Molecular Analysis of the t(8;21) Chromosomal Translocation in Acute Myeloid Leukemia

Hiroyuki Miyoshi

Summary

The t(8;21) translocation is one of the most frequent chromosomal abnormalities in acute myeloid leukemia (AML). A *Not* I linking clone specific for chromosome 21 was first used to detect the t(8;21) rearrangement on pulsed-field gel electrophoresis analysis. Chromosome walking from the *Not* I linking clone and subsequent cDNA screening resulted in the isolation of a novel gene, *AML1*, that was rearranged by the t(8;21) translocation. The *AML1* gene is highly homologous to the *Drosophila* segmentation gene *runt*, which encodes a nuclear protein and regulates the expression of other pair-rule genes. Detailed analysis of the breakpoints in 24 patients and a cell line with t(8;21), including three cases with complex translocations, revealed that all the breakpoints on chromosome 21 clustered within a single specific intron of *AML1*. Isolation and characterization of a fusion cDNA from the cell line with t(8;21) demonstrated that this translocation juxtaposed the *AML1* gene with a novel gene, *MTG8*, on chromosome 8, resulting in the synthesis of an *AML1-MTG8* fusion transcript. The AML1-MTG8 fusion protein consists of the *runt*-homology region of AML1 and the most part of MTG8 containing putative zinc finger DNA-binding motifs and proline-rich regions. On the basis of the structural features, both AML1 and MTG8 are inferred to be transcription factors. The expression of the *AML1-MTG8* fusion transcript is a consistent feature of the t(8;21) translocation, and thus the production of the chimeric AML1-MTG8 protein, probably a chimeric transcription factor, may contribute to myeloid leukemogenesis.

Introduction

Consistent chromosomal aberrations are frequently observed in human neoplasia (for reviews see Heim and Mitelman, 1989; Mitelman *et al.*, 1991). The neoplasia-associated chromosomal aberrations are generally classified as numerical or structural. Numerical abnormalities are gains or losses of whole chromosomes, whereas structural alterations include translocations, inversions, deletions, insertions, and amplifications. For technical reasons, our knowledge of chromosomal aberrations is most extensive for the hematopoietic malignancies. The quality of chromosome preparations for the solid tumors has been poor; the chromosomes are often contracted and fuzzy, the spreading is poor, and the banding is unsatisfactory. In addition, since the solid tumors carry numerous chromosomal abnormalities, it is difficult to determine whether these are consistent cytogenetic changes in the malignant cells. However, with improved techniques, the information on chromosomal abnormalities in solid tumors has now increased. Specific chromosomal abnormalities are closely associated with particular types or subtypes of cancer, suggesting that these rearrangements pinpoint the location of genes involved in tumor initiation or progression.

There appear to be at least two distinct mechanisms of tumorigenesis (for review see Bishop, 1991). One mechanism is associated with the activation of proto-oncogenes normally involved in the control of cellular growth and differentiation. The proto-oncogenes can be activated by dominant mutations, such as point mutations, small insertions and deletions, and juxtaposition to other chromosome sequences. This last event can be visualized cytogenetically as a translocation or inversion. Since many of these chromosomal changes involve previously uncharacterized regions, their molecular elucidation has resulted in the discovery of a number of novel proto-oncogenes with interesting and sometimes surprising relationships to familiar genes. An alternative mechanism is related to the loss or inactivation of genes referred to as tumor suppressor genes. The tumor suppressor genes are recessive, and so the inactivation of both alleles is required to initiate malignant growth. Loss of a chromosome, chromosome arm, or subchromosomal band is considered to be a cytogenetic hallmark of this mechanism. Consistent chromosomal deletions have provided the essential information regarding the

chromosome location of tumor suppressor genes and have led to the cloning of several these loci (for reviews see Marshall, 1991; Weinberg, 1991). Careful comparison of tumor DNA samples with normal DNA from the same individuals by means of restriction fragment length polymorphisms (RFLPs) have provided a particularly powerful alternative approach. Loss of constitutional heterozygosity detected by RFLPs indicated that loss of genetic information also occurred by cytogenetically invisible mechanisms. This technique has been successfully applied to numerous human neoplasms, and has led to an almost exponential accumulation of data on allele losses. In general terms, structural rearrangements that consistently juxtapose two different chromosomal regions are thought to contain dominantly acting proto-oncogenes. Deletions or monosomies are believed to be the site of recessive tumor suppressor genes.

Specific chromosomal translocations are the most striking chromosomal abnormalities in hematopoietic malignancies. The first consistent chromosomal aberration observed in human neoplasia was the Philadelphia (Ph^1) chromosome in chronic myeloid leukemia (CML) in 1960 (Nowell and Hungerford, 1960). The proof that the Ph^1 chromosome occurred as the result of a translocation, rather than a deletion, had to wait until the development of chromosome banding techniques became available in 1973 (Rowley, 1973a). It became possible to identify each human chromosome, and parts of chromosomes, precisely. Thus it was shown that the Ph^1 chromosome arises from a reciprocal translocation between the long arm (q) of chromosome 9, band 34, and band q11 of chromosome 22 [t(9;22)(q34;q11)]. This was one of two translocations reported in 1973; the other one was the t(8;21)(q22;q22) translocation actually identified about 6 months earlier than the Ph^1 translocation and occurred in patients with acute myeloid leukemia (AML) (Rowley, 1973b). These are the first consistent translocations to be discovered in any human or animal tumors. Up to now, a large number of recurring translocations in hematopoietic malignancies have been described, and many of these are closely associated with specific histologic or immunologic subtypes. Molecular studies of chromosomal translocations have led to the identification of new cellular proto-oncogenes located at or near the translocation breakpoints and have provided important insights into the mechanisms of oncogenesis.

Two general mechanisms are well known for the roles of translocation in activation of proto-oncogenes. The first involves deregulated expression by juxtaposition with the immunoglobulin (Ig) or T cell receptor (TCR) loci in lymphoid malignancies. This results from errors in the gene rearrangement process, thereby implicating the Ig/TCR recombinase and class switch enzymes as mediators of interchromosomal translocations. However, the extent and mechanistic details of their role remain to be determined. As a result of recombination errors, a variety of cellular genes have been juxtaposed with the Ig or TCR loci. The best studied examples are translocations involving the c-*MYC* and *BCL-2* genes. In the t(8;14), t(8;22), and t(2;8) translocations associated with Burkitt's lymphoma, the c-*MYC* gene is deregulated by the Ig heavy chain enhancer or light chain λ or κ enhancer (Dalla-Favera *et al.*, 1982; Taub *et al.*, 1982; for reviews see Croce and Nowell, 1985; Cory, 1986). The c-*MYC* gene encodes a DNA-binding protein containing a transcriptional activation domain, a basic DNA-binding region, a helix-loop-helix motif, and a leucine zipper dimerization motif. In follicular lymphoma, the t(14;18) translocation fuses the *BCL-2* gene to the Ig heavy chain enhancer, resulting in aberrantly increased *BCL-2* expression (Tsujimoto *et al.*, 1985; Bakhshi *et al.*, 1985; Cleary and Sklar, 1985). The *BCL-2* gene encodes a cytoplasmic protein (Tsujimoto *et al.*, 1987; Chen-Levy *et al.*, 1989) that seems to be associated with the inner mitochondrial membrane (Hockenberry *et al.*, 1990). The property of BCL-2 of preventing programmed cell death (apoptosis) (Hockenberry *et al.*, 1990) represents a new mechanism in the etiology of tumors. Transgenic mice carrying constructs that mimic these translocation rearrangements recreate a pathology similar to the human disease counterpart (for review see Adams and Cory, 1991). Walking from the Ig or TCR genes has identified more than ten similarly deregulated proto-oncogenes (for reviews see Rowley, 1990; Rabbitts, 1991; Solomon *et al.*, 1991).

The second general mechanism involves formation of fusion genes coding for chimeric proteins as a result of interchromosomal recombination occurred within introns of the genes. The t(9;22) translocation in the Ph[1] chromosome positive CML is the most extensively studied example for this type, generating the *BCR-ABL* fusion gene and the chimeric protein p210

(Heisterkamp *et al.*, 1983; Konopka *et al.*, 1984; Shtivelman *et al.*, 1985; for review see Kurzrock *et al.*, 1988). Interestingly, the Ph[1] positive acute lymphoblastic leukemia (ALL) also has the *BCR-ABL* but results in a smaller protein p190. Both chimeric proteins have an elevated tyrosine kinase activity (Konopka *et al.*, 1984; Lugo *et al.*, 1990) and are capable of inducing CML or ALL in transgenic mice (Daley et al., 1990; Heisterkamp *et al.*, 1990). Other examples of fusion gene by this mechanism are the *E2A-PBX1* fusion in the t(1;19) translocation observed in a significant percentage of childhood pre-B cell ALL (Nourse *et al.*, 1990; Kamps *et al.*, 1990), the *DEK-CAN* fusion in the t(6;9) translocation associated with a small subset of AML (von Lindern *et al.*, 1992), and the fusion of the retinoic acid receptor α gene with PML in the t(15;17) translocation in acute promyelocytic leukemia (APL) (Kakizuka *et al.*, 1991; de Thé *et al.*, 1991; Pandolfi *et al.*, 1991; Goddard *et al.*, 1991). To date, molecular studies of translocations have established an important paradigm whereby cellular genes involved in the regulation of proliferation and/or differentiation may be converted to oncogenic forms by certain chromosomal translocations altering their expression or structure. However, exactly how these translocated genes contribute to the pathogenesis of these neoplasms is still poorly understood.

The t(8;21)(q22;q22) translocation is one of the most frequent chromosomal abnormalities in AML and is morphologically associated with the M2 subtype according to the French-American-British (FAB) classification (Rowley, 1984). Leukemic cells with this translocation are uniquely characterized by a high frequency of Auer rods, maturation of the granulocytic line (Berger *et al.*, 1982), and occasionally form tumors called myeloblastomas (Abe *et al.*, 1986). Cytogenetically, this translocation is often accompanied by a loss of sex chromosome which is rarely observed in acute leukemias without t(8;21) (Sakurai *et al.*, 1974). Chromosomes 8 and 21 can participate in three-way translocations similar to those involving chromosomes 9 and 22 in CML (Rowley, 1982). In general, patients with t(8;21) show a relatively good response to chemotherapy and have durable remissions. The t(8;21) translocation is one of the first consistent translocations specifically associated with human diseases, but molecular characterization of this translocation has not been successful (Drabkin

et al., 1985; Sacchi *et al.*, 1986), owing to a lack of closely linked probes.

In this study, a *Not* I linking clone library was used to generate probes that allowed detection of the t(8;21) rearrangement on pulsed-field gel electrophoresis (PFGE) analysis. Subsequently, chromosome walking and cDNA screening led to the identification of the *AML1* gene located at the breakpoint region on chromosome 21. The breakpoints on chromosome 21 consistently occur within a single specific intron of *AML1*. Further analyses were carried out to isolate and characterize a fusion cDNA from a cell line with t(8;21), demonstrating that the 5' portion of the *AML1* gene is juxtaposed with the most part of the *MTG8* gene on chromosome 8 following the t(8;21) translocation. The present work establishes a new member of fusion gene involved in leukemogenesis.

Results

Detection of the t(8;21) rearrangement on PFGE analysis

Not I linking clones specific for chromosome 21, which are genomic fragments containing rare-cutter *Not* I restriction sites, were analyzed by regional mapping. Seven clones mapped in the region of 21q21 to 21q22.3 were tested on pulsed-field gel electrophoresis (PFGE) blots. A subclone (LL263L; a 3.2 kb *Not* I-*Hin*dIII fragment) of the linking clone LL263 detected an abnormal band of about 1,500 kb in *Not* I-digested leukemic cell DNAs from two AML patients with t(8;21) (Figure 1). The abnormal band was not detected in the bone marrow cell DNAs from same patients in remission (Figure 1), leukemic cell DNAs from other 13 AML patients without t(8;21) and lymphocyte DNAs from 12 normal individuals (data not shown). A partial *Not* I digest of normal lymphocyte DNA did not show the same size abnormal band. After complete digestion of the leukemic cell DNA with *Not* I, complete or partial digestion with *Bss*HII demonstrated band shifts from 1,500 kb to 1,100 kb and 900 kb to 270 kb, respectively (Figure 2). These findings indicate that the abnormal band was due to the t(8;21) translocation, not actual restriction fragment length polymorphism (RFLP), DNA methylation differences, or incomplete digestion.

Physical mapping of the t(8;21) breakpoint region

Using the somatic cell hybrids established by fusion of the leukemic cells from patient 2 and mouse myeloma X63 (HPRT$^-$) cells, the relationship between the LL263 location and the t(8;21) breakpoint was examined. The LL263L probe hybridized to DNA from hybrid cells containing a derivative chromosome 8 [der(8)], but not to DNA from hybrid cells containing a normal chromosome 8 or a derivative chromosome 21 [der(21)]. These findings show that LL263 is distal to the breakpoint on chromosome 21 (Figure 3).

LL263S (a 3.0 kb *Not* I-*Hin*dIII fragment) is the other subclone of the LL263 linking clone and detected a 1,300 kb *Not* I fragment adjacent to the 900 kb *Not* I fragment detected with the LL263L probe. A DNA marker D21S17 (pGSH8) (Stewart *et al.*, 1985) (supplied by the Japanese Cancer Research Resources Bank) was mapped in the same 1,300 kb fragment.

