cDNAs

We previously reported that Northern blotting revealed *AML1* mRNAs of different sizes (1,7). However, the expression pattern of these mRNAs is similar in various human hematopoietic cell lines as well as in bone marrow and peripheral blood cells (7). To determine the origin of these various mRNA isoforms, several cDNA libraries (mainly a Namalwa cDNA library) were screened with a *AML1* cDNA probe (H2), and multiple positive clones were isolated and characterized by restriction mapping. Of these, a total of 18 independent clones were selected for further analysis. Nucleotide sequence analysis of the cDNA clones showed that the predicted AML1 proteins could be classified into two types, designated AML1b (453 amino acids) and AML1c (480 amino acids), with different N-terminal regions (Fig. 1). The N-terminal region of AML1b is identical to that of the previously reported AML1 protein (1), designated AML1a (250 amino acids). AML1b and AML1c contain the same large C-terminal region, which is rich in proline, serine and threonine. This C-terminal region has been suggested to be a transcriptional activation domain (29,30). The Runt domain is fully conserved among the three species. The AML1b protein is very similar to the recently reported mouse AML1, PEBP2αB (31), with 96% homology. Comparison of the nucleotide sequence of *AML1b* cDNA to *PEBP2αB* cDNA demonstrated significant homology, even in the 5′- and 3′-untranslated regions (data not shown). The sequence of the longest (7.3 kb) *AML1b* cDNA is shown in Figure 2. The sequences of other cDNAs are indicated relative to this sequence. The previously reported *AML1a* cDNA clone ended 27 nucleotides after a consensus polyadenylation signal (AATAAA) with a long poly(A) tail. The extremely large 3′-untranslated region of *AML1b* and *AML1c* contained three polyadenylation signals and more than half of the cDNA clones ended immediately downstream of the first polyadenylation signal with a poly(A) tail, suggesting differential usage of polyadenylation signals.

Genomic structure of the *AML1* gene

To characterize the genomic structure, we isolated three YAC clones (154, 155 and 158) containing the *AML1* gene and constructed a cosmid library from yeast DNA harboring YAC clone 154 (see Materials and methods). A cosmid contig was constructed by restriction mapping and riboprobe walking (Fig. 3). Southern blot hybridization of cosmid clones with various parts of *AML1* cDNAs provided information on the location of each exon on the physical map. The 5′ end of *AML1c* cDNA was not present within the cosmid contig and library. We therefore screened a chromosome 21-specific P1 library with an *AML1c*-specific probe (N17X0.5) and two P1 clones (T674 and S687) were isolated. When exon–intron boundary sequence analysis was performed, it became apparent that the 3′ end of the *AML1c*-specific region was not present within these P1 clones. Therefore, three YAC clones (316E12, 790C02 and 806E05) were isolated from CEPH YAC libraries using the N17X0.5 probe. The ligation-mediated PCR method with these YAC clones was then used to determine the exon–intron boundaries of exon 2. The other exon–intron boundaries were determined by cycle sequencing of cosmid and P1 clones using oligonucleotide primers corresponding to the cDNAs. The *AML1* gene contains nine exons, and all exon–intron boundaries conform to the consensus splice donor and acceptor sequences (32) including the GT–AG motif (Table 1). The approximate location of each exon on the physical map is shown in Figure 3. Based on this map, we estimated that the *AML1* gene is larger than 150 kb, although the lengths of introns 1 and 2 remain unknown. According to the gene organization, the t(8;21) breakpoint cluster region is located in intron 5. The partial genomic structure reported previously (1), which was based on Southern blot hybridization of genomic DNA with *AML1a* cDNA, was found to contain two extra exons, probably due to cross-hybridizing of the Runt domain-encoding region to closely related genes.

