Leucine-zipper dimerization motif encoded by the *AF17* gene fused to *ALL-1* (*MLL*) in acute leukemia

[t(11;17) chromosome translocation/oncogenes/protein–protein interactions]

R. PRASAD*, D. LESHKOWITZ†, Y. GU*, H. ALDER*, T. NAKAMURA*, H. SAITO*, K. HUEBNER*, R. BERGER‡, C. M. CROCE*, AND E. CANAANI*†§

*Jefferson Cancer Institute, Jefferson Cancer Center, and Department of Microbiology and Immunology, Thomas Jefferson University, Philadelphia, PA 19107;

†Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel; and ‡Institut National de la Santé et de la Recherche Médicale, Institut de Genetique Moléculaire, Paris 75010, France

Communicated by Sidney Weinhouse, April 28, 1994

ABSTRACT Chromosome region 11q23 is involved in reciprocal chromosome translocations associated with human acute leukemias. These aberrations fuse the *ALL-1* gene located at 11q23 to a series of partner genes positioned on a variety of human chromosomes. The fused genes encode chimeric proteins. Here we report the cloning and characterization of the *ALL-1* partner at 17q21, the *AF17* gene. The *AF17* gene encodes a protein of 1093 amino acids, containing a leucine-zipper dimerization motif located 3' of the fusion point and a cysteine-rich domain at the N terminus. The latter can be arranged in three zinc fingers and shows homology to a domain within the protein Br140 (peregrin). *AF17* contains stretches of amino acids previously associated with domains involved in transcriptional repression or activation. Based on features of *AF17* and of the proteins encoded by the other partner genes analyzed and in conjunction with other recent studies, we propose a model in which *ALL-1* rearrangements result in loss of function of the gene. In this model, the partner polypeptide plays an accessory role either by repressing activity of the truncated *ALL-1* protein or by blocking the function of the normal protein presumably present in the leukemic cells.

Chromosome translocations involving band 11q23 are associated with ≈10% of patients with acute lymphoblastic leukemia (ALL) and >5% of patients with acute myeloid leukemia (AML) (1). The most common translocations are t(4;11), t(11;19), t(9;11), and t(6;11); less-frequent aberrations involve specific regions on chromosomes 1, 2, 5, 10, 12, 15, 17, 18, 22, and X (2, 3). The ALL patients are often very young, <1 year of age; 11q23 aberrations are the most common cytogenetic alteration among infants with acute leukemia (4). These aberrations are also common in secondary myeloid leukemias developing in children with ALL treated with epipodophyllotoxins (5). Leukemias associated with 11q23 abnormalities show distinct features, such as frequent expression of both lymphoid and myeloid markers, massive cell burden, and poor prognosis (6, 7).

The gene at 11q23 was recently cloned (8–10) and designated *ALL-1* (8, 9), *HRX* (10), *MLL* (11), or *Htrx* (12). *ALL-1* has regions of homology with the *Drosophila* trithorax gene and probably is the human homologue of the latter. Trithorax is involved in segment determination in *Drosophila* through positive regulation of the genes of the Antennapedia and Bithorax complexes. The *ALL-1* protein contains at least four domains that are likely to interact with target genes: two centrally located domains with multiple zinc fingers, several AT hooks at the N terminus of the protein, and a motif positioned upstream of the zinc fingers and implicated in

protein–DNA interactions that may discriminate between methylated and unmethylated DNA (13).

11q23 aberrations cleave *ALL-1* within a breakpoint cluster region of 8.3 kb encompassing exons 5–11 (8, 9) and fuse it to specific genes on the partner chromosomes. This results in production of fused RNAs encoding chimeric proteins (9, 10). Each translocation gives rise to two reciprocal products; however, the identification in some patients of *ALL-1* transcripts originating from the derivative chromosome 11, der(11), but not from the reciprocal derivative (14), implicates the fusion product containing the N terminus of *ALL-1* as the oncogene. This segment of *ALL-1* encompasses the AT hooks and the motif with homology to methyltransferase but does not include the bulk or any of the zinc fingers. To date, the partner genes from chromosomes 1, 4, 6, 9, and 19 were cloned and designated *AF1P*, *AF4*, *AF6*, *AF9*, and *ENL*, respectively (10, 15–18). In this paper, we describe the properties of *AF17*, the fusion partner on chromosome 17, and discuss a possible role for the partner-gene-encoded polypeptides in conferring a malignant potential to the chimeric proteins.