Further analysis of the t(8;21) breakpoint region was performed with lymphocyte DNA

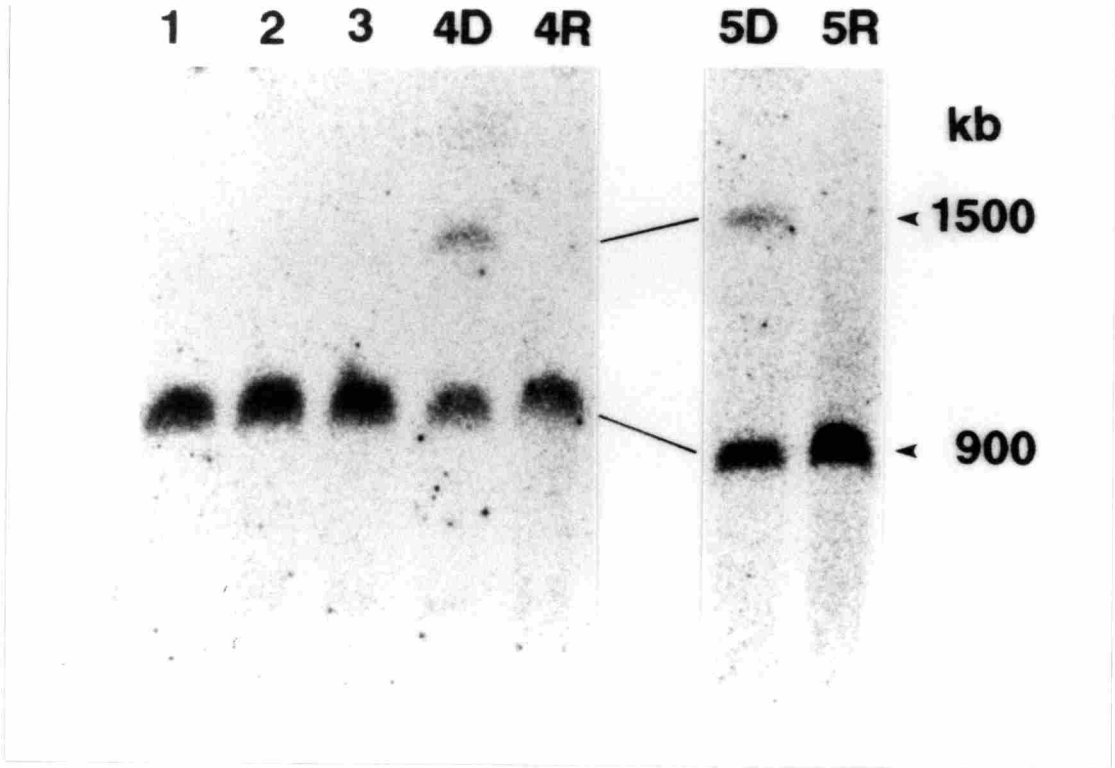


Figure 1. PFGE analysis of *Not* I-digested DNAs with the LL263L probe. Lanes 1-3, lymphocyte DNAs from three normal individuals; lanes 4D and 5D, leukemic cell DNAs from patients 1 and 2, respectively, obtained at the time of diagnosis; lanes 4R and 5R, bone marrow cell DNAs from patients 1 and 2, respectively, obtained at the time of remission. LL263L detected 900 kb germline band in all lanes and 1,500 kb abnormal band only in lanes 4D and 5D.

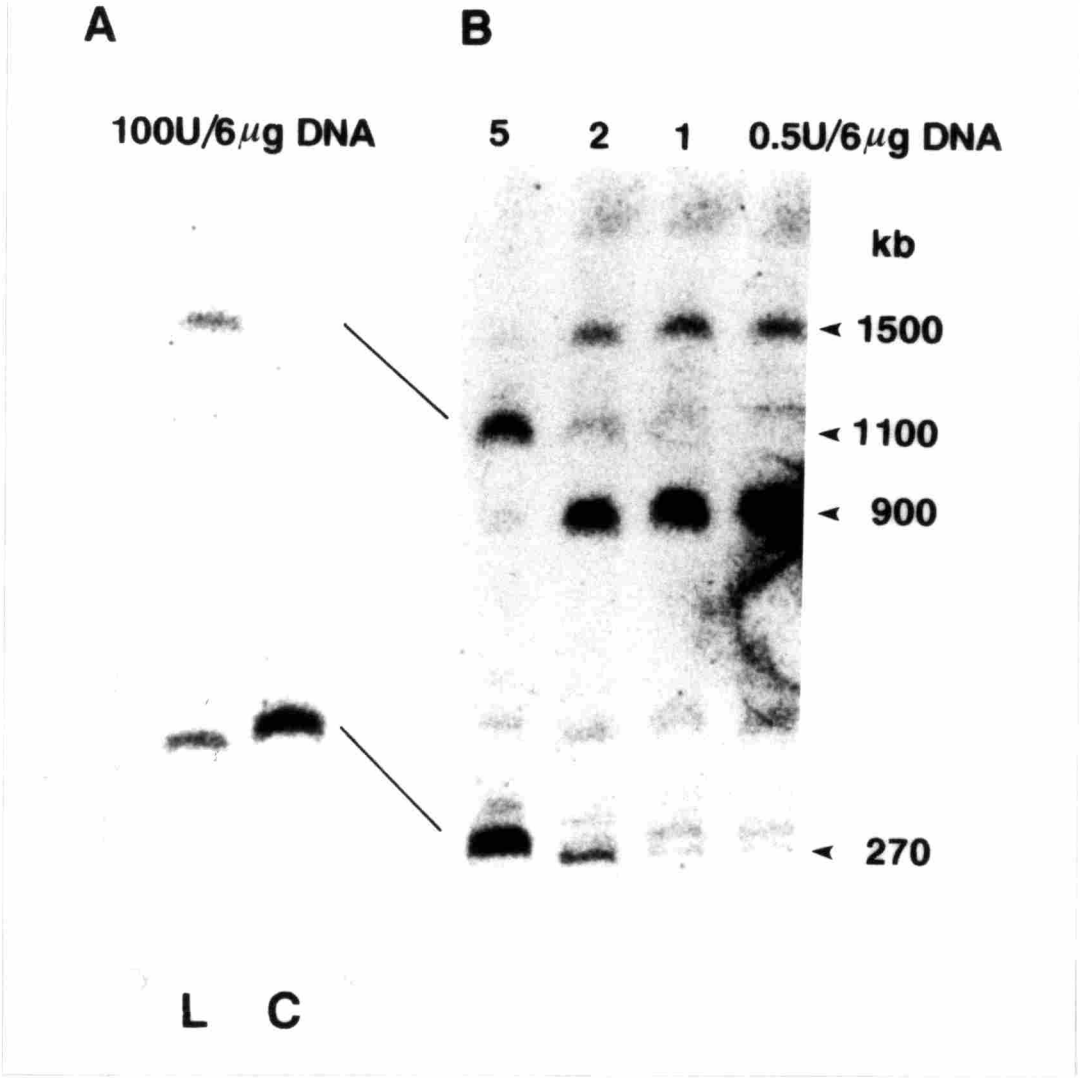


Figure 2. PFGE analysis of *Not* I and *Bss*HII doubly digested DNAs with the LL263L probe. DNA from normal lymphocytes and leukemic cell DNA from patient 2 were completely digested with *Not* I, and then completely (A) or partially (B) digested with *Bss*HII. (A) The LL263L probe detected a 270 kb band in the control DNA (C), and a 1,100 kb band as well as the 270 kb band in the leukemic cell DNA (L). (B) With gradually increasing units of *Bss*HII, the 900 kb and 1,500 kb bands shifted to the 270 kb and 1,100 kb bands, respectively.

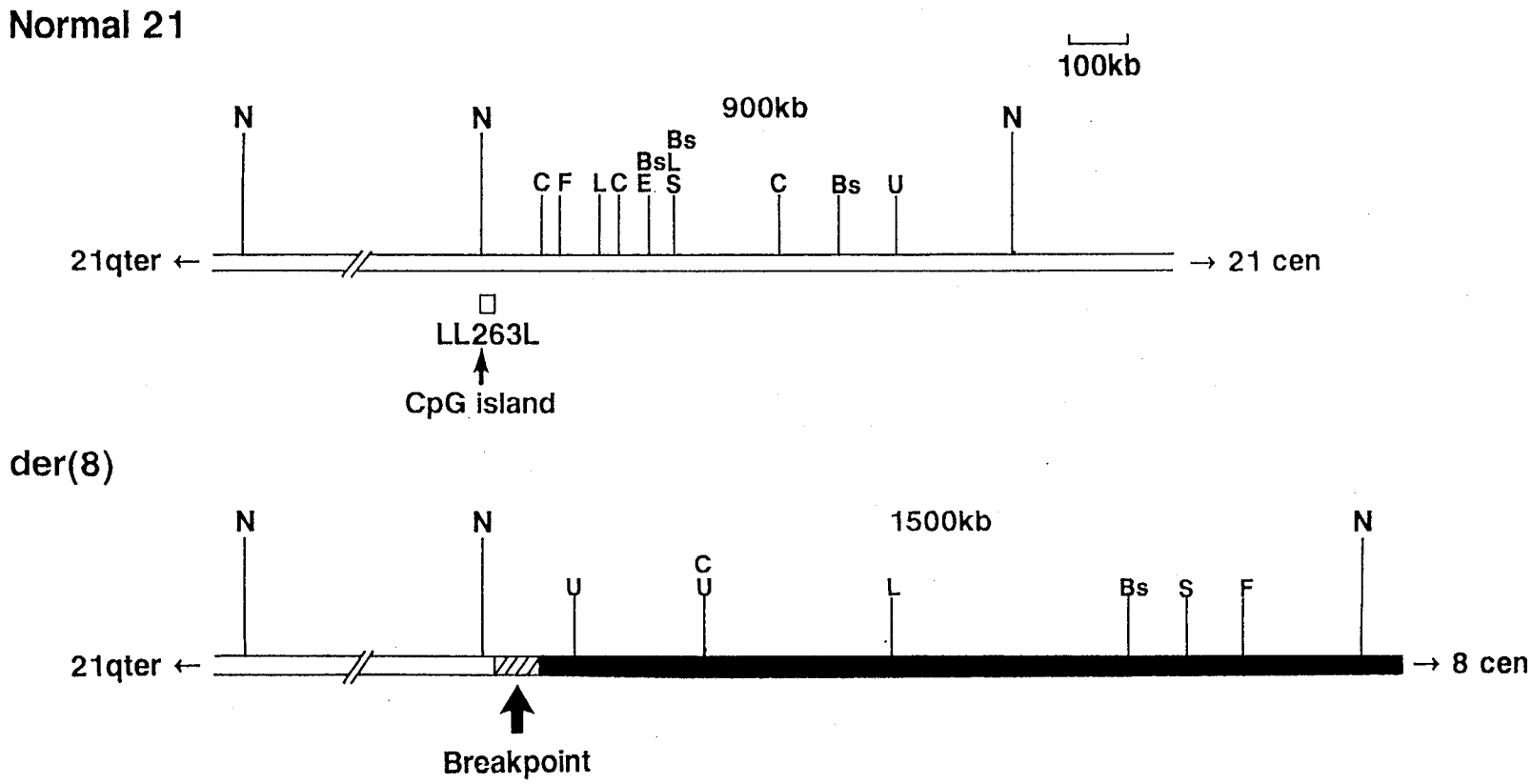


Figure 3. Pulsed-field map of the t(8;21) breakpoint region in normal chromosome 21 and the der(8) chromosome. Chromosomes 21 and 8 are shown by open and solid bars, respectively. The cross-hatched region represents the area where the t(8;21) breakpoint occurs. This map was constructed using the LL263L probe and DNA digested with rare-cutter enzymes on PFGE blots. Abbreviations for restriction sites are as follows: B, *Bam*HI; Bs, *Bss*HII; C, *Cla* I; E, *Eag* I; F, *Sfi* I; H, *Hin*dIII; L, *Sal* I; N, *Not* I; S, *Sac* II; U, *Nru* I; Xb, *Xba* I. No *Mlu* I sites were detected in the region.

from a normal individual and leukemic cell DNA from patient 2. The LL263L probe detected the same fragments in both samples doubly digested with *Not* I and *Xba* I, *Bam*HI, *Hin*dIII, or *Eco*RI (Figure 4A). The largest 13 kb fragment was detected in the *Not*I and *Bam*HI digested DNA. The probe detected a 100 kb fragment in the control DNA and a 390 kb fragment as well as the 100 kb fragment in the leukemic cell DNA after their digestion with *Not* I and *Cla* I (Figure 4B). Thus the t(8;21) breakpoint in chromosome 21 was located about 13 kb to 100 kb proximal to the LL263 *Not* I site (Figure 3).

The 6.7 kb *Hin*dIII fragment, from which LL263 was derived, had two *Not* I restriction sites with a distance of 0.5 kb. LL263L was digested by many restriction enzymes recognizing CpG sequences including *Bss*HII, *Eag* I, *Nae* I, *Nar* I, *Sac* II, and *Sma* I. This finding indicates the presence of a CpG island in the fragment.

Cloning of cDNA located at the breakpoint region on chromosome 21

The t(8;21) breakpoint was close to the LL263 *Not* I site. In addition, LL263 contains a CpG island that is present in the 5' regions of most housekeeping genes (Bird, 1987), suggesting that LL263 itself is a part of candidate gene. Accordingly, a human bone marrow cDNA library was screened with the LL263 clone, but no cDNA clones were isolated. Then, chromosome walking was performed from LL263L and three genomic clones (λE3, λE4, and λE12) were isolated (Figure 5). Genomic clone λE4 insert was used to rescreen the cDNA library, and cDNA clone C6 was identified (Figure 5).

Leukemic cell DNAs from t(8;21) AML patients digested with *Bam*HI were examined by Southern blot analysis with a 0.24 kb *Hin*dIII-*Eco*RI fragment (C6E6H2) of clone C6 as a probe. Rearranged *Bam*HI fragments were detected in leukemic DNAs from four patients but not in their DNAs in remission (Figure 6A). This finding showed that the translocation breakpoints in the four patients occurred within the *Bam*HI fragments which were recognized by the cDNA probe (C6E6H2). Comparison of the intensities of the germline bands in DNAs of leukemic cell and of cells in remission suggested that the rearranged DNA fragment of patient 2 was derived from a germline *Bam*HI fragment of 11 kb, while those of patients 1, 3, and 4 were from a fragment of 19 kb.