Exons 1 and 3 contain the 5′-untranslated region and the ATG initiation codon. *AML1c* arises from the use of an internal splice acceptor site in exon 3, while *AML1a* and *AML1b* begin at exon 3, probably due to utilization of an alternative promoter. This results in a protein with a different N-terminal. The Runt domain is encoded by part of exon 3, exon 4 and exon 5. Exons 7A and 8 contain the TAA and TGA termination codons, respectively, and the entire 3′-untranslated region. As a result of alternative splicing,

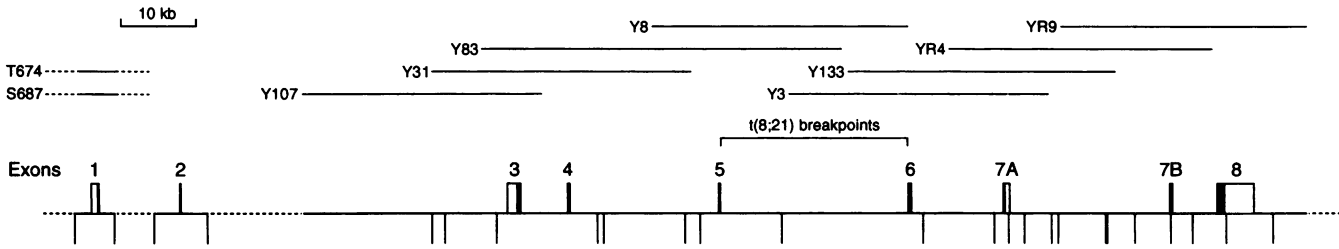

Figure 3. Genomic organization of the *AML1* gene. The *Bam*HI restriction sites are indicated by vertical lines. Exons are numbered and indicated by vertical boxes, with coding regions indicated by solid boxes. The P1 and cosmid clones are shown above. The t(8;21) breakpoint cluster region is located in intron 5.

AML1a contains exon 7A. On the other hand, *AML1b* and *AML1c* contain exons 7B and 8, resulting in proteins with a larger C-terminal. The three patterns of alternative splicing are schematically shown in Figure 4. Recently, Levanon *et al.* (33) reported an AML1 protein sequence which is almost identical to AML1b but differs at the C-terminal end due to a 99 bp deletion in the cDNA as compared with our sequence analysis. We have also isolated three cDNAs which have the same deletion. However, since there is no consensus splice sequences around the deleted sequences and the deleted region is extremely GC-rich, these clones could result from a cloning artifact.

Expression of alternatively spliced transcripts

To determine the existence and relative expression levels of alternative forms of the *AML1* transcript, the series of cDNA probes shown in Figure 1 were hybridized to the Northern blot of RNA from Namalwa cells. As shown in Figure 5, probes (SS6, H2 and EX6) derived from the conserved region among the three species detected six major transcripts of 7.5, 7.0, 6.3, 5.8, 3.5 and 2.2 kb. The N17X0.5 probe specific for *AML1c* detected three transcripts of 6.3, 5.8 and 2.2 kb. Probes (E0.8 and S1) derived from the 5′-untranslated region of *AML1a* and *AML1b* detected three transcripts of 7.5, 7.0 and 3.5 kb. The SS2 probe detected the six major transcripts, although the 6.3, 5.8 and 2.2 kb bands were weaker than the 7.5, 7.0 and 3.5 kb bands because this probe contains not only a common region to the three species but also includes an *AML1a*- and *AML1b*-specific region. No detectable bands were observed with the *AML1a*-specific probe (EX7A) corresponding to exon 7A. On the other hand, the B0.7 probe corresponding to part of exons 7B and 8 detected the six major transcripts. Therefore, we conclude that the predominant species are the 7.5, 7.0 and 3.5 kb *AML1b* transcripts and the 6.3, 5.8 and 2.2 kb *AML1c* transcripts, and that *AML1a* is a minor transcript which can be detected by RT–PCR amplification (Fig. 6A). Since the 3′-untranslated region of *AML1b* and *AML1c* encoded by exon 8 contained three polyadenylation signals, we tested whether the different sizes of the transcripts were due to differential usage of the polyadenylation sites using two probes; one (X2.1) from the region between the first and the third polyadenylation signals and the other (X0.2) from the region downstream of the second polyadenylation signal. Consistent with this hypothesis, the X2.1 probe detected four transcripts of 7.5, 7.0, 6.3 and 5.8 kb, and the X0.2 probe detected only two transcripts of 7.5 and 6.3 kb, representing transcripts which terminated at the third polyadenylation site (Fig. 5). Furthermore, the location of the three polyadenylation signals was consistent with the difference in the sizes of each of the three transcripts for *AML1b* and *AML1c*. Although the precise transcription start site has not yet been determined, considering the sizes of transcripts detected on Northern blots, the cloned cDNAs should represent almost full-length clones.

