AF6q21, a Novel Partner of the MLL Gene in (t(6;11)(q21;q23), Defines a Forkhead Transcriptional Factor Subfamily

By Josette Hillion, Maryvonne Le Coniat, Philippe Jonveaux, Roland Berger, and O.A. Bernard

Fusion genes implicating the MLL gene have been recently demonstrated in various 11q23 chromosomal abnormalities in human hematopoietic malignancies. We analyzed a t(6;11)(q21;q23) translocation detected in a secondary acute myeloblastic leukemia. This translocation results in fusion of the MLL gene on 11q23 to a previously unknown gene on chromosome 6 that differs from the previously reported MLL partner gene AF6q. The novel gene, named AF6q21, encodes a forkhead (FH) protein with strong similarities to the two FH family members whose genes are already known to be involved in chromosomal translocations of human malignancies, AFX and FKHR. Strikingly, in these translocations the breakpoints are located at the same position within the FH domains. Therefore, AF6q21, AFX, and FKHR could define a new FH subfamily particularly involved in human malignancies.

© 1997 by The American Society of Hematology.

CHROMOSOMAL BAND 11q23 is nonrandomly rearranged with approximately 30 other bands in chromosomal abnormalities of malignant blood disorders. Most rearrangements of 11q23 occur within a clustered region of 8.3 kb and result in the fusion of the MLL gene, normally localized to 11q23, to a partner gene on the other chromosome. Based on molecular and cytogenetic data, the crucial chimeric gene has been hypothesized to be the 5′-MLL-partner 3′ fusion, and this has been recently demonstrated using “knock-in” mice expressing a fused MLL-AF9 product. Ten of these chimeric genes have been identified to date, and no common features or homologies have been found that might explain their involvement in these nonrandom events. Some partner genes are clearly related to each other, such as AF9/ENL and AF17/AF10. Interestingly, AF10 is also found to be rearranged independently of MLL in the t(10;11)(p13;q14) chromosomal translocation observed in the human leukemia cell line U937. Identification of other partners of MLL is now in progress to try to understand the consequences of the fusion of MLL to so many various partner genes.

The detection of t(6;11)(q21;q23) in a patient with secondary acute leukemia (AL) prompted us to identify the resulting molecular rearrangement. This region of chromosome 6q is indeed rearranged in a variety of hematopoietic disorders.

SUBJECTS AND METHODS

Patient. M.D., a 40-year-old man, was examined for AL in the Department of Hematology of Hospital Saint-Louis. He had been treated for Hodgkin’s disease. When AL was discovered, the bone marrow was hypercellular with 100% blast cells. The immunophenotype was DR+, CD22+, CD29+, CD33+, and CD34+. From cytologic and immunologic features, the diagnosis was undifferentiated AL. Chromosomal studies were performed after 24 and 48 hours of bone marrow culture cells, and the RHG-banding procedure was performed. The karyotype was 46,XY,t(6;11)(q21;q23)[12]/idem,-del(1)(q41)[10] (Fig 1).

Nucleic acid analysis. DNA was extracted by standard procedures, and RNA was isolated using Trizol reagent (GIBCO-BRL, Cergy Pontoise, France) following the manufacturer’s instructions. Nucleic acid experiments were performed using standard protocols. Northern blot membranes were from Clontech Laboratory (Clontech, Palo Alto, CA). The FA4 probe (a gift from Y. Gu, Jefferson Cancer Institute, Philadelphia, PA, and E. Canaani, Weizmann Institute of Science, Rehovot, Israel) encompasses exon 8 of MLL, and I5 is a 0.4-kb Avui fragment located within intron 5 of MLL. The K562 cDNA library (in λgt11) was purchased from Clontech Laboratory, and a genomic DNA library was established from peripheral blood cells of a healthy male. The libraries were screened using standard protocols. The nucleotide sequence was determined using the T7 sequencing kit (Pharmacia, Les Ulis, France).

Polymerase chain reaction. Briefly, 5 μg total RNA was reverse-transcribed with 50 pmol primer 1, GATTAGATCCTAATATC (N5), using the cDNA cycle kit (Invitrogen, San Diego, CA). One one-hundredths was incubated in a 50-μL reaction mixture with 20 pmol of each synthetic primer, 100 μmol · L−1 of each dNTP, 1.5 mmol · L−1 MgCl2, 50 mmol · L−1 KCl and 1 U AmpliTag polymerase (Cetus, Emeryville, CA). Denaturation was performed at 94°C for 90 minutes in the first cycle and for 1 minute in 35 subsequent cycles. Annealing and extension steps were performed for 1 minute at 56°C and 90 seconds at 72°C, respectively. The oligonucleotides were MLL Ex7 (TCACGACCTCTCCTCAAATGG) and oligo AF61 (CCTTTGTGCTATTTGCCA).