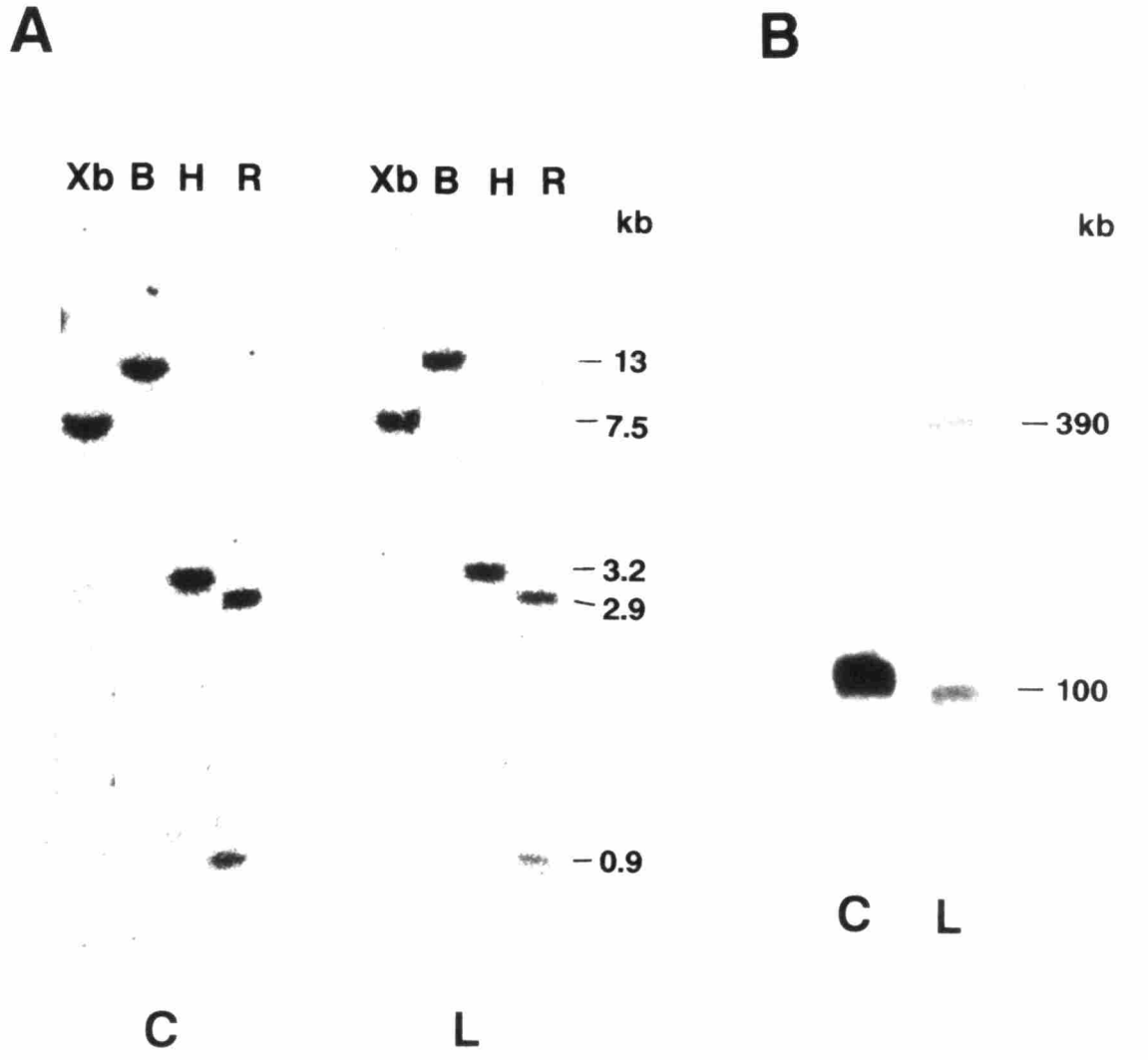


Figure 4. (A) Southern blot analysis of doubly digested DNAs with *Not* I and *Xba* I, *Bam*HI, *Hin*dIII, or *Eco*RI. About 4 μg of leukemic cell DNA of patient 2 (L), and control lymphocyte DNA of normal individual (C) in each agarose plug was completely digested with the restriction enzymes, separated on 1% agarose gel, and hybridized with the LL263L probe. No abnormal bands were detected in the leukemic cell DNA and the control DNA. (B) PFGE analysis of *Not* I and *Cla* I doubly digested DNA with the LL263L probe. LL263L detected only 100 kb germline band in the control DNA (C), but also 390 kb abnormal band in the leukemic cell DNA (L).

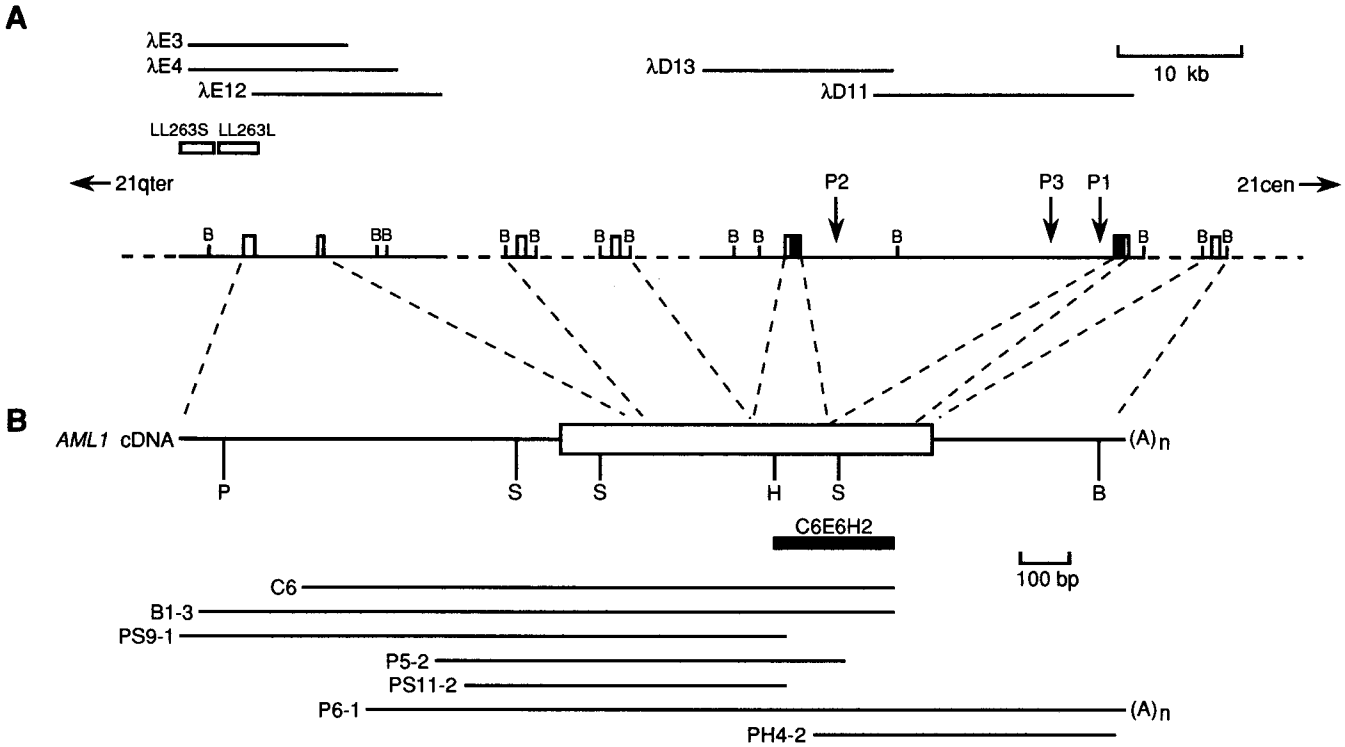


Figure 5. Restriction maps of the t(8;21) breakpoint region on chromosome 21 and the *AML1* cDNAs. (A) Genomic clones λE3, λE4, λE12, λD11 and λD13 are represented by horizontal bars above the map. LL263L and LL263S are fragments of the linking clone LL263 digested with *Not* I. The partial restriction map of *Bam*HI was derived from Southern blot analysis of genomic DNA and genomic clones using several restriction fragments of the cDNA clones as probes. Exons of the *AML1* (not shown to scale) are represented by boxes; solid boxes correspond to the C6E6H2 fragment. The breakpoints in patients 1, 2, and 3 (P1, P2, and P3) are indicated by vertical arrows. (B) Open box represents the open reading frame deduced from the sequence of cDNA clones indicated below. The cDNA clone P6-1 contains a poly(A) tract of >90 nucleotides, and several other overlapping cDNA clones with distinct sequences in the 3' region were also isolated (not shown). Dashed lines show the approximate locations of exons in genomic DNA. C6E6H2, a *Hin*dIII-*Eco*RI fragment of C6 clone, is indicated by a solid bar. B, *Bam*HI; H, *Hin*d III; P, *Pst* I; S, *Sma* I; bp, base pairs; kb, kilobases.

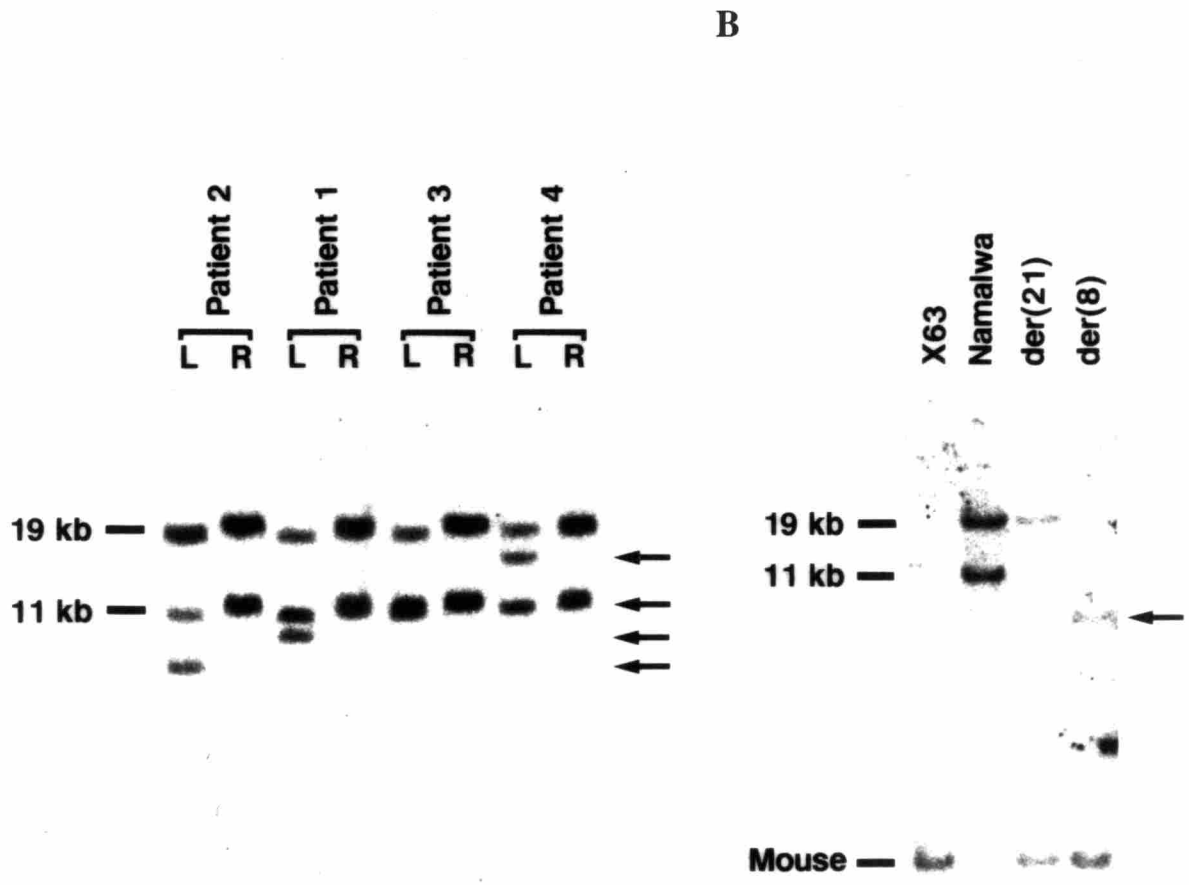


Figure 6. Detection of the t(8;21) breakpoints by Southern blot analysis of *Bam*HI-digested DNAs with cDNA probe C6E6H2. (A) DNAs from leukemic cells (L) and cells in remission (R) from four patients are compared. Rearranged bands are indicated by arrows. In DNAs obtained in remission, only germline fragments of 11 kb and 19 kb were detected. The rearranged band in patient 3 overlapped the 11 kb germline band. (B) Mouse-human somatic cell hybrid DNA containing the der(8) or der(21) chromosome of patient 2. X63 (mouse myeloma parental cell line) and Namalwa (Burkitt's lymphoma cell line) were used as controls for mouse and human DNAs, respectively.

The chromosomal breakpoint of patient 2 was subsequently confirmed by Southern blotting of mouse-human somatic cell hybrid DNA containing only the der(8) or der(21) chromosome from leukemic cells of patient 2 (Figure 6B). Both the hybrid containing the der(8) and leukemic cell DNAs of patient 2 shared the same rearranged band, whereas the hybrid containing the der(8) lacked the 11 kb band, indicating that the breakpoint in patient 2 occurred in the 11 kb *Bam*HI fragment. Furthermore, detection of the 19 kb *Bam*HI fragment in the hybrid containing the der(21) chromosome indicated that this 19 kb fragment was located centromeric to the 11 kb fragment. The localizations of breakpoints in patients 1 and 3 were confirmed by use of genomic clone λD11, as described later.

Nucleotide sequence and predicted amino acid sequence of *AML1*

With the C6E6H2 probe, human bone marrow and peripheral blood leukocyte cDNA libraries were screened, and several overlapping cDNA clones were isolated (Figure 5). The insert of clone P6-1 contained the poly(A) sequence. Sequencing of these cDNA clones revealed a long open reading frame encoding a protein of 250 amino acids (Figure 7). The nucleotide sequence surrounding the predicted initiation codon agreed well with the Kozak consensus sequence (Kozak, 1987). The predicted amino acid sequence showed no significant homology to any known protein sequences in the SWISS-PROT database. Thus, this novel gene was named *AML1*. Afterward, two genes highly homologous to *AML1* have been reported (see Discussion and Figure 18).

The predicted *AML1* gene product is rich in proline (9.6%) and arginine (9.2%). An ATP- or GTP-binding site motif A (Moller and Amons, 1985) is found at amino acid residues 138-145 by using the PC/GENE PROSITE program. The 5' untranslated region of *AML1* is very long (more than 700 nucleotides) and similar in length to that of most proto-oncogenes, in contrast with the 5' untranslated region of most vertebrate mRNAs, which fall in the size range of 20 to 100 nucleotides (Kozak, 1987). The 3' untranslated region contains an A+U-rich sequence involving the AUUUA motif, which has been shown to mediate selective degradation of mRNA (Shaw and Kamen, 1986).

With several fragments of cDNA clones as probes, Southern blot analysis of the

Figure 7. Nucleotide sequence of *AML1* cDNA and deduced amino acid sequence (single-letter code). Numberings of nucleotides and amino acids are shown at right. The underlined amino acid sequence represents an ATP- or GTP-binding site motif A. The 3' untranslated region contains the AUUUA motif (underlined) implicated in selective destabilization of mRNA. The polyadenylylation signal, AATAAA, is boxed. An arrow indicates the exon boundary that was interrupted by the t(8;21) translocation.

hybrids demonstrated the orientation of *AML1* on chromosome 21 as 5' telomeric and 3' centromeric (see Figure 5).

The t(8;21) breakpoints on chromosome 21 cluster within a specific intron of *AML1*

By screening a human lymphocyte genomic library with C6E6H2 probe, two overlapping clones (λD11 and λD13) were obtained. Southern blot analysis using various subfragments of these genomic clones as probes identified the locations of breakpoints in patients 1, 2, and 3 on chromosome 21 (Figure 5). Since the *AML1* cDNA hybridized to six genomic *Bam*HI fragments, and one of them covered by λE4 contained at least two exons, the *AML1* cDNA contained at least seven exons. Exon mapping on λD11 and λD13 with the C6E6H2 probe and subsequent sequencing of intron/exon boundaries of the cloned DNA revealed that each of the two clones contained one exon. Consequently, the breakpoints of patients 1, 2, and 3 occurred within the same intron between these two exons.