Several cDNA clones for *AML1b* and *AML1c* have an internal deletion of 192 bp, which corresponds to exon 6, resulting in proteins with a deletion of 64 amino acids immediately downstream of the Runt domain without disrupting the correct reading frame. However, the EX6 and B0.7 probes recognized the same transcripts on Northern blots (Fig. 5), suggesting that transcripts lacking exon 6 are expressed at low levels. To determine whether such transcripts could be detected, we performed RT–PCR analysis using oligonucleotide primers located on each side of exon 6. As shown in Figure 6A, PCR products of the two expected sizes were detected and differed by ~200 bp, which corresponds to the size of exon 6. The EX6 probe recognized only the larger PCR product on Southern blotting (data not shown). Thus, the transcripts lacking exon 6 are expressed. RT–PCR analysis using a primer set designated to amplify *AML1a* transcript also revealed the existence of transcript lacking exon 6 (Fig. 6A). In addition, a small number of *AML1c* cDNA clones showed a 122 bp (nucleotides 264–385) deletion or a 208 bp (nucleotides 178–385) deletion in the 5′-untranslated region encoded by exon 1. Both ends of the deleted sequences conform to the consensus splice donor and acceptor sites (Fig. 2), which supports these alternative transcription events. As expected, RT–PCR analysis of the 5′-untranslated region of *AML1c* transcripts yielded two additional PCR products representing these deletions (Fig. 6B).

The tissue distribution of the different *AML1* transcripts was examined by Northern blot analysis of RNA from various human tissues (Fig. 7). The expression of *AML1* was detected in all tissues examined except for brain and heart; however, the expression levels of the six major transcripts varied among the different tissues. The ratio of the *AML1c* transcripts to the *AML1b* transcripts was considerably higher in thymus and spleen than in other tissues. This may result from the existence of tissue-specific regulatory elements.

DISCUSSION

In this report, we have elucidated the genomic structure of the *AML1* gene and identified various forms of its transcript. The major mRNA species encode the AML1b and AML1c proteins, while the previously reported cDNA sequence encoding AML1a is derived from a minor species. The differently sized mRNAs of AML1b and AML1c are the result of differential usage of three polyadenylation signals. It is interesting to note that the 3′-untrans-

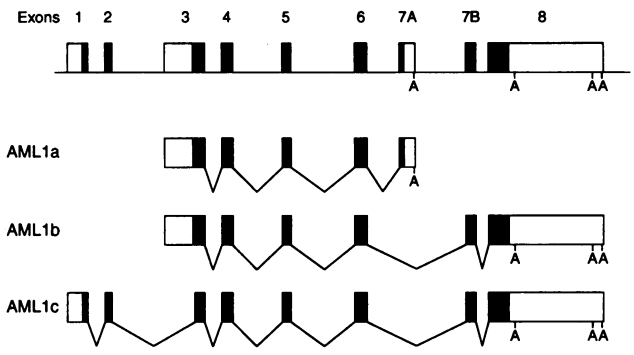

Figure 4. Schematic representation of alternative splicing of the *AML1* transcripts. A schematic representation of the exons is shown at the top. Solid boxes represent the coding regions, and open boxes represent the 5′- and 3′-untranslated regions. Positions of polyadenylation signals are shown by A.