Anchored polymerase chain reaction. Random primed cDNA was purified on chromaspin (Clontech). One fifth was submitted to polymerase chain reaction (PCR) amplification for 30 cycles with oligo 4 (GATTAGATCCTAATATC) and oligo 1, 11q23 Ex5 (CACGCCTATGGACCCAG). After gel purification, a second PCR was performed for 30 cycles between oligo 4 and a second nested primer 2, oligo 11q23 Ex5/6 (ggcaattCAGAGCGAACAAACAG). The amplified cDNAs were subcloned and isolated using a third, internal, primer: oligo 11q23 Ex6 (CCCAGTATCCTTGAACAAACA).

PCR-mediated library screening. Gene-specific and λgt11-specific primers were used in the first round of PCRs; a second round of PCRs were performed using the same λgt11-specific primer with a nested gene-specific primer. PCR products were subcloned and submitted to sequence analysis.

Fluorescence in situ hybridization studies. Fluorescence in situ hybridization (FISH) to metaphase chromosomes was performed as...
MLL-FORKHEAD GENE FUSION IN HUMAN LEUKEMIA

HYBRIDIZATION OF **Bam**HI- AND **Hind**III-DIGESTED DNA WITH THE FA4 AND I5 PROBES SHOWED RARRANGED FRAGMENTS IN ADDITION TO NORMAL-SIZED FRAGMENTS, WHEREAS IN **Eco**RI-DIGESTED DNA, ONLY FA4 DETECTED AN ABNORMAL FRAGMENT (FIG 2). THE 11q23 BREAKPOINT WAS thus LOCATED WITHIN THE COMMON BREAKPOINT CLUSTER REGION OF MLL. SINCE THE RARRANGED FRAGMENTS DETECTED BY THE TWO PROBES DIFFERED IN SIZE, THE BREAKPOINT WAS PREDICTED TO BE BETWEEN MLL EXONS 6 AND 8.

CHARACTERIZATION OF 6q21 SEQUENCES FUSED TO MLL. TO CHARACTERIZE THE PUTATIVE FUSED cDNA BETWEEN MLL AND THE 6q PARTNER GENE, ANCHORED PCR EXPERIMENTS WERE PERFORMED STARTING FROM RANDOMLY PRIMED RNA FROM THE PATIENT (SEE SUBJECTS AND METHODS). SEQUENCE ANALYSIS OF A SUBSET OF MLL CDNAS REVEALED A NOVEL EXON SUBSTITUTING FOR EXON 8.16-KB GENOMIC INSERT ENCOMPASSING THE FIRST INTRON OF **AF6q21**, AND T3, CONTAINING A 20-KB GENOMIC INSERT CORRESPONDING TO THE 3' PART OF THE GENE.

RESULTS

INvolvement of MLL in the t(6;11)(q21;q23) translocation. TO ESTABLISH THE INVOLVEMENT OF MLL IN THIS TRANSLOCATION, SOUTHERN BLOT ANALYSIS OF HIGH-MOLECULAR-WEIGHT DNA FROM THE PATIENT'S BONE Marrow CELLS WAS PERFORMED USING THE PROBES COVERING THE BREAKPOINT CLUSTER REGION OF MLL RARRANGEMENTS IN ALs. HYBRIDIZATION OF **Bam**HI- AND **Hind**III-DIGESTED DNA WITH THE FA4 AND I5 PROBES SHOWED RARRANGED FRAGMENTS IN ADDITION TO NORMAL-SIZED FRAGMENTS, WHEREAS IN **Eco**RI-DIGESTED DNA, ONLY FA4 DETECTED AN ABNORMAL FRAGMENT (FIG 2). THE 11q23 BREAKPOINT WAS thus LOCATED WITHIN THE COMMON BREAKPOINT CLUSTER REGION OF MLL. SINCE THE RARRANGED FRAGMENTS DETECTED BY THE TWO PROBES DIFFERED IN SIZE, THE BREAKPOINT WAS PREDICTED TO BE BETWEEN MLL EXONS 6 AND 8.

CHARACTERIZATION OF 6q21 SEQUENCES FUSED TO MLL. TO CHARACTERIZE THE PUTATIVE FUSED cDNA BETWEEN MLL AND THE 6q PARTNER GENE, ANCHORED PCR EXPERIMENTS WERE PERFORMED STARTING FROM RANDOMLY PRIMED RNA FROM THE PATIENT (SEE SUBJECTS AND METHODS). SEQUENCE ANALYSIS OF A SUBSET OF MLL CDNAS REVEALED A NOVEL EXON SUBSTITUTING FOR EXON 8.16-KB GENOMIC INSERT ENCOMPASSING THE FIRST INTRON OF **AF6q21**, AND T3, CONTAINING A 20-KB GENOMIC INSERT CORRESPONDING TO THE 3' PART OF THE GENE.