To determine whether rearrangements constantly occurred in the single specific intron of the *AML1* gene, the breakpoints in 21 AML patients with t(8;21) including three with complex t(8;V;21) translocations, t(8;4;21), t(8;12;21) and t(8;20;21) (Maseki *et al.*, 1993), and an AML cell line with t(8;21), Kasumi-1 (Asou *et al.*, 1991), were analyzed. Figure 8 shows Southern blot analysis on *Bam*HI-digested DNAs with the *AML1* cDNA probe C6E6H2. In 18 of 22 cases (21 patients and Kasumi-1), abnormal rearranged bands were detected in addition to germline bands of 11 kb and 19 kb. No rearrangements in 4 cases (patients HMO, FO, MU, and KK) detected with the C6E6H2 probe may be due to comigration of the rearranged bands with the germline ones. The 5' and 3' segments of the exonic probe C6E6H2 were both located close to the external ends of the region covered by the 11 kb and 19 kb *Bam*HI fragments. On the other hand, the intronic probe D11X2, which originated from the internal boundary region of both *Bam*HI fragments, revealed rearranged bands in 18 cases (data not shown) including three (patients FO, MU, and KK) of four cases mentioned above. Thus, using the C6E6H2 and D11X2 probes in *Bam*HI digestion, *AML1* rearrangements were detected in 21 of 22 cases. In 15 cases (patients KH, TR, IMK, IK, IT,

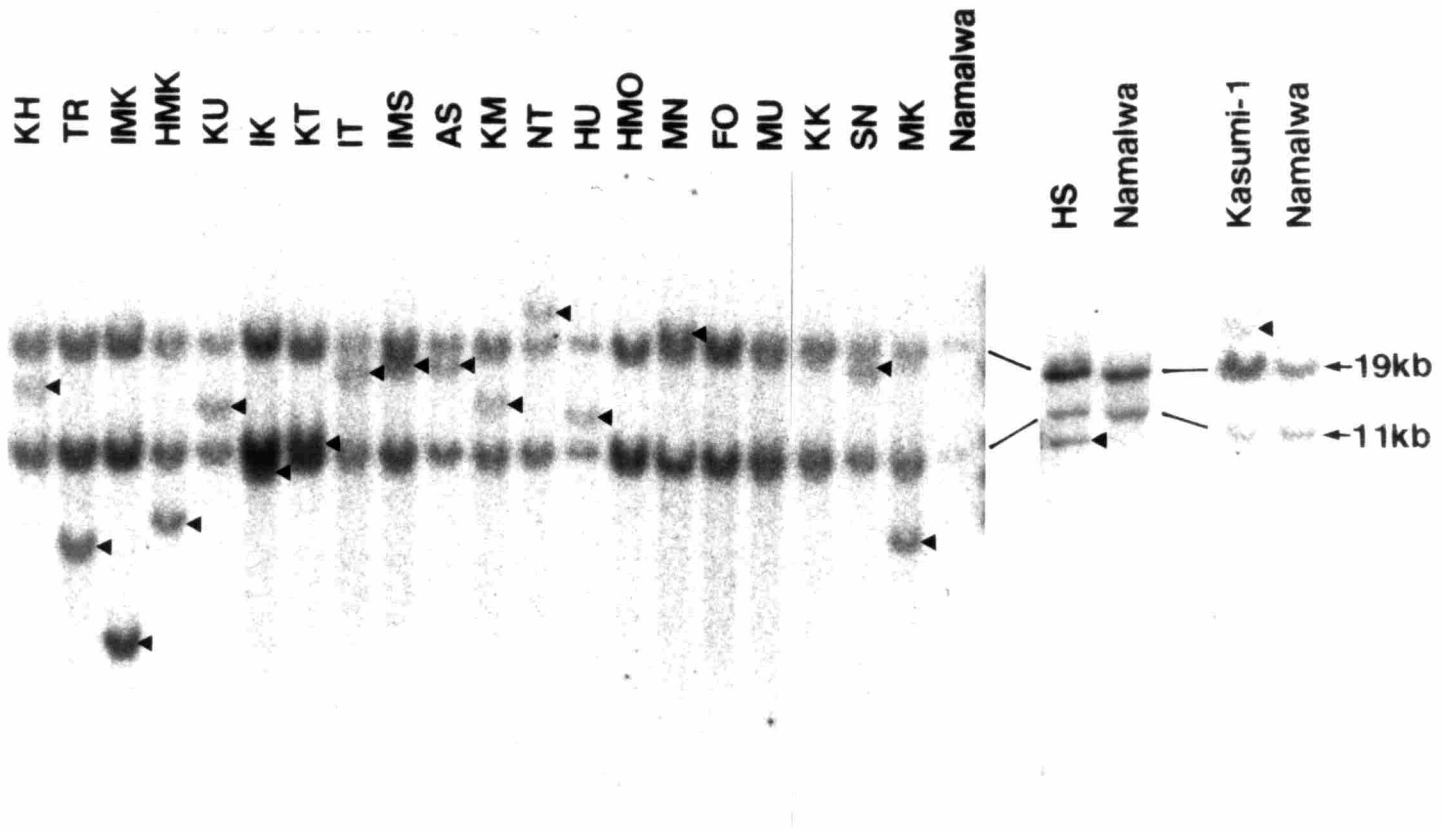


Figure 8. Southern blot analysis of leukemic cell DNAs with the C6E6H2 probe. Leukemic cell DNAs from 21 AML patients with t(8;21) or t(8;V;21) and Kasumi-1 were digested with *Bam*HI and fractionated by field inversion gel electrophoresis on 1.0% agarose gel. Patients KK, SN, and MK have complex translocations, t(8;4;21), t(8;12;21), and t(8;20;21), respectively. Namalwa (Burkitt's lymphoma cell line) cells were used as normal control. Rearranged bands are indicated by arrowheads.

NT, AS, KT, HMK, KU, HU, IMS, KM, SN, and MK), both C6E6H2 and D11X2 detected rearranged bands of different sizes, which were derived from the der(8) and der (21) chromosomes, indicating that the breakpoints occurred within the same intron. Patient MK was first diagnosed as having t(8;20) by cytogenetic analysis. However, after detection of the *AML1* gene rearrangement, reexamination clarified that patient MK had the complex translocation t(8;20;21).

Subsequently, detailed mapping of the breakpoints in 22 cases was performed by Southern blot analysis using several restriction enzymes and multiple genomic probes derived from λD11 and λD13 genomic clones. Representative Southern blots are shown in Figure 9. In patient FO, the D11X1 probe detected two rearranged *Eco*RI bands of 1.8 kb and 6.0 kb in addition to the germline *Eco*RI band of 7.3 kb (Figure 9A), indicating that the breakpoint on chromosome 21 occurred in the 7.3 kb *Eco*RI fragment. In patient HMO, the D11X3 and D11X4 probes detected two rearranged *Hin*dIII bands of 3.5 kb and 6.2 kb that derived from the der(8) and der(21) chromosome, respectively (Figure 9B). Therefore, the breakpoint in patient HMO occurred in the region indicated in Figure 10. The locations of the breakpoints in 25 cases including those in three patients (patients 1, 2, and 3) are summarized in Figure 10. The rearrangements were identified in more than two digests with different restriction enzymes in most cases (22 of 25), and so the abnormal bands were probably not due to RFLP. In three cases (patients MU, NT, and HMO), the rearrangements were analyzed with a single enzyme: *Bam*HI for patients MU and NT, *Hin*dIII for patient HMO. However, thus far as examined, no RFLP sites for *Bam*HI and *Hin*dIII have been detected in this region. The results shows that all breakpoints occurred within the same 25 kb intron of the *AML1* gene with no specific localization of breaks; therefore the existence of translocation hot spot seems unlikely. Considering the clustering of breakpoints in the single specific intron, the *AML1* probes used here should be useful for the diagnosis and monitoring of this type of leukemia.

Cloning and sequencing of the breakpoint region

The breakpoint regions of the der (8) and der (21) chromosomes as well as the corresponding germline regions from patient KH were cloned and sequenced (Figure 11). Comparison of the

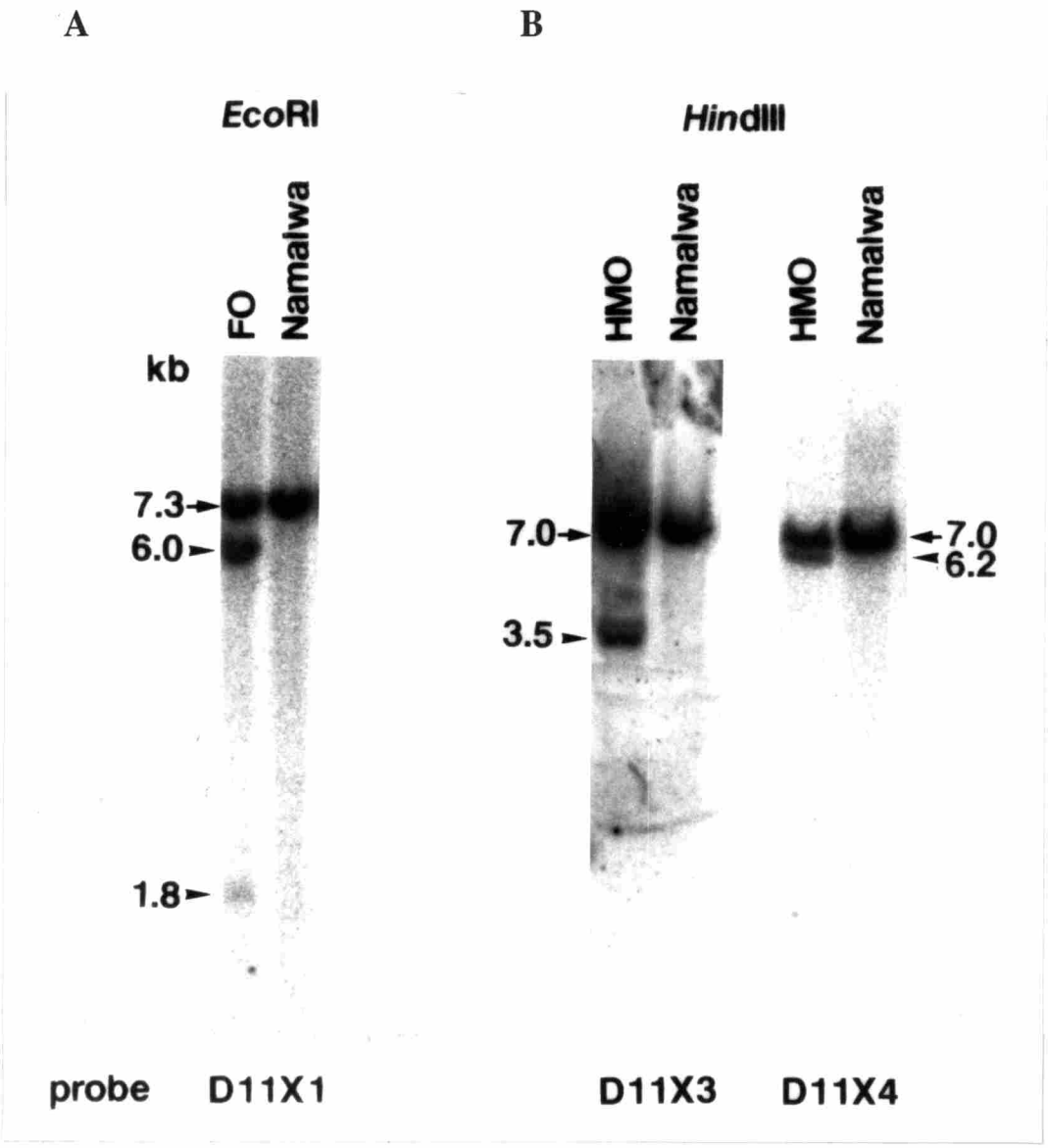


Figure 9. Southern blot analysis of leukemic cell DNAs. (A) Leukemic cell DNA from patient FO was digested with *Eco*RI and hybridized to the D11X1 probe. (B) Leukemic cell DNA from Patient HMO was digested with *Hin*dIII and hybridized to the D11X3 or D11X4 probe. Arrows and arrowheads indicate germline and rearranged bands, respectively.

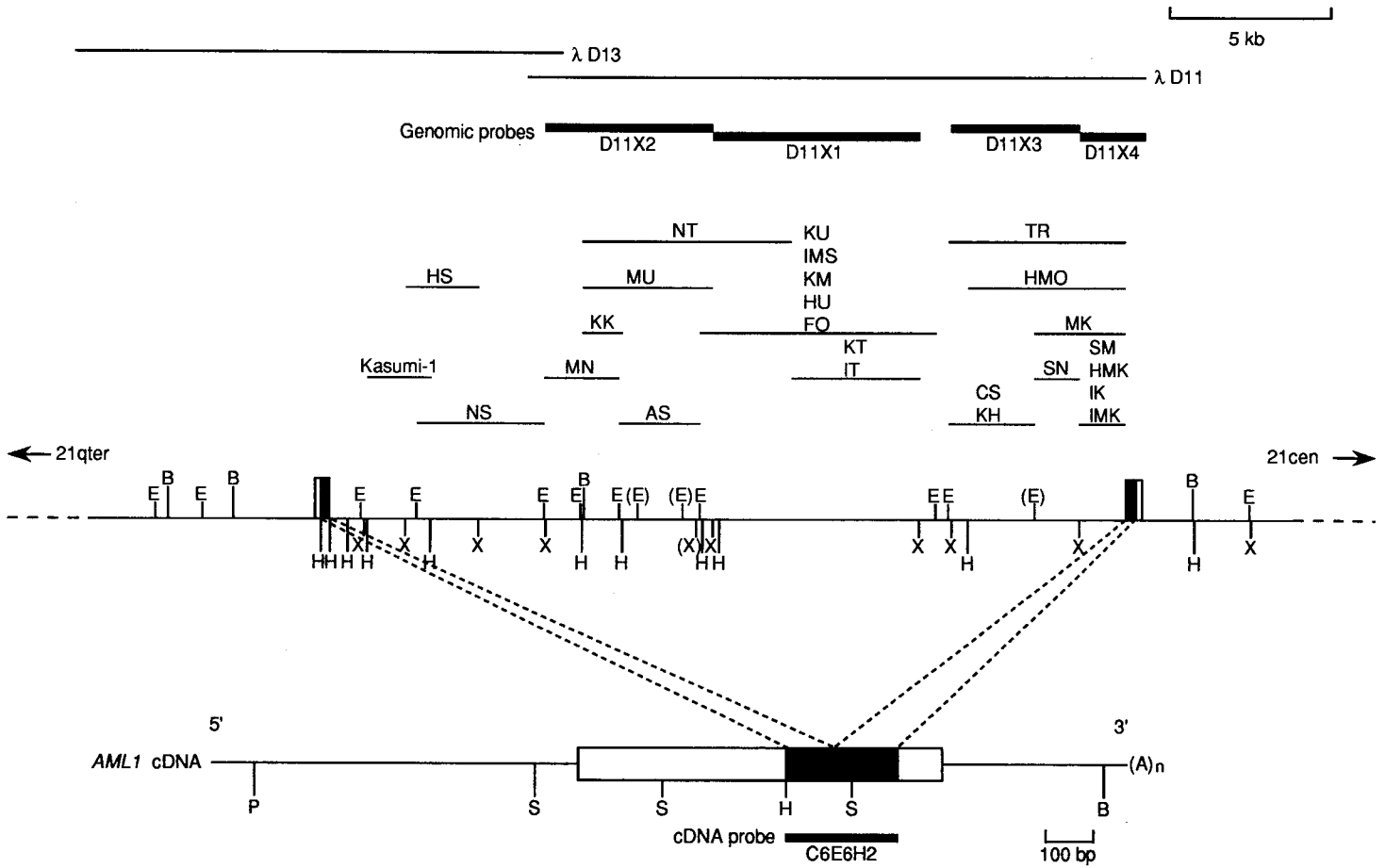


Figure 10. Locations of the t(8;21) translocation breakpoints and restriction map of the *AML1* locus with relation to the *AML1* cDNA. Patients SM, NS, and CS are identical to patients 1, 2, and 3, respectively. Boxes on the genomic map represent exons and shaded parts in boxes correspond to the *AML1* cDNA probe C6E6H2. Box in the cDNA represents the open reading frame. Lines above the map indicate the regions in which the breakpoints occurred. 5' end of the region for patients KT and IT is the *Spe*I site. Bold horizontal lines represent the genomic probes. RFLP sites are shown in parentheses. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sma*I; P, *Pst*I; X, *Xba*I.