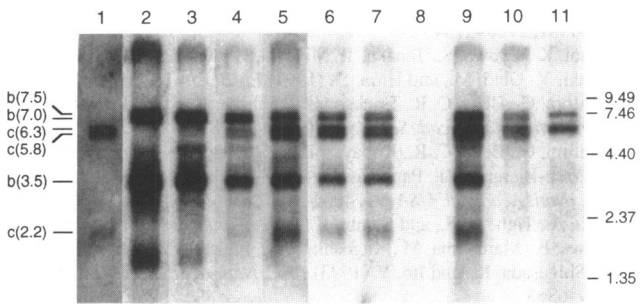

Figure 5. Northern blot analysis of the *AML1* transcripts. Northern blots of RNA from Namalwa cells were hybridized with various cDNA probes shown in Figure 1. Lanes: 1, N17X0.5; 2, E0.8; 3, S1; 4, SS2; 5, SS6; 6, H2; 7, EX6; 8, EX7A; 9, B0.7; 10, X2.1; 11, X0.2. The locations of *AML1b* (7.5, 7.0 and 3.5 kb) and *AML1c* (6.3, 5.8 and 2.2 kb) transcripts are shown on the left. The locations of RNA size markers (kb) are shown on the right.

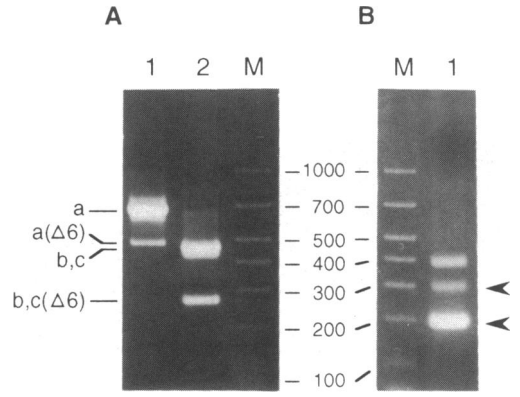

Figure 6. RT–PCR analysis. (**A**) Detection of *AML1* transcripts lacking exon 6. Primers C and 3A were used to preferentially amplify the *AML1a* transcript (lane 1). Primers C and 2B were used to preferentially amplify the *AML1b* and *AML1c* transcripts (lane 2). Δ6 represents the deletion of exon 6. (**B**) Detection of alternatively spliced *AML1c* transcripts. Primers CF2 and CR1 were used to preferentially amplify the 5′-untranslated region of the *AML1c* transcript (lane 1). PCR products derived from the alternatively spliced transcripts are indicated by arrowheads. M, DNA size markers (bp).

lated region, particularly between the first and the second polyadenylation signals, contains multiple copies of the AUUUA sequence motif, which is thought to be responsible for destabilization of mRNA (34). However, a recent report has shown that the AUUUA motif is not sufficient to destabilize mRNA, but the UUAUUUA(U/A)(U/A) motif can function as a destabilizing element (35). In this regard, the region between the first and the second polyadenylation signals still contains two copies of this motif. Therefore, the two longer transcripts which used the second and third polyadenylation signals may be unstable, even though they are readily detected by Northern blotting in most tissues. Several studies have shown that the 3′-untranslated regions of mRNAs are implicated in many functions, such as translational regulation (36,37), regulation of cell growth and differentiation (38), mRNA localization during oogenesis (39–41), and intracellular mRNA localization (42). Although no experimental evidence is currently available, variations in the lengths of the 3′-untranslated regions of *AML1* could be of functional importance. The expression levels of *AML1b* and *AML1c* transcripts vary between different tissues, suggesting the existence of tissue-specific regulatory elements in their promoter regions. Recently, it has been shown that the expression of the *AML1* gene in skeletal muscle increases greatly (50–100-fold) following denervation (43). It would be of interest to identify both these regulatory elements and the promoter regions.