MATERIALS AND METHODS

Patients and Cells. AML patients, GUS and GE, showed the chromosome translocation t(11;17) (q23;q21) in their leukemic cells. The cell lines used for RNA analysis included K562 and KCl-22 (erythroid and myeloid acute phase of chronic myeloid leukemia); MV4:11 and B-1 (ALLs with the 4;11 translocation); 380, ALL-1, 697, and GM607 (ALLs); GM1500 (Epstein–Barr virus-transformed lymphoblastoid cell line); T98G (glioblastoma); and PC3 (prostate carcinoma). These cell lines are described in refs. 8 and 16.

Cloning and Sequencing. The junction fragment of patient GUS was cloned from a library prepared from a partial digest of genomic DNA cloned into the EMBL-3 phage vector. The library was screened with a 0.86-kb *Bam*HI cDNA probe spanning *ALL-1* exons 5–11. cDNA libraries were prepared from ALL-1 and KCl-22 cytoplasmic RNAs utilizing a kit manufactured by Pharmacia and the λ ZAPII vector of Stratagene. Reverse transcription–PCR was performed as described (15) utilizing as primers an *ALL-1* oligonucleotide with a *Bam*HI site attached at the 5' end (CGGGATCCGACCTACTACAGGACCGCCAAG) and an *AF17* oligonucleotide with an *Eco*RI site at the 5' end (ATCTGAAT-TCTGGTGGAGATAGAAGCAGAA). Sequencing was per-

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia.

§Present address: Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. U07932).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

formed in the Applied Biosystem automatic sequencer with cDNAs and genomic fragments excised from phage vectors and cloned into the SK plasmid vector. The sequence was analyzed using the FASTA, TFASTA, and MOTIFS programs.

RESULTS

DNA from patient GUS with AML and t(11;17) was partially digested with *Mbo* I enzyme, and after size selection was cloned into the EMBL-3 phage vector. The library was screened with a cDNA probe spanning the breakpoint cluster region. A clone composed of a rearranged *ALL-1* segment was identified among positive clones. Comparison between the physical maps of this clone and the corresponding normal *ALL-1* DNA (Fig. 1A) indicated that *ALL-1* sequences upstream of exon 6 were substituted with the non-*ALL-1* DNA; the latter was subsequently found to be derived from chromosome 17 (data not shown). Within the non-*ALL-1* segment of the junction clone, a 1.7-kb *Eco*RI fragment (R1.7) was found to be devoid of repetitive sequences. This fragment was used as a probe to analyze, by the Southern blot technique, genomic DNA from a second patient (GE) with AML and the t(11;17) aberration. In that DNA, we detected an 11.5-kb rearranged *Eco*RV fragment (Fig. 1B, lane b). This indicated that in both patients the breaks occurred in the same region on chromosome 17.

Fragment R1.7 was next used as a probe on cDNA libraries derived from RNAs of the cell lines KCL-22 and ALL-1. Inserts from positive clones were subcloned into the SK plasmid vector and mapped. Clones 1, 3, 13, and a4 (Fig. 2A) were subjected to sequencing analysis. *AF17* cDNA contains an open reading frame spanning 3279 nucleotides. The first ATG shows a good fit to a Kozak consensus sequence and is preceded by an in-frame termination codon. The predicted protein spans 1093 amino acids. It contains relatively high concentrations of Ser, Gly, Ala, Leu, and Pro (15, 11, 10, 10, and 10%, respectively) often concentrated in short stretches. In addition, it has a Gln-rich region (41%) between amino acids 935 and 984 (Fig. 2B). The same region shows high concentration of hydrophobic amino acids, in particular Leu. We note that domains rich in Ala (19), Gly (20), Gln, and Pro (21) were implicated in transcriptional repression. Also,

regions with high concentrations of Ser and Gln (22) or Pro (23) were found to be involved in transcriptional activation.