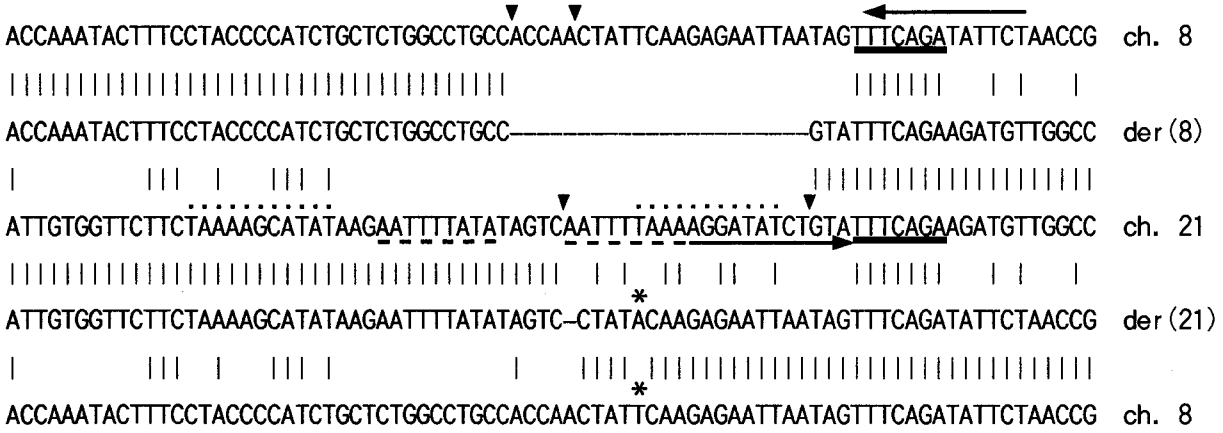


Figure 11. Nucleotide sequences at the breakpoint region in patient KH. Bold lines represent the homologous sequence on chromosome 8 and chromosome 21. Arrows represent the inverted homologous sequences. Dots and dashed lines represent the repeated sequences on chromosome 21. Chromosomal breakage occurred in the regions between the arrowheads on both normal chromosomes. Asterisks indicate nucleotide polymorphism.

rearranged sequences with that for germline showed that the small deletions of 18 bases of chromosome 21 and 5 bases of chromosome 8 were accompanied in chromosomal breakage and rejoining. Chromosomal breakage occurred in the regions between the arrowheads on both chromosomes shown in Figure 11. Some small homologous regions and direct repeats were noticed around the recombination site, however, no tandem repeats as often seen in chromosomal translocations were found (Heisterkamp *et al.*, 1985; Bakhshi *et al.*, 1987; McKeithan *et al.*, 1990).

Cloning of fusion cDNAs from an AML cell line with t(8;21)

Cytogenetic studies on complex translocations have indicated that the der(8) chromosome is the critical constant rearrangement (Rowley, 1982; Maseki *et al.*, 1993). In addition, considering the orientation of the *AML1* gene and the conservation of breakpoints in the standard t(8;21) and complex translocations within the single specific intron of *AML1*, it was suggested that the t(8;21) translocation, in the breakpoint region of the der(8) chromosome, fused a common 5' portion of the *AML1* gene to a presumable counterpart gene on chromosome 8.

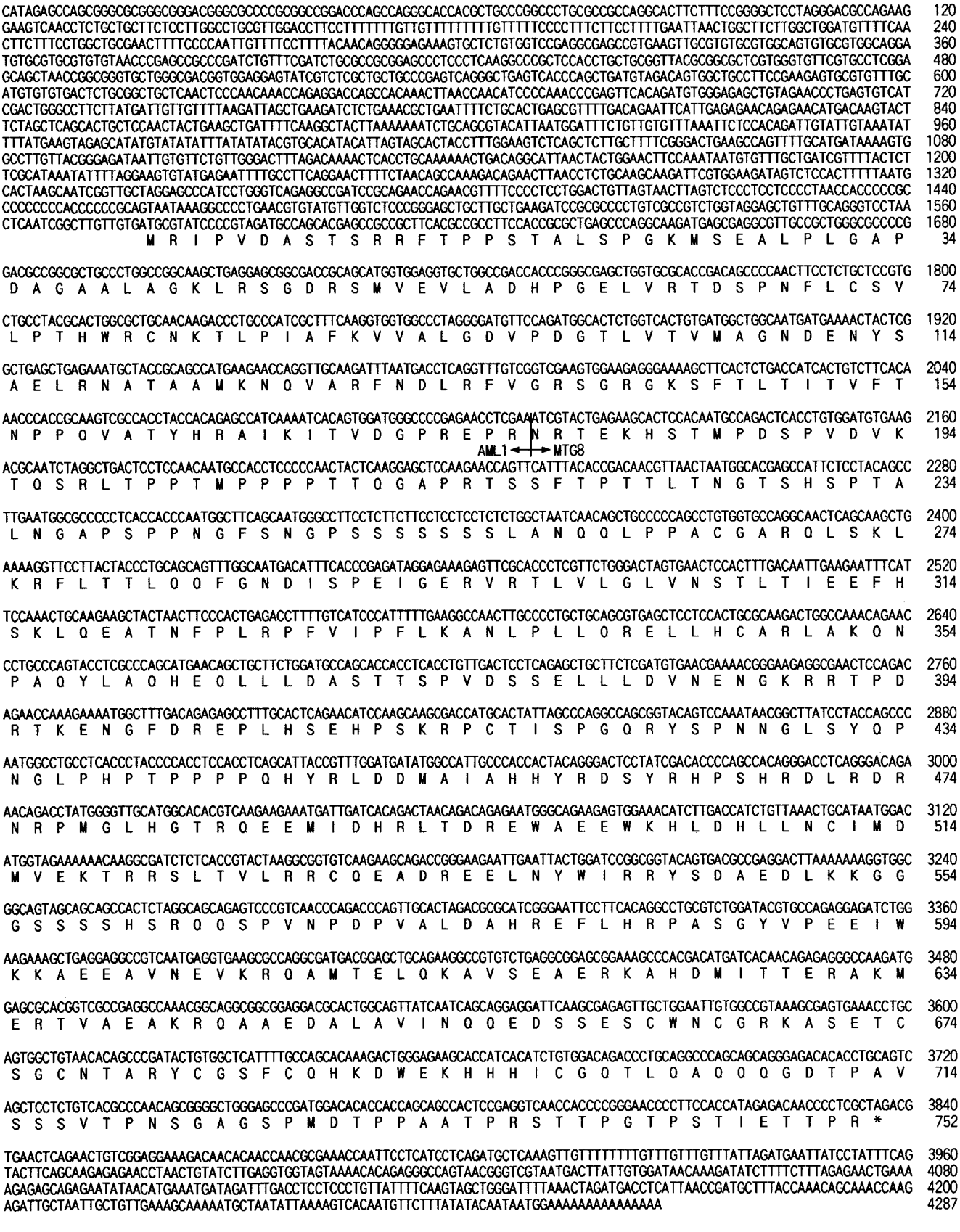
To amplify the fused cDNA sequences from an AML cell line with t(8;21), Kasumi-1, the polymerase chain reaction (PCR) was performed with an oligo(dT) primer and *AML1*-specific primer, resulting in the isolation of a part of the fusion gene (Kozu *et al.*, 1993). A 0.4 kb *Hin*cII fragment (CH15H2S, identical to nucleotides 2,247-2,701 in Figure 12A) of the PCR product was shown to be derived from chromosome 8. This was confirmed by hybridization of this fragment to Southern blots containing DNAs from a panel of human-hamster cell hybrids and a mouse-human somatic cell hybrid line containing human chromosome 8 (data not shown). A cDNA library was constructed from the Kasumi-1 cell line and screened with the *AML1*-specific probe and the CH15H2S probe. Numerous positive clones were obtained and five of these cDNA clones hybridizing with both the probes were finally isolated (see Materials and methods). Since the library was not amplified, each isolate represented a unique cDNA. Restriction enzyme mapping and partial sequence analyses revealed that the overlapping regions of these cDNA clones were identical and that four of them contained a poly(A) tract. The longest cDNA clone (K1) was selected for following sequence

analysis.

The fusion cDNA sequence shows an in-frame joining of the 5' portion of the *AML1* gene and a counterpart gene on chromosome 8, as could be expected (Figure 12A). The 5' portion of this sequence was 810 nucleotides longer than the *AML1* sequence in Figure 7 and was identical to the *AML1* up to nucleotide 2110. This nucleotide position indicates the exon boundary in the *AML1* gene interrupted by the t(8;21) translocation. The deduced amino acid sequence of the fusion cDNA consists of 752 amino acids with a relative molecular mass of 83,174. The predicted amino acid sequence derived from chromosome 8 (amino acids 179-752) showed no significant homology to any known protein sequences in the SWISS-PROT (August, 1992) and Protein Identification Resource (June, 1992) databases, and hence this novel gene was named *MTG8* (myeloid translocation gene on chromosome 8).

The C-terminal region (amino acids 663-699) of the predicted MTG8 protein contains two putative zinc finger DNA-binding motifs: one resembling a Cys/Cys-Cys/Cys class of zinc finger typified by the glucocorticoid receptor family, and the other resembling a Cys/Cys-His/Cys class found in the retroviral nucleic acid binding proteins (for reviews see Evans and Hollenberg, 1988; Berg, 1990). These two potential zinc fingers are most similar to that of a cell death-associated protein RP8 (Owens *et al.*, 1991) with 41% amino acid identity (Figure 12B), although the amino acid similarity does not extend beyond this region. In addition, the C-terminal region (amino acids 738-751) carries a PEST region (PEST score, 16.2 according to the PC/GENE PESTFIND program) that confers rapid intracellular degradation of protein (Rogers *et al.*, 1986). Another notable feature of the MTG8 protein is its unusually high proline content in three regions: amino acids 187-264, 405-445 and 712-751 containing 22%, 29% and 23% proline, respectively. A proline-rich region of CTF/NF-I has been identified as a transcriptional activation domain (Mermod *et al.*, 1989), and such regions have also been noted in many other transcription factors. Two of the three proline-rich regions (amino acids 187-264 and 712-751) are also rich in serine and threonine. The serine/threonine-rich segments might be sites of phosphorylation, which would potentially play a role in regulation of the MTG8 protein.

A



B

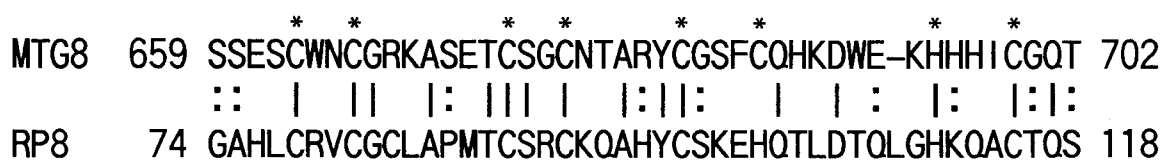


Figure 12. (A) Nucleotide and deduced amino acid sequences of the *AML1-MTG8* fusion cDNA. The fusion point is indicated by a vertical line. The sequence data reported here have been deposited in the DDBJ, EMBL and GenBank sequence databases under the accession number D13979. (B) Comparison of amino acid sequence of MTG8 with cell death-associated protein RP8. Vertical lines indicate identical amino acids. Double dots indicate conservative amino acid substitutions. Cysteines or histidines that are implicated in formation of zinc fingers are marked by asterisks.

Cloning and characterization of the wild-type *MTG8* cDNAs

The wild-type *MTG8* cDNA clones were isolated from a human fetal brain cDNA library using the CH15H2S probe since the *MTG8* gene was expressed in mouse brain (see Figure 17). Nucleotide sequence analysis of eight overlapping clones identified two types of composite cDNA sequence, named *MTG8a* and *MTG8b*. The sequence of *MTG8* portion of the *AML1-MTG8* fusion cDNA was conserved in both types. As shown in Figure 13, their sequences upstream of the fusion point were divergent each other. Each of the types was determined by sequencing at least two independent clones. Both *MTG8a* and *MTG8b* types were also isolated from Raji (Burkitt's lymphoma) cell line cDNA library and from HEL (erythroleukemia) cell line using anchored PCR, respectively, noting that the *MTG8* gene was expressed in both cell lines (see Figure 16). Accordingly these two types of *MTG8* cDNAs are unlikely cloning artifacts but alternatively spliced forms. In fact, several alternative 5' cDNA clones were isolated. These results suggest that the fusion site with *AML1* is an alternative splice acceptor site in the *MTG8* transcript. The predicted open reading frames of *MTG8a* and *MTG8b* cDNAs code for 577 and 604 amino acid proteins, respectively. An upstream in-frame termination codon was identified in both types. However, the second common ATG codon located immediately downstream of the fusion point may in fact act as a translation initiation site, because the sequence surrounding the second ATG is more favorable for the Kozak consensus sequence (Kozak, 1987). In either case, the *AML1-MTG8* fusion gene encodes a chimeric protein which retains the *runt*-homology region of AML1 protein and most of the MTG8 protein. The characteristics of AML1, MTG8, and AML1-MTG8 proteins are schematically shown in Figure 14.