AML1b has been shown to function as a transcriptional activator (29,30). Since AML1c differs from AML1b by only 32 amino acids in the N-terminal, which have no significant functional features, it is most likely that AML1c has the same function as AML1b. On the other hand, AML1a has no transcriptional activation property because it lacks a putative transcriptional activation domain in the C-terminal region (29,30). Although AML1a is a minor product, since it has a higher affinity to DNA than AML1b and suppresses transcriptional activation by AML1b (30), AML1a may act as a negative regulator. In fact, overexpression of AML1a blocks granulocytic differentiation and stimulates proliferation in 32Dcl3 cells treated with granulocyte colony-stimulating factor, whereas overexpression of AML1b has no significant effects (30). As has been recently shown, AML1 regulates the expression of several hematopoietic lineage-specific genes, such as T cell receptors, myeloperoxidase and macrophage colony-stimulating factor (colony-stimulating factor 1) receptor genes (29,44,45). These facts raise the possibility that AML1 contributes, by regulating the expression of target genes, to hematopoietic cell growth and/or differentiation. Further studies are required to clarify the exact role of *AML1* gene products and to elucidate how AML1–MTG8(ETO) fusion protein resulting from t(8;21) translocation contributes to leukemogenesis.

Human and mouse *AML1* transcripts are of similar sizes in most cell lines and tissues (7,31,46). A remarkable nucleotide sequence conservation between *AML1b* and *PEBP2αB*, which is the mouse homolog of *AML1*, is observed even in the 5′- and 3′-untranslated regions. Furthermore, based on the limited structural information regarding the *PEBP2αB* gene (29), the genomic structure of the *AML1* gene is likely to be highly conserved between human and mouse. These facts suggest that similar *AML1* transcripts are expressed in human and mouse. We have shown here that human *AML1* transcripts lacking exon 6, which results in proteins with a

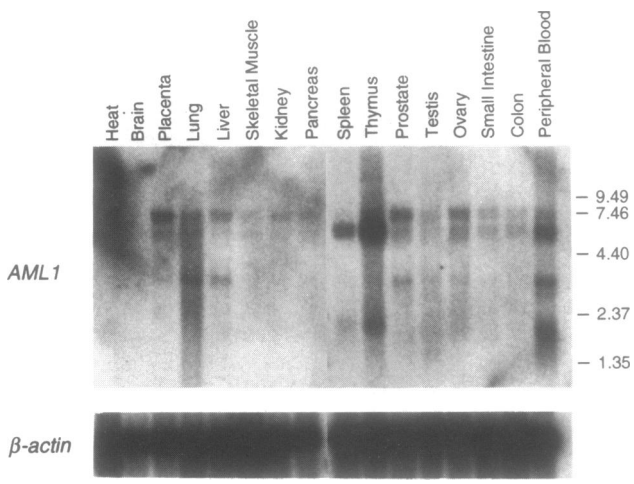

Figure 7. Tissue distributions of the *AML1* transcripts. Northern blots of RNA from various human tissues were sequentially hybridized with the *AML1* cDNA probe (H2) and the β-actin probe. The locations of RNA size markers (kb) are shown on the right.

deletion of 64 amino acids immediately downstream of the Runt domain, are minor products. On the other hand, *PEBP2αB2*, which has the same deletion, is one of two major mouse *AML1* transcripts in Ha-*ras*-transformed NIH 3T3 cells (29). PEBP2αB2 has a lower activity of transcriptional activation and a higher affinity to DNA than the other major protein, PEBP2αB1(mouse AML1b). Interestingly, t(3;21) breakpoints in several patients are located not in intron 5 but in intron 6, and the resultant fusion proteins therefore contain 64 amino acids corresponding to exon 6 (9,11,12). Further studies are needed to determine whether this region encodes a functionally important regulatory domain.

There are currently three closely related genes, *AML1(PEBP2αB)*, *AML2(PEBP2αC)* and *AML3(PEBP2αA)*, which code for proteins containing the Runt domain (33,47). Proteins encoded by this gene family have an extensive region of similarity besides the Runt domain. In addition, all of the presently available structural information suggests that the genomic structures of the members of this gene family would be well conserved. It is noteworthy that human AML2(PEBP2αC) has an internal deletion corresponding to exon 7B of AML1, and yet still possesses transcriptional activation property (47). The distinct proteins containing the Runt domain may have a variety of functions and differentially regulate the expression of target genes.