RESULTS

INvolvement of MLL in the t(6;11)(q21;q23) translocation. TO ESTABLISH THE INVOLVEMENT OF MLL IN THIS TRANSLOCATION, SOUTHERN BLOT ANALYSIS OF HIGH-MOLECULAR-WEIGHT DNA FROM THE PATIENT'S BONE Marrow CELLS WAS PERFORMED USING THE PROBES COVERING THE BREAKPOINT CLUSTER REGION OF MLL RARRANGEMENTS IN ALs. HYBRIDIZATION OF **Bam**HI- AND **Hind**III-DIGESTED DNA WITH THE FA4 AND I5 PROBES SHOWED RARRANGED FRAGMENTS IN ADDITION TO NORMAL-SIZED FRAGMENTS, WHEREAS IN **Eco**RI-DIGESTED DNA, ONLY FA4 DETECTED AN ABNORMAL FRAGMENT (FIG 2). THE 11q23 BREAKPOINT WAS thus LOCATED WITHIN THE COMMON BREAKPOINT CLUSTER REGION OF MLL. SINCE THE RARRANGED FRAGMENTS DETECTED BY THE TWO PROBES DIFFERED IN SIZE, THE BREAKPOINT WAS PREDICTED TO BE BETWEEN MLL EXONS 6 AND 8.
HILLION ET AL

Fig 3. (A) Nucleotide sequence and predicted amino acid sequences of fused MLL-AF6q21 cDNAs. Fusion point is shown by arrows. (B) Genomic sequence surrounding exon 1 of AF6q21. The deduced protein is shown above the sequence. The upstream in-frame stop codon is shown by an asterisk. The longest AF6q21 cDNA isolated from the K562 cDNA library starts at nucleotide 90. Further upstream sequences are of genomic origin. Intronic sequences downstream of the putative splice site appear as lowercase letters. (C) Predicted AF6q21 protein starting at the second in-frame ATG. The NH2 region conserved with AFX and FKHR and the FH domain are underlined. The arrows border the fusion point with MLL.

to human chromosome 6. A weak signal was also observed with hybrid DNA containing chromosome 17 (data not shown).

To confirm the presence of this chimeric MLL RNA in the patient’s material, bispecific PCR was performed using oligo MLL Ex7 on the 5' side and oligo AF61 on the 3' side. The amplified product had the predicted size (247 bp) and was not seen in control cDNA. Its specificity was confirmed by nucleotide sequencing. The partial sequence encompassing the t(6;11) fusion point is shown in Fig 3A.

Therefore, t(6;11)(q21;q23) results in fusion between MLL and a gene on chromosome 6, provisionally named AF6q21.

Isolation of AF6q21 cDNAs. To further characterize the AF6q21 gene, a cDNA library established from the K562 cell line was screened using the AF6q21 probe. Fourteen cDNAs were isolated, and the longest was chosen for nucleotide sequencing. None of the cDNAs harbored a poly(A) tail. No ATG initiator codons were identified 5' to the putative open reading frames (ORFs) in this set of cDNAs. Further 5' AF6q21 sequences were obtained upon PCR screening of the K562 cDNA library up to nucleotide 90 (Fig 3). Probes covering almost all of the AF6q21 ORF were also used to isolate genomic sequences from a normal genomic library. Sequencing of the genomic counterpart confirmed the nucleotide sequence of the PCR-isolated fragments and allowed us to extend further 5' of the AF6q21 sequences. A cDNA probe containing almost all of the coding sequences reacted with a 6-kb RNA species, as did the 199-bp PCR probe, in RNA from various cell lines (Fig 4). Hybridization of a multiple tissue blot with the same probe revealed the same 6-kb RNA species in all organs examined, suggesting a ubiquitous expression of the AF6q21 gene.

Compilation of sequence data allowed the establishment of a 2,549-bp consensus sequence. Analysis of this consensus...
Fig 4. Northern blot analysis of AF6q21 transcription in several human leukemic cell lines (left) and several human tissues (right) using a cDNA probe corresponding to almost all of the AF6q21 ORF. Control hybridization with an actin probe is shown at bottom. The position of size markers is indicated. Additional smaller species are detected in some samples upon longer exposure (data not shown). They could be due to variation in alternate splicing within the AF6q21 gene or to expression of a homologous gene such as the one postulated for chromosome 17. Loading is as follows: Left panel: 1, HL60; 2, HeLa; 3, K562; 4, MOLT4; 5, Raji; 6, SW480; 7, A549; 8, G361. Right panel: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas.