To confirm the rearrangements in *MTG8* gene, Southern blot analysis of leukemic cell DNAs from 17 AML patients with t(8;21) was performed with the *MTG8* cDNA probes. An *MTG8a*-specific probe (R4H2RN) detected the rearranged *Eco*RI bands in nine cases including one complex t(8;4;21) translocation (Figure 15). However, the CH15H2S probe was not able to detect any rearrangements with *Eco*RI nor *Bam*HI digestion. These results suggest that the breakpoints on chromosome 8 are clustered within a limited region, probably at the 5' end of

MTG8a

```
TGAATAAATTATTCGGCATTTAGCTTATCATTCTGAATTTCACTTTTTGCTTTTTGGTGCTCTGAAACTTGCAGAGAGAGGTGGAGTGCT    90
AAGGGGAGGGTAGGGGTTGTGATACTTTGCACACACATCCCTGTCATTGTTCTGCCTAAAGAGACAGGGCTGGGTTCAAGGCCACATGTG   180
CTCCTGTCATCCTCCACATTTCTGCTCCAAGTGCAATCCGGAGTGTCAGCTCTCCATCTGTCTCTGCCTGGCAGGCGCACGCGCCCAGCA   270
CCCTGCCTCCGGCGATGCCGCCCCAGCCCCTCTGATGGCCCTCCTCTCTGCTGCCACTCATTCCAGAACAGGAGGCATGAGCCCGGAACG   360
CGCTTGCTTTTAGGAGACAGCCACTTTCTGTGTGGTACGCTGGATTCAAGGATGCCTGATCGTACTGAGAAGCACTCCACAATGCCAGAC   450
                                                    M  P  D  R  T  E  K  H  S  T  M  P  D      13
```

MTG8b

```
                                                                          AAAGATTGATCTCTGGGCTGGTGA    24
ACATAATCTCTGTCCCAGTCAGAAAAGGAGAGAGGAAATTAGCAGAGCGATTGGTGGAGAATGATATCTGTCAAAAGAAACACTTGGAGA   114
                                                                  M  I  S  V  K  R  N  T  W  R     10
GCACTGAGTTTAGTAATAGGTGACTGCCGGAAAAAAGGGAACTTTGAATATTGTCAAGATCGTACTGAGAAGCACTCCACAATGCCAGAC   194
 A  L  S  L  V  I  G  D  C  R  K  K  G  N  F  E  Y  C  Q  D  R  T  E  K  H  S  T  M  P  D     40
```

Figure 13. Partial nucleotide and deduced amino acid sequences of the *MTG8a* and *MTG8b* cDNAs. The in-frame termination codons preceding the putative initiation codon are underlined. Arrows indicate the point of fusion with *AML1*. The sequence downstream of the fusion point is identical to that of the *AML1-MTG8* fusion cDNA shown in Figure 12A. The sequence data reported here have been deposited in the DDBJ, EMBL and GenBank sequence databases under the accession numbers D14820 for *MTG8a* and D14821 for *MTG8b*.

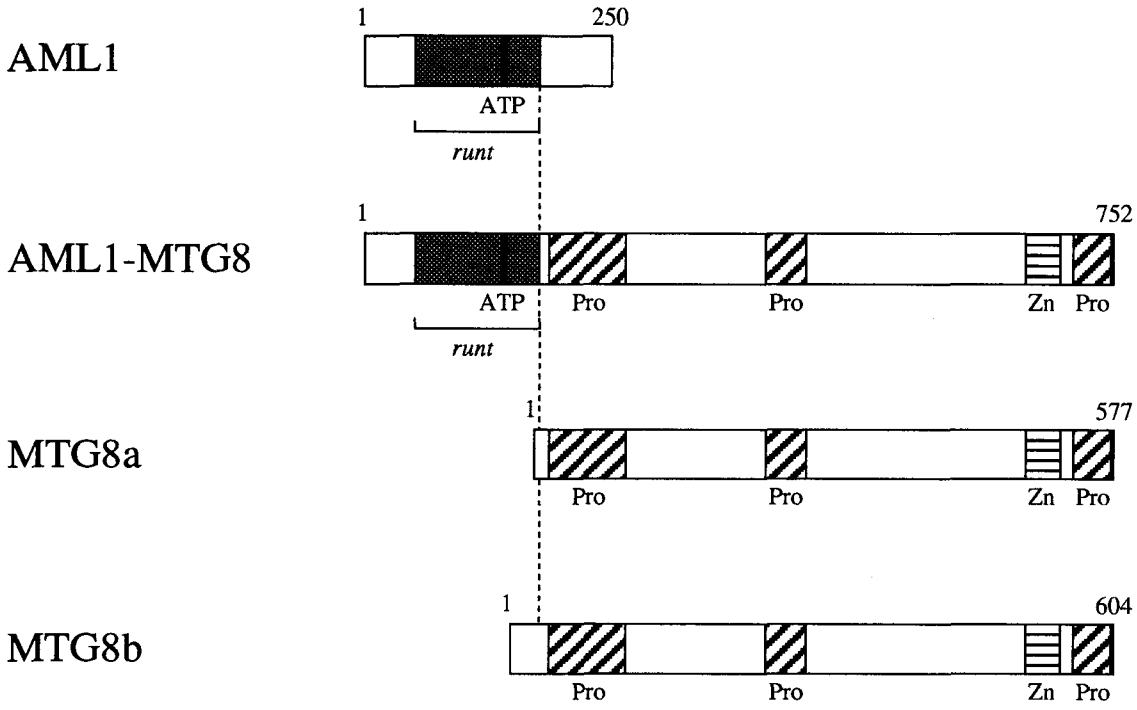


Figure 14. Schematic representation of the AML1, MTG8, and AML1-MTG8 proteins. *runt*, *runt*-homology region; ATP, ATP-binding site motif; Pro, proline-rich region; Zn, zinc finger DNA-binding motifs. Amino acid numbers are shown above the box.

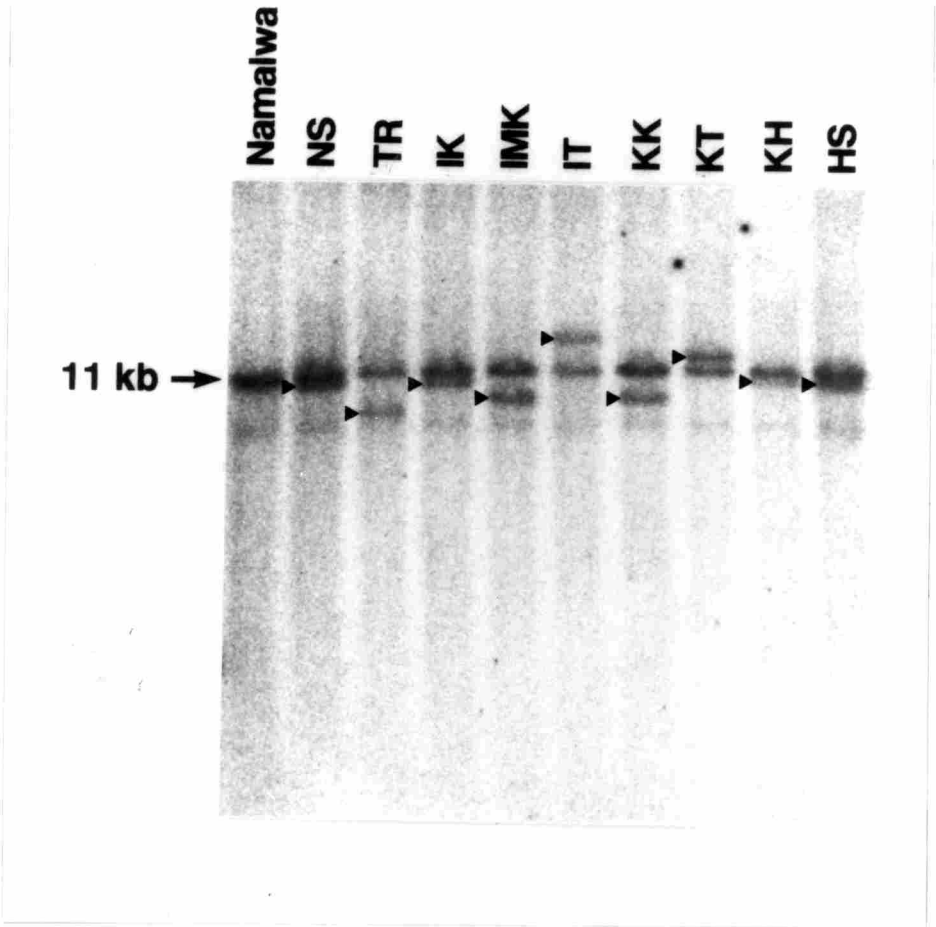


Figure 15. Rearrangements of the *MTG8* gene. Leukemic cell DNAs from nine AML patients with t(8;21) were digested with *Eco*RI and hybridized to the R4H2RN probe (nucleotides 51-396 of *MTG8a* in Figure 13). Rearranged bands and germline bands (11 kb) are indicated by arrowheads and arrow, respectively. Namalwa (Burkitt's lymphoma cell line) cells were used as a normal control. Patient KK has complex t(8;4;21) translocation.

the *MTG8* gene.

Expression of the *AML1-MTG8* fusion gene in t(8;21) AML

The expression of the *AML1*, *MTG8*, and *AML1-MTG8* genes was examined by Northern blot analysis of RNA isolated from blood samples from the t(8;21) AML patients, other AML patients and normal individuals, and several hematopoietic cell lines. An *AML1*-specific probe (C6E3SS6) identified four major transcripts in all samples examined except for the Raji cell line, although abnormal bands were not clearly observed in the t(8;21) AML patients and the Kasumi-1 cell line because of multiple transcripts expressed at high levels (Figure 16). The expression of *AML1* seems to be constitutive at various stages of hematopoietic differentiation, because it was detected in all hematopoietic cell lines examined, especially in myeloid lineages. On the other hand, using an *MTG8*-specific portion of the *AML1-MTG8* fusion cDNA (CH15H2S) as a probe, two major transcripts of 6.2 and 7.8 kb were detected in Kasumi-1 and all four t(8;21) AML samples examined (Figure 16 shows two samples), but not in normal peripheral blood samples, several cell lines without t(8;21) and the t(8;21) AML sample in remission. A 5.7 kb transcript was detected only in the Raji and HEL cell lines, possibly corresponding to a normal *MTG8* transcript. Expression of the *AML1-MTG8* fusion transcript in the t(8;21) AML patients was confirmed by a reverse PCR analysis with primers flanking the fusion point (Kozu *et al.*, 1993). The cDNAs representative of the full-length 6.2 and 7.8 kb fusion transcripts were not cloned. However, all five independent fusion cDNA clones included the same coding frame with different lengths in the 5' untranslated region. Thus the major fusion transcripts probably encode the same fusion protein as that encoded by the cloned portion of cDNAs.

Expression of the *AML1* and *MTG8* genes in mouse tissues

The tissue distributions of the *AML1* and *MTG8* transcripts in various mouse tissues were examined by Northern blot analysis (Figure 17), since the *AML1*- and *MTG8*-specific probes both hybridized strongly with mouse DNA. Various sizes of the *AML1* transcripts were detected at different expression levels in lung, heart, spleen, thymus and ovary, possibly due to tissue-specific alternative splicing, but were not detected in brain, liver, kidney and testis. In

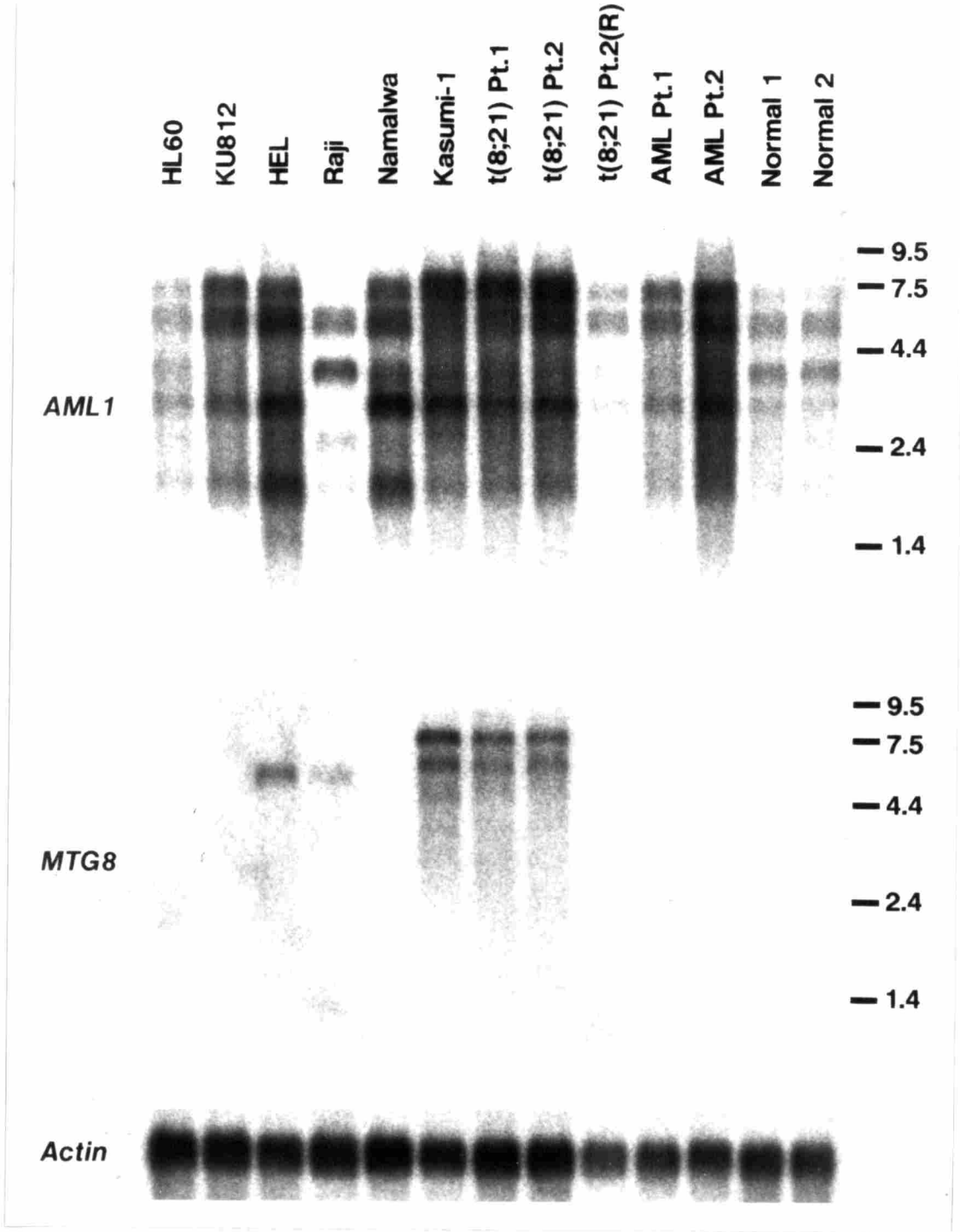


Figure 16. Northern blot analysis of the *AML1*, *MTG8*, and *AML1-MTG8* expressions. Poly(A)+ RNA was isolated from the following cell lines and blood samples: HL60, promyelocytic leukemia cell line; KU812, chronic myelogenous leukemia cell line; HEL, erythroleukemia cell line; Raji and Namalwa, Burkitt's lymphoma cell lines; Kasumi-1, AML cell line with t(8;21); t(8;21) Pt. 1 and Pt. 2, bone marrow samples from the t(8;21) AML patients, t(8;21) Pt. 2 is identical to patient HS in Figure 15; t(8;21) Pt. 2(R), peripheral blood sample from Pt. 2 in remission; AML Pt. 1 and Pt. 2, bone marrow samples from the AML

patients without t(8;21); Normal 1 and 2, peripheral blood samples from normal individuals. Northern blot was sequentially hybridized with the indicated probes: *AML1*-specific probe (C6E3SS6, a 0.4 kb *Sma* I fragment of *AML1* cDNA corresponding to nucleotides 945-1,335 in Figure 7), *MTG8*-specific probe (CH15H2S) and β-actin probe. Size markers (in kb) are shown at the right of the figure.