The structural information regarding the *AML1* gene provided in this report should facilitate future studies on evolutionary relationships between the members of the Runt domain-encoding gene family. This information should also provide a basis for a detailed analysis of the functional domains.

ACKNOWLEDGEMENTS

We thank M. Fukushima for providing technical assistance and N. Kobayashi for help in preparing the manuscript. This work was supported in part by Grants-in-Aid for Creative Basic Research (Human Genome Program) and for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture; by a grant from the Special Coordination Funds for the Promotion of Science and Technology from the Science and Technology Agency; and by a Grant-in-Aid for the Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health and Welfare of Japan.

REFERENCES

1 Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991) *Proc. Natl. Acad. Sci. USA*, **88,** 10431–10434.
2 Shimizu, K., Miyoshi, H., Kozu, T., Nagata, J., Enomoto, K., Maseki, N., Kaneko, Y., and Ohki, M. (1992) *Cancer Res.*, **52,** 6945–6948.
3 Tighe, J. E., Daga, A., and Calabi, F. (1993) *Blood*, **81,** 592–596.
4 Nucifora, G., Birn, D. J., Erickson, P., Gao, J., LeBeau, M. M., Drabkin, H. A., and Rowley, J. D. (1993) *Blood*, **81,** 883–888.
5 Maseki, N., Miyoshi, H., Shimizu, K., Homma, C., Ohki, M., Sakurai, M., and Kaneko, Y. (1993) *Blood*, **81,** 1573–1579.
6 Erickson, P., Gao, J., Chang, K.-S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992) *Blood*, **80,** 1825–1831.
7 Miyoshi, H., Kozu, T., Shimizu, K., Enomoto, K., Maseki, N., Kaneko, Y., Kamada, N., and Ohki, M. (1993) *EMBO J.*, **12,** 2715–2721.
8 Kozu, T., Miyoshi, H., Shimizu, K., Maseki, N., Kaneko, Y., Asou, H., Kamada, N., and Ohki, M. (1993) *Blood*, **82,** 1270–1276.
9 Nucifora, G., Birn, D. J., Espinosa III, R., Erickson, P., LeBeau, M. M., Roulston, D., McKeithan, T. W., Drabkin, H., and Rowley, J. D. (1993) *Blood*, **81,** 2728–2734.
10 Mitani, K., Ogawa, S., Tanaka, T., Miyoshi, H., Kurokawa, M., Mano, H., Yazaki, Y., Ohki, M., and Hirai, H. (1994) *EMBO J.*, **13,** 504–510.
11 Nucifora, G., Begy, C. R., Erickson, P., Drabkin, H. A., and Rowley, J. D. (1993) *Proc. Natl. Acad. Sci. USA*, **90,** 7784–7788.
12 Nucifora, G., Begy, C. R., Kobayashi, H., Roulston, D., Claxton, D., Pedersen-Bjergaard, J., Parganas, E., Ihle, J. N., and Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA*, **91,** 4004–4008.
13 Daga, A., Tighe, J. E., and Calabi, F. (1992) *Nature*, **356,** 484.
14 Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K., and Ito, Y. (1993) *Proc. Natl. Acad. Sci. USA*, **90,** 6859–6863.
15 Kania, M. A., Bonner, A. S., Duffy, J. B., and Gergen, J. P. (1990) *Genes Dev.*, **4,** 1701–1713.
16 Duffy, J. B. and Gergen, J. P. (1991) *Genes Dev.*, **5,** 2176–2187.
17 Duffy, J. B., Kania, M. A., and Gergen, J. P. (1991) *Development*, **113,** 1223–1230.
18 Wang, S., Wang, Q., Crute, B. E., Melnikova, I. N., Keller, S. R., and Speck, N. A. (1993) *Mol. Cell. Biol.*, **13,** 3324–3339.
19 Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y., and Shigesada, K. (1990) *J. Virol.*, **64,** 4808–4819.
20 Meyers, S., Downing, J. R., and Hiebert, S. W. (1993) *Mol. Cell. Biol.*, **13,** 6336–6345.
21 Satake, M., Inuzuka, M., Shigesada, K., Oikawa, T., and Ito, Y. (1992) *Jpn. J. Cancer Res.*, **83,** 714–722.
22 Suzow, J. and Friedman, A. D. (1993) *Mol. Cell. Biol.*, **13,** 2141–2151.
23 Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y., and Shigesada, K. (1993) *Virology*, **194,** 314–331.
24 Liu, P., Tarle, S. A., Hajra, A., Claxton, D. F., Marlton, P., Freedman, M., Siciliano, M. J., and Collins, F. S. (1993) *Science*, **261,** 1041–1044.
25 Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M., and Gergen, P. (1993) *Trends Genet.*, **9,** 338–341.
26 Imai, T. and Olson, M. V. (1990) *Genomics*, **8,** 297–303.
27 Mueller, P. R. and Wold, B. (1989) *Science*, **246,** 780–786.
28 Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M., and Schlessinger, D. (1992) *Genomics*, **14,** 241–248.
29 Bae, S. C., Ogawa, E., Maruyama, M., Oka, H., Satake, M., Shigesada, K., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., and Ito, Y. (1994) *Mol. Cell. Biol.*, **14,** 3242–3252.
30 Tanaka, T., Tanaka, K., Ogawa, S., Kurokawa, M., Mitani, K., Nishida, J., Shibata, Y., Yazaki, Y., and Hirai, H. (1995) *EMBO J.*, **14,** 341–350.
31 Bae, S. C., Yamaguchi-Iwai, Y., Ogawa, E., Maruyama, M., Inuzuka, M., Kagoshima, H., Shigesada, K., Satake, M., and Ito, Y. (1993) *Oncogene*, **8,** 809–814.
32 Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.*, **50,** 349–383.
33 Levanon, D., Negreanu, V., Bernstein, Y., Bar-Am, I., Avivi, L., and Groner, Y. (1994) *Genomics*, **23,** 425–432.
34 Shaw, G. and Kamen, R. (1986) *Cell*, **46,** 659–667.