The genomic sequences further upstream show a high GC content. Since such GC-rich regions are frequently associated with the 5′ end of genes, this could represent the 5′ end of the AF6q21 gene (Fig 3B).

Of the first two in-frame ATGs, the second occurs within a better Kozak consensus (ACCATGG) because of purines at positions −3 and +4. We therefore postulate that the second in-frame ATG is the major site of translational initiation of the AF6q21 RNA allowing the synthesis of a 381–amino acid protein. This second ATG is in a position similar to the ATGs of two proteins homologous to the AF6q21 protein (description follows), further supporting this hypothesis.

AF6q21 encodes a Forkhead domain protein. Sequence comparisons of the deduced AF6q21 protein to protein databases showed significant similarity to the forhead (FH) DNA-binding domain proteins (Fig 3C). FH proteins belong to a rapidly growing family of proteins with site-specific DNA-binding and transcription-regulation properties.

Among the FH protein family, particularly high homology within the FH domain is observed with the AFX and FKHR proteins, whose genes are recombined following chromosomal translocations in human malignancies. AFX was isolated as an MLL translocation partner following the study of a t(X;11)(q13;q23) translocation in a human ALL cell line. The FKHR gene is fused to PAX3 in t(2;13) (q35;q14) and to PAX7 in t(1;13)(p36;q14) translocations associated with alveolar rhabdomyosarcoma. Amino acid alignment of the FH domain of these three FH proteins is shown in Fig 5, and a derived consensus sequence is compared with a previously defined HNF3 FH consensus. AF6q21, AFX, and FKHR share several features that differ from other FH proteins, as previously noted. They lack a KPPY tetrapeptide at the NH2 side of the FH domain, and exhibit several additional or missing amino acids with respect to the FH consensus (Fig 5).

A second region at amino acids 16 to 39 of AF6q21 is found to be conserved between these three FH proteins (17 of 21 amino acids are identical within the three proteins; Fig 5). The corresponding region starts at amino acid 12 of AFX and amino acid 8 of FKHR. Database comparison did not reveal other proteins with significant homology to this amino acid sequence. The significance of the common features of the three FH domain proteins remains to be determined, but they may characterize a particular subfamily of FH proteins.

Partial genomic characterization of the AF6q21 gene. Examination of nucleotide sequences upstream of the putative AF6q21 translational start sites showed a high CG content, several recognition sites for rare cutters, and methylation-sensitive enzymes. Consistently, 93 nucleotides 645 bp upstream of the first nucleotide shown in Fig 3B were found to be identical to a reported CpG island genomic DNA fragment (cpg86e6, accession no. Z63553). This 119-bp DNA fragment has been isolated on the basis of its methylated bases showed significant similarity to the forkhead (FH) DNA-binding domain proteins. Among the FH protein family, particularly high homology within the FH domain is observed with the AFX and FKHR proteins, whose genes are recombined following chromosomal translocations in human malignancies. AFX was isolated as an MLL translocation partner following the study of a t(X;11)(q13;q23) translocation in a human ALL cell line. The FKHR gene is fused to PAX3 in t(2;13) (q35;q14) and to PAX7 in t(1;13)(p36;q14) translocations associated with alveolar rhabdomyosarcoma. Amino acid alignment of the FH domain of these three FH proteins is shown in Fig 5, and a derived consensus sequence is compared with a previously defined HNF3 FH consensus. AF6q21, AFX, and FKHR share several features that differ from other FH proteins, as previously noted. They lack a KPPY tetrapeptide at the NH2 side of the FH domain, and exhibit several additional or missing amino acids with respect to the FH consensus (Fig 5).
**Fig 5.** Comparison of AF6q21, AFX, and FKHR amino acid sequences. Top: Homology region within the NH2 part of the genes. Bottom: Alignment of FH domains. A HNF3 FH consensus is shown for comparison. Only amino acids identical in AF6q21, AFX, and FKHR appear in the consensus. Gaps introduced to maximize alignment are shown as dashes. Asterisks indicate identity between the FH consensus defined in this study and the HNF3 FH consensus.16 Numbering is with respect to the AF6q21 protein starting at the second in-frame ATG.