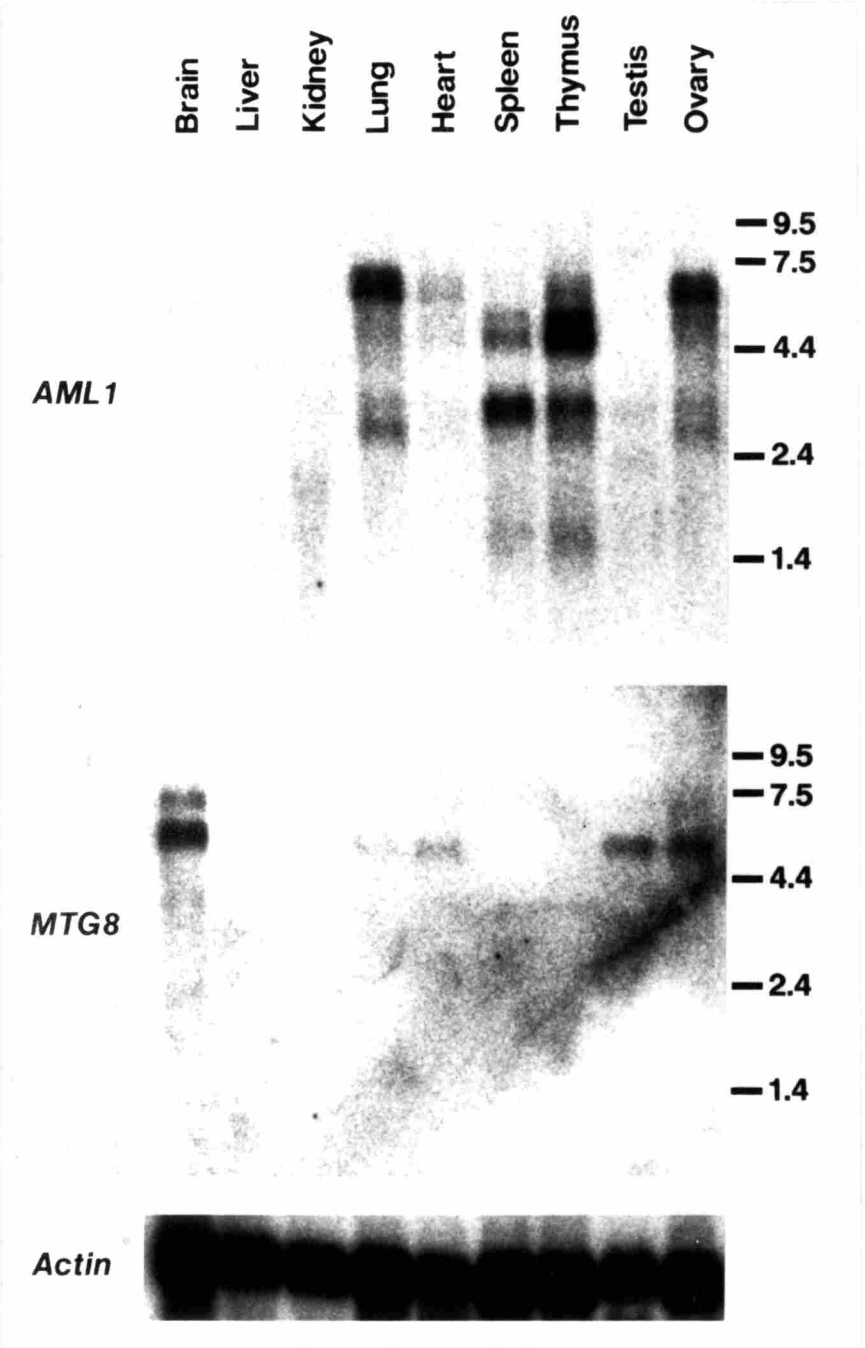


Figure 17. Tissue distributions of the *AML1* and *MTG8* transcripts. Poly(A)+ RNA was isolated from various mouse tissues. Northern blot was prepared and sequentially hybridized with the indicated probes as described in Figure 16. The autoradiographic exposure time was six times that for Figure 16.

contrast, the *MTG8* expression was detected at a high level only in brain and at low levels in lung, heart, testis and ovary. Interestingly, no detectable levels of *MTG8* expression were observed in hematopoietic organs such as spleen and thymus.

Discussion

Chromosomal rearrangements can be visualized as alterations in rare-cutter fragment sizes on PFGE. The use of linking clones, which span rare-cutter sites, simplifies the detection of rearrangements and helps in the construction of long-range physical maps. In addition, the linking clones often mark the position of expressed genes. Through a combination of *Not* I linking clones and PFGE analysis in this initial study, it was possible to detect the t(8;21) rearrangement. This rearrangement could also be detected with yeast artificial chromosome (YAC) clones spanning the breakpoint on chromosome 21 (Gao *et al.*, 1991; Kearney *et al.*, 1991).

The t(8;21) translocation occurs in variant forms (three-way translocations) in a limited number of patients whose clinical course is the same as that of patients with the standard t(8;21) translocation. Since the der(8) chromosome is constant in the variant forms of t(8;21), similar to the Philadelphia (Ph^1) chromosome in the t(9;22) translocation in CML, the critical recombinant chromosome is the der(8) chromosome (Rowley, 1982; Maseki *et al.*, 1993). The studies described here show that the t(8;21) breakpoints as well as the complex translocation breakpoints on chromosome 21 consistently occur in the same intron between two coding exons of the *AML1* gene, and that the t(8;21) translocation juxtaposes the 5' portion of *AML1* gene with the most part of *MTG8* gene, resulting in the expression of the *AML1-MTG8* fusion transcripts from the breakpoint region of the der(8) chromosome. Therefore, it is suggested that the resultant AML1-MTG8 fusion protein may be involved in the pathogenesis of t(8;21) AML.

The *AML1* gene is highly homologous to the *Drosophila* segmentation gene *runt* (Daga *et al.*, 1992), which encodes a nuclear protein containing an ATP-binding site and regulates the expression of other pair-rule genes, although it does not contain any identifiable transcription factor motifs (Kania *et al.*, 1990). The region of highest homology between AML1 (amino acids 50-177) and runt (amino acids 105-232) has 70% amino acid identity. Interestingly, the AML1 protein is disrupted at the C-terminal end of this homologous region by the t(8;21) translocation. This homologous region may be an evolutionary conserved functional domain.

Recently, homology between AML1 and the α subunit of polyomavirus enhancer binding protein 2 (PEBP2α) has been shown (Ogawa *et al.*, 1993). In the region of highest homology, AML1 is more similar to PEBP2α (91% identity in amino acids 93-220) than runt. PEBP2 is a transcription factor and a heterodimer of α and β subunits. PEBP2α is capable of binding to the core sequences of polyomavirus enhancer. The consensus sequence for binding, PuACCPuCA (Kamachi *et al.*, 1990), is also found in enhancer regions of murine leukemia virus and several T cell specific genes such as T cell receptor genes (Satake *et al.*, 1992). Alignment of the homologous domains of AML1, runt and PEBP2α is schematically shown in Figure 18. Based on the structural homology, AML1 is inferred to be a transcription factor with sequence-specific DNA-binding properties and potential to dimerize with other factors.

The predicted structure of MTG8 protein is characterized by two putative zinc fingers and proline-rich domains, which have been shown to have transcriptional activation properties. The MTG8 protein may act as a transcription factor by binding to DNA and regulating the transcriptional activity of specific target genes. Hence, the AML1-MTG8 fusion protein is suggested to be a chimeric transcription factor, containing two putative DNA-binding domains derived from AML1 and MTG8, though it is uncertain whether these domains recognize and bind to specific DNA sequences. Several chimeric transcription factors involved in leukemogenesis have been identified in recent molecular studies on nonrandom chromosomal translocations in hematopoietic malignancies (for reviews see Rabbitts, 1991; Cleary, 1991; Solomon *et al.*, 1991). For example, the t(1;19) translocation in pre-B cell acute lymphoblastic leukemia results in the fusion of the transcriptional activation domain of E2A and the DNA-binding homeodomain of PBX1 (Nourse *et al.*, 1990; Kamps *et al.*, 1990). In acute promyelocytic leukemia, the t(15;17) translocation fuses the retinoic acid receptor α gene with the *PML* gene, which encodes a novel zinc finger protein (Kakizuka *et al.*, 1991; de Thé *et al.*, 1991; Pandolfi *et al.*, 1991; Goddard *et al.*, 1991).

Northern blot analysis of *AML1* and *MTG8* expressions may help to elucidate the role of *AML1-MTG8* fusion gene in myeloid leukemogenesis. Expression of *AML1* gene is

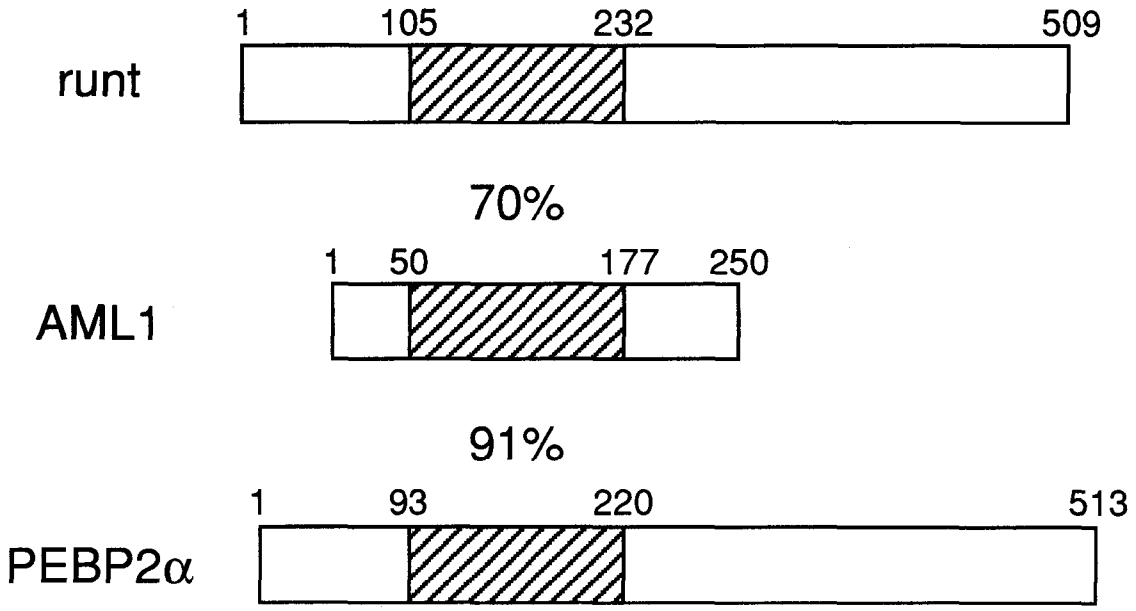


Figure 18. Schematic alignment of homologous domain in the amino acid sequences for the AML1, runt, and PEBP2. The cross-hatched regions indicate homologous domain within the three proteins. The numbers above the boxes represent the amino acid positions.

probably constitutive in multiple hematopoietic lineages, suggesting that *AML1* may play an important role in hematopoietic cell growth and/or differentiation. If so, the AML1-MTG8 fusion protein might interfere with the function of wild-type AML1 as a dominant negative mutant. Alternatively, aberrant expression of truncated *MTG8* following juxtaposition with the 5' portion of *AML1* as a promoter may be involved in leukemogenic transformation as in the case of c-*MYC* in Burkitt's lymphoma, since the *MTG8* gene is not expressed in normal hematopoietic cells. However, it is most likely that the resultant AML1-MTG8 fusion protein contributes directly to leukemogenesis, because the breakpoints on chromosome 21 consistently occurred in the single specific intron of *AML1* and AML1-MTG8 fusion protein retains presumable functional domains of AML1 protein, such as ATP-binding and DNA-binding domains, which would act as functional parts of chimeric oncoprotein. The AML1-MTG8 protein presumably acts as a chimeric transcription factor and alters the transcriptional activity of AML1- or MTG8-responsive genes which contribute to neoplastic process of hematopoietic origin. It is also possible that AML1-MTG8 would have different biological properties from those of the respective wild-type proteins. Functional analysis of the AML1-MTG8 fusion protein is required to clarify its role in AML as well as the characterization of wild-type AML1 and MTG8 proteins.

Materials and methods

Not I linking clones and regional mapping

A *Not* I linking library was constructed from a lambda phage library containing *Hin*dIII inserts prepared from flow-sorted human chromosome 21 as described (Saito *et al.*, 1991). The *Not* I linking clones were selected by inserting a supF fragment into *Not* I sites as a selection maker (Ichikawa *et al.*, 1992). Regional mapping of the *Not* I linking clones were performed by quantitative Southern blot analysis using two cell lines carrying an unbalanced chromosome 21 translocation (GM09552 and GM06136), and chromosomal in situ hybridization techniques using lymphoblastoid cell lines carrying a balanced chromosome 21 translocation (GM06135, GM09528A, GM09542). This allocated linking clones to one of three regions, pter-q21, q21-q22.3, and q22.3-qter.

PFGE analysis

High molecular weight DNAs were prepared in 0.8% agarose gel plugs and treated with ESP (0.5 M EDTA, 1% Na-sarcocine, and 1 mg/ml proteinase K) at 50°C for 48 hours (Smith and Cantor, 1987). About 6 μg of DNA in each plug was digested to completion with rare-cutter restriction enzymes, and the digests were separated by PFGE in a BioRad apparatus (CHEFDRII) with 1% agarose gel in 0.5 x TBE (90 mM tris base, 90 mM boric acid, and 2.5 mM Na_2H_2EDTA) at 200 V (6 V/cm) for 24 hr. The pulse time in the linear gradient was 30 to 120 seconds. DNAs were stained with ethidium bromide, nicked with 0.25 M HCl for 8 minutes, and transferred to Hybond-N membrane (Amersham). The sizes of the fragments were estimated with *Saccharomyces cerrevisiae* YNN295 (BioRad).