35 Lagnado, C. A., Brown, C. Y., and Goodall, G. J. (1994) *Mol. Cell. Biol.*, **14,** 7984–7995.
36 Wharton, R. P. and Struhl, G. (1991). *Cell*, **67,** 955–967.
37 Sirito, M., Lin, Q., Maity, T., and Sawadogo, M. (1994) *Nucleic Acids Res.*, **22,** 427–433.
38 Rastinejad, F. and Blau, H. M. (1993) *Cell*, **72,** 903–917.
39 Macdonald, P. M. and Struhl, G. (1988) *Nature*, **336,** 595–598.
40 Mowry, K. L. and Melton, D. A. (1992) *Science*, **255,** 991–994.
41 Gavis, E. R. and Lehmann, R. (1992) *Cell*, **71,** 301–313.
42 Davis, I. and Ish-Horowicz, D. (1991) *Cell*, **67,** 927–940.
43 Zhu, X., Yeadon, J. E., and Burden, S. J. (1994) *Mol. Cell. Biol.*, **14,** 8051–8057.
44 Nuchprayoon, I., Meyers, S., Scott, L. M., Suzow, J., Hiebert, S., and Friedman, A. D. (1994) *Mol. Cell. Biol.*, **14,** 5558–5568.
45 Zhang, D., Fujioka, K., Hetherington, C. J., Shapiro, L. H., Chen, H., Look, A. T., and Tenen, D. G. (1994) *Mol. Cell. Biol.*, **14,** 8085–8095.
46 Satake, M., Nomura, S., Yamaguchi-Iwai, Y., Takahama, Y., Hashimoto, Y., Niki, M., Kitamura, Y., and Ito, Y. (1995) *Mol. Cell. Biol.*, **15,** 1662–1670.
47 Bae, S. C., Takahashi, E., Zhang, Y. W., Ogawa, E., Shigesada, K., Namba, Y., Satake, M., and Ito, Y. (1995) *Gene*, in press.