**FISH studies with AF6q21 probes.** FISH experiments were performed to confirm the localization of the breakpoint on chromosome 6. The two phases isolated as indicated earlier were first hybridized to metaphase chromosomes of a normal control and localized to band 6q21. FISH was then performed on metaphase chromosomes of the patient. The phage probe T10 (5' part of the gene) hybridized to the t(2;13) and t(1;13), characteristic of alveolar rhabdomyosarcoma.14,15 The amino acid sequence of the FH domain of der(6) chromosome, and T3 (3' part) to rearranged 11. This confirmed the orientation as 5' centromere ± 3' telomere of AFX is closely related to those of the human FH gene, the hepatocyte transcription factors, and the oncogene AF6q21. In the patient’s metaphase, hybridization signals were also observed on normal chromosome 6 with the two probes. Consistently, with somatic cell hybrid analysis, a secondary signal was also observed on chromosome band 17p11 (data not shown).

**DISCUSSION**

Various chromosomal rearrangements involving band 11q23 have been described in hematopoietic disorders.1,19 Most rearrangements found in leukemias involve the MLL gene, which is fused to various partner genes in diverse translocations. Ten different fusion types have been described to date.1 Translocation t(6;11)(q21;q23) has rarely been reported in AL, whereas t(6;11)(q27;q23) fusing the MLL and AF6 genes appears to be relatively more common. Interestingly, the undifferentiated myeloid leukemia analyzed here was therapy-related, since it appeared after treatment for Hodgkin’s disease. Chromosomal abnormalities found in leukemia after Hodgkin’s disease frequently involve chromosomes 7 and 5, but abnormalities of band 11q23 rearranging MLL are also known to occur in secondary leukemia.20

The AF6q21 gene contains a FH DNA binding domain. The FH gene was initially described as a homeogene encoding a nuclear protein expressed in the extremities of the Drosophila embryo.21 The hepatocyte nuclear factor 3α containing a FH domain was then isolated from rat liver and shown to be a transcription factor.22 Other transcription regulator genes encoding a FH domain protein were then discovered in eukaryotes, including man.23,24 Two genes, AFX and FKHR, have recently been repeated to be involved in chromosomal translocations of human malignancies. The first, AFX, is fused to MLL in t(X;11) (q13;q23) of AL.6,12,13 The second gene, FKHR, is fused to members of the paired box or PAX transcription factor family, PAX3 and PAX7, as a result of two translocations, t(2;13) and t(1;13), characteristic of alveolar rhabdomyosarcoma.14,15 The amino acid sequence of the FH domain of AFX is closely related to those of the human FH gene, the hepatocyte transcription factors, and the oncogene AF6q21. It is striking but could be partly explained by the structure of these genes, since all translocations seem to occur within the first intron of the genes.13,18 FH homologues have been found to be expressed in a lineage-restricted manner in human hematopoietic cell lines.27 The FH genes are defined as belonging to the family of transcription regulator genes, but their functions in hematopoietic cells remain to be established, as well as the functional consequences of the fusion genes of malignant cells. The Drosophila gene trithorax, which has homology with the human MLL gene, has recently been claimed to directly regulate the region-specific homeotic gene FH.28 Whether a similar interaction exists in human homologous genes is not known, nor is it known if such regulation could favor the occurrence of translocations between MLL and genes possessing FH domains.

Homology regions lying outside of the FH domain have been noted between several FH protein subfamilies.1,19,26 Interestingly, an NH2 region conserved within the HNF3 protein family has been demonstrated to contribute to the transcriptional activation properties of HNF3α.29 The functional significance of the similarities observed between AFX, AF6q21, and FKHR gene products remains unknown. However, their existence suggests that they represent a particular FH protein subgroup particularly involved in chromosomal translocations of human malignancies.

Possible FKHR homologous genes have been localized to
human chromosomes 5, 6, 13, 17, and X. Our preliminary data suggest that an AF6q21-like sequence is assigned to chromosome band 17p11. It is worth determining if these sequences correspond to true AF6/FKHR/AF6q21 homologous genes, and if they are also involved in human malignancy.

ACKNOWLEDGMENT

We sincerely thank Prof B.D. Young for fruitful discussion. We also thank Dr M. Lanotte for the Northern blot membranes.

REFERENCES

11. Kozak M: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283, 1986
20. Pedersen-Bjergaard J, Philip P: Balanced translocations involving bands 11q23 and 21q22 are highly characteristic of myelodysplasia and leukemia following therapy with cytostatic agents targeting at DNA-topoisomerase II. Blood 78:1147, 1991
AF6q21, a Novel Partner of the MLL Gene in t(6; 11)(q21; q23), Defines a Forkhead Transcriptional Factor Subfamily

Josette Hillion, Maryvonne Le Coniat, Philippe Jonveaux, Roland Berger and O.A. Bernard