Southern blot analysis

Genomic DNA (5μg) was digested with appropriate restriction enzymes, fractionated by conventional or field inversion gel electrophoresis with a Bio-Rad apparatus (CHEF MAPPER), and transferred to Hybond-N membrane (Amersham). Hybridization with random primer-labeled probe was carried out in 6X SSC, 10% dextran sulfate, 1% SDS, 1X Denhardt's solution, denatured salmon sperm DNA (100 μg/ml), and 50% formamide at 42°C with or without human placental DNA (100μg/ml). The final washing was in 0.1X SSC,

0.1% SDS at 65°C. Autoradiography was performed using a bioimage analyzer, Fujix BAS 2000.

Northern blot analysis

Total RNA was isolated by the AGPC method (Chomozynski and Sacchi, 1987). Poly(A)+ RNA was selected by using Oligotex-dT30 (Roche). Samples of poly(A)+ RNA (1.5 μg) were separated on 1% agarose-formaldehyde gel. Blotting, hybridization, washing, and autoradiography were performed as described for Southern blot analysis.

Somatic cell hybrids

Somatic cell hybrids were obtained by fusing mouse myeloma X63 cells deficient in hypoxanthine phosphoribosyl transferase (HPRT) and leukemic cells from patient 2 using polyethylene glycol 3500. Hybrid clones were first selected by HAT medium and maintained in RPMI 1640 medium supplemented with 15% fetal calf serum without HAT. Clones containing the der(8) or der(21) chromosome were identified and screened for the presence or absence of DNA markers specific for human chromosome 8 and chromosome 21; c-*mos* (8q11-q12), c-*myc* (8q24), D21S16 (21q11) and D21S15 (21q22).

Genomic and cDNA cloning of *AML1*

Genomic clones λE3 and λE4 were isolated from a human leukocyte genomic library (Clontech, EMBL-3) with the LL263L as a probe, which is one of two fragments of the linking clone LL263 digested with *Not* I (13). λE12 was isolated from the same library with λE4. The cDNA clones C6 and B1-3 were isolated from a human bone marrow cDNA library (Clontech, λgt10) with λE4 and C6E6H2, a *Hin*dIII-*Eco*RI fragment of C6 clone, as probes, respectively. Other cDNA clones shown in Figure 5B were from a human peripheral blood leukocyte cDNA library (Clontech, λgt10). Genomic clones λD11 and λD13 were isolated from a human lymphocyte genomic library (Stratagene, λDASH) using C6E6H2 probe. Library screening and plaque purification were performed by the standard procedures (Sambrook *et al.*, 1989).

Cloning of rearranged fragments

Leukemic cell DNA from patient KH was digested with *Bam*HI and cloned into λDASH II

phage vector (Stratagene). The library was screened with the C6E6H2 and D11X1 probes to isolate the clones containing the breakpoint region of the der(8), der(21) and normal chromosome 21. The chromosome 8-specific portion of the clone derived from the der(21) was used as a probe to isolate the clone containing the corresponding region of normal chromosome 8. The inserts of phage clones containing the breakpoint region were digested with *Eco*RI, subcloned into the pBluescript II (KS+) (Stratagene) and sequenced.

Cloning of *AML1-MTG8* cDNAs

Poly(A)+ RNA (5 μg) from the Kasumi-1 cell line was converted to double-stranded cDNA. Size-selected cDNA library (>2.0 kb) was constructed in the λ ZAPII vector (Stratagene) with *Eco*RI-*Not* I adaptors (Pharmacia). The library was screened with the C6E6H2 and CH15H2S probes. Approximately 1 x 10^6 recombinant phages were screened, and twenty of more than 100 positive plaques were initially isolated. The cDNA inserts from each positive clone were checked by *Eco*RI and *Not* I digestion of the purified phage DNA and agarose gel electrophoresis. Subsequently, the inserts of five clones were subcloned into the *Not* I site of pBluescript II (KS+) (Stratagene) for further analysis.

Cloning of *MTG8* cDNAs

Wild-type *MTG8* cDNAs were isolated from the human fetal brain (17-18 weeks gestation) cDNA library (Stratagene) using the CH15H2S probe. Twenty positive clones were initially isolated and characterized. Eight overlapping clones were then selected for sequence analysis, and two types of cDNA, *MTG8a* and *MTG8b*, were identified. *MTG8a* type cDNA could also be isolated from the Raji cell line cDNA library using the CH15H2S probe. *MTG8b* type cDNA was also isolated from the HEL cell line using anchored PCR with antisense primer downstream of the fusion point with *AML1*.

DNA sequencing

Purified phage DNAs from either cDNA or genomic phage libraries were restriction mapped, and appropriate fragments were subcloned into the pBluescript II (KS+) (Stratagene). When needed, a nested series of deletions was generated by using Exo/Mung Kit (Stratagene). DNA sequencing was performed by the dideoxy chain-termination method, using Sequenase version

2.0 DNA sequencing kit (United States Biochemical) with [α-^{35}S] dCTP or A.L.F. DNA Sequencer (Pharmacia) with fluorescent labeled primers.

References

Abe,R., Umezu,H., Uchida,T., Kariyone,S., Maseki,N., Kaneko,Y. and Sakurai,M. (1986) *Cancer*, **58**, 1260-1264.

Adams,J.M. and Cory,S. (1991) *Biochim. Biophys. Acta*, **1072**, 9-32.

Asou,H., Tashiro,S., Hamamoto,K., Otsuji,A., Kita,K. and Kamada, N. (1991) *Blood*, **77**, 2031-2036.

Bakhshi,A., Jensen,J.P., Goldman,P., Wright,J.J., McBride,O.W., Epstein,A.L. and Korsmeyer,S.J. (1985) *Cell*, **41**, 899-906.

Bakhshi,A., Wright,J.J., Graninger,W., Seto,M., Owens,J., Cossman,J., Jensen,J.P., Goldman,P. and Korsmeyer,S.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2369-2400.

Berg,J.M. (1990) *J. Biol. Chem.*, **265**, 6513-6516.

Berger,R., Bernheim,A., Daniel,M.-T., Valensi,F., Sigaux,F. and Flandrin,G. (1982) *Blood*, **59**, 171-178.

Bird, A. P. (1987) *Trends Genet.* **3**, 342-347.

Bishop,J.M. (1991) *Cell*, **64**, 235-248.

Chen-Levy,Z., Nourse,J. and Cleary,M.L. (1989) *Mol. Cell. Biol.*, **9**, 701-710.

Chomozynski,P. and Sacchi,N. (1987) *Analyt. Biochem.* **162**, 156-159.

Cleary,M.L. (1991) *Cell*, **66**, 619-622.

Cleary,M.L. and Sklar,J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7439-7443.

Cory,S. (1986) *Adv. Cancer Res.* **47**, 189-234.

Croce,C.M. and Nowell,P.C. (1985) *Blood*, **65**, 1-7.

Daga,A., Tighe,J.E. and Calabi,F. (1992) *Nature*, **356**, 484.

Daley,G.Q., Van Etten,R.A. and Baltimore,D. (1990) *Science*, **247**, 824-830.

Dalla-Favera,R., Bregni,M., Erickson,J., Patterson,D., Gallo,R.C. and Croce,C.M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7824-7827.

de Thé,H., Lavau,C., Marchio,A., Chomienne,C., Degos,L. and Dejean,A. (1991) *Cell*, **66**, 675-684.

Drabkin,H.A., Diaz,M., Bradley,C.M., Le Beau,M.M., Rowley,J.D. and Patterson,D.

(1985) *Proc. Natl. Acad. Sci. USA* **82**, 464-468.

Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1-3.

Gao,J., Erickson,P., Gardiner,K., Le Beau,M.M., Diaz,M.O., Patterson,D., Rowley,J.D. and Drabkin,H.A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4882-4886.

Goddard,A.D., Borrow,J., Freemont,P.S. and Solomon,E. (1991) *Science*, **254**, 1371-1374.

Heim,S. and Mitelman,F. (1989) *Adv. Cancer Res.*, **52**, 1-42.

Heisterkamp,N., Stephenson,J.R., Groffen,J., Hansen,P.F., de Klein,A., Bartram,C.R. and Grosveld,G. (1983) *Nature*, **306**, 239-242.

Heisterkamp,N., Stam,K., Groffen,J., de Klein, A. and Grosveld,G. (1985) *Nature*, **315**, 758-761.

Heisterkamp,N., Jenster,G., ten Hoeve,J., Zovich,D., Pattengale,P.K. and Groffen,J. (1990) *Nature*, **344**, 251-253.

Hockenberry,D., Nunez,G., Milliman,C., Schreiber,R.D. and Korsmeyer,S.J. (1990) *Nature*, **348**, 334-336.

Ichikawa,H., Shimizu,K., Saito,A., Wang,D., Oliva,R., Kobayashi,H., Kaneko,Y., Miyoshi,H., Smith,C.L., Cantor,C.R. and Ohki,M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 23-27.

Kakizuka,A., Miller,Jr.,W.H., Umesono,K., Warrell,Jr.,R.P., Frankel,S.R., Murty, V.V.V.S., Dmitrovsky,E. and Evans,R.M. (1991) *Cell*, **66**, 663-674.

Kamachi,Y., Ogawa,E., Asano,M., Ishida,S., Murakami,Y., Satake,M., Ito,Y. and Shigesada,K. (1990) *J. Virol.*, **64**, 4808-4819.

Kamps,M.P., Murre,C., Sun,X.-h. and Baltimore,D. (1990) *Cell*, **60**, 547-555.

Kania,M.A., Bonner,A.S., Duffy,J.B. and Gergen,J.P. (1990) *Genes Dev.*, **4**, 1701-1713.

Kearney,L., Watkins,P.C., Young,D. and Sacchi,N. (1991) *Cancer Genet. Cytogenet.*, **57**, 109-119.

Konopka,J.B., Watanabe,S.M. and Witte,O.N. (1984) *Cell*, **37**, 1035-1042.

Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125-8148.

Kozu,T., Miyoshi,H., Shimizu,K., Maseki,N., Kaneko,Y., Asou,H., Kamada,N. and Ohki,M. (1993) *Blood*, in press.

Kurzrock,R., Gutterman,J.U. and Talpaz,M. (1988) *New Engl. J. Med.*, **319**, 990-998.

Lugo,T.G., Pendergast,A.-M., Muller,A.J. and Witte,O.N. (1990) *Science*, **247**, 1079-1082.

Maseki,N., Miyoshi,H., Shimizu,K., Homma,C., Ohki,M., Sakurai,M. and Kaneko,Y. (1993) *Blood*, **81**, 1573-1579.

Marshall,C.J. (1991) *Cell*, **66**, 313-326.

McKeithan,T.W., Ohno,H. and Dias,M.O. (1990) *Genes Chrom. Cancer*, **1**, 247-255.

Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741-753.

Mitelman,F., Kaneko,Y. and Trent,J. (1991) *Cytogenet. Cell Genet.*, **58**, 1053-1079.

Moller,W. and Amons,R. (1985) *FEBS Lett.* **186**, 1-7.

Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,J., Stanbridge,E., Smith,S.D. and Cleary, M. L. (1990) *Cell*, **60**, 535-545.

Nowell,P.C. and Hungerford,D.A. (1960) *Science*, **132**, 1497.

Ogawa,E., Maruyama,M., Kagoshima,H., Inuzuka,M., Lu,J., Satake,M., Shigesada,K. and Ito,Y. (1993) *Proc. Natl. Acad. Sci. USA*, in press.

Owens,G.P., Hahn,W.E. and Cohen,J.J. (1991) *Mol. Cell. Biol.*, **11**, 4177-4188.

Pandolfi,P.P., Grignani,F., Alcalay,M., Mencarelli,A., Biondi,A., LoCoco,F., Grignani,F. and Pelicci,P.G. (1991) *Oncogene*, **6**, 1285-1292.

Rabbitts,T.H. (1991) *Cell*, **67**, 641-644.

Rogers,S., Wells,R. and Rechsteiner,M. (1986) *Science*, **234**, 364-368.

Rowley,J.D. (1973a) *Nature*, **243**, 290-293.

Rowley,J.D. (1973b) *Ann. Genet.*, **16**, 109-112.

Rowley,J.D. (1982) *Science*, **216**, 749-751.

Rowley,J.D. (1984) *Cancer Res.*, **44**, 3159-3168.

Rowley,J.D. (1990) *Cancer Res.*, **50**, 3816-3825.

Sacchi,N., Watson,D.K., Guerts van Kessel,A.H.M., Hagemeijer,A., Kersey,J.,

Drabkin,H.D., Patterson, D. and Papas,T.S. (1986) *Science* **231**, 379-382.

Saito,A., Abad,J.P., Wang,D., Ohki,M., Cantor,C.R. and Smith,C.L. (1991) *Genomics*, **10**, 618-630.

Sakurai,M., Oshimura,M., Kakati,S. and Sandberg,A.A. (1974) *Lancet*, **2**, 227-228.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).

Satake,M., Inuzuka,M., Shigesada,K., Oikawa,T. and Ito,Y. (1992) *Jpn. J. Cancer Res.*, **83**, 714-722.

Shaw,G. and Kamen,R. (1986) *Cell* **46**, 659-667.

Shtivelman,E., Lifshitz,B., Gale,R.P. and Canaani,E. (1985) *Nature*, **315**, 550-554.

Smith,C.L. and Cantor,C.R. (1987) *Methods Enzymol.*, **155**, 449-467.

Solomon,E., Borrow,J. and Goddard,A.D. (1991) *Science*, **254**, 1153-1160.

Stewart,G.D., Harris,P., Galt,J. and Ferguson-Smith,M.A. (1985) *Nucleic Acids Res.*, **13**, 4125-4132.

Taub,R., Kirsch,I., Morton,C., Lenoir,G.M., Swan,D., Tronick,S., Aaronson,S. and Leder,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7837-7841.

Tsujimoto,Y., Cossman,J., Jaffe,E. and Croce,C.M. (1985) *Science*, **228**, 1440-1443.

Tsujimoto,Y., Ikegaki,N. and Croce,C.M. (1987) *Oncogene*, **2**, 3-7.

von Lindern,M., Fornerod,M., van Baal,S., Jaegle,M., de Wit,T., Buijs,A. and Grosveld,G. (1992) *Mol. Cell. Biol.*, **12**, 1687-1697.

Weinberg,R.A. (1991) *Science*, **254**, 1138-1146.

Acknowledgements

I warmly thank M. Ohki for continuing support and encouragement, members of the Ohki laboratory, especially K. Shimizu for valuable discussions and sharing data, H. Ichikawa and T. Kozu for their excellent assistance with parts of this work and helpful discussions, M. Ohira, J. Nagata, and K. Enomoto for experimental help, F. Hosoda and Y. Arai for discussions and advice, M. Mori, Y. Sugiyama and M. Fukushima for technical assistance and N. Kobayashi for help in the preparation of the manuscript. I also thank Y. Kaneko and N. Maseki for patient samples and helpful comments, K. Nakachi for critical reading of the manuscript, O. Gotoh for computer analysis, M. Oshimura and N. Kamada for providing cell lines, K. Sakurada and M. Miyamoto for technical advice.