Homology search in GenBank indicated 90% identity over amino acids 45–139 between *AF17* and an anonymous human cDNA sequence (GenBank accession no. TO6113). Furthermore, 188 residues (Fig. 3A) of *AF17* showed 48% identity and 67% similarity to a region within the protein Br140 (34), previously named peregrin. This domain is Cys-rich in both proteins and can be arranged into three zinc fingers according to the consensus C-X₂-C-X₁₀₋₁₃-C-X₂₋₄-C (Fig. 3B). Related consensus sequences are present in the adenovirus E1A protein and in the steroid receptor superfamily (24). The human Br140 protein has a second Cys-rich domain and is located in the nucleus; the function of this protein is unknown. Inspection of *AF17*-predicted protein sequence revealed a leucine-zipper dimerization motif between amino acids 729 and 764 (Fig. 2B). Unlike many leucine zippers (25, 26), the one in *AF17* is not preceded by a basic region.

To prove that the *ALL-1-AF17* fused gene is transcribed into a chimeric RNA, we used cDNA and genomic DNA sequence information to design primers for amplification by reverse transcription-PCR of a putative *ALL-1-AF17* RNA junction from the leukemic cells of patient GUS. An amplification product was indeed found to contain the RNA junction (Fig. 2C). Within the fused RNA, the open reading frames of the two genes were found to be linked in phase. Thus, the t(11;17) abnormality results in production of an RNA encoding a chimeric *ALL-1-AF17* protein.

To examine the expression of the normal *AF17* gene, we performed a Northern blot analysis. A major transcript of 7.5 kb and a minor diffuse species of 5 kb were detected in a variety of hematopoietic and nonhematopoietic cell lines (Fig. 4).

DISCUSSION

The cloning and sequence analysis of the partner genes that recombine with *ALL-1* in 11q23 translocations provides information and reagents that can be used in the diagnosis, prognosis, and monitoring of human acute leukemias. In addition, this cloning enables construction of biologically active molecules and might provide insights into the mech-

FIG. 1. Genomic analysis of the t(11;17) chromosome translocation. (A) Physical map of the genomic junction of patient GUS [der(17)] and a map of the corresponding normal region (Chr. 11q23). Numbered open boxes in the top line represent *ALL-1* exons. Thickened segment of der(17) corresponds to chromosome 17 sequences, and the open box therein represents an exon. Fragment R1.7 was used as a probe for the genomic Southern blot analysis and for cDNA screening. Cen and Tel show directions of the centromeres and telomeres, respectively. R, *Eco*RI; H, *Hind*III; B, *Bam*HI; G, *Bgl* II; X, *Xba* I. (B) Southern blot genomic analysis of a DNA from patient GE with AML and t(11;17) and a normal DNA (lanes b and a, respectively). DNAs were digested with *Eco*RV and hybridized with the R1.7 probe. The germ-line fragment is 18 kb.

FIG. 2. Cloning and sequencing of *AF17* cDNA and of the junction within *ALL-1*–*AF17* fusion transcript. (A) Physical map of *AF17* cDNA clones. S, *Sac*I; H, *Hind*III; H2, *Hinc*II. Initiation (ATG) and termination (TAA) are shown by arrows. (B) Predicted amino acid sequence of *AF17* protein. Cys residues within the Cys-rich region at the N terminus are underlined. Also underlined is the leucine zipper at positions 729–764. Arrow indicates point of fusion with the *ALL-1* protein. (C) *ALL-1*–*AF17* RNA junction cloned from the leukemic cells of patient GUS.

anism of leukemogenesis. The most notable feature of *AF17* protein is the leucine-zipper protein dimerization motif. After t(11;17) chromosome translocation, this motif will be included in the *ALL-1*–*AF17* chimeric protein, which is presumed to be the critical product of the aberration. Since the leucine zipper of *AF17* is not preceded by a basic region required for interaction with DNA and because leucine zippers are found not only in transcription factors (25, 26) but also in other proteins with diverse functions (27–29), the function of *AF17* is currently not known.

AF17 is the sixth partner gene involved in 11q23 abnormalities to be cloned and characterized. Schematic representation of the proteins encoded by five of these genes and by *ALL-1* is shown in Fig. 5. Inspection of the sequences within the segments of the partner proteins (right side of the arrows) linked to *ALL-1* sequences (left side of the fusion point within the top scheme) in the chimeric proteins thought to be critical for leukemogenesis does not reveal a common motif. *AF9* and *ENL* are the only partner genes that share sequence homology (15). The two proteins are rich in Ser and Pro. *AF9* and *ENL* contain nuclear-targeting sequences and are probably nuclear proteins. The *AF4* polypeptide within the *ALL-1*–*AF4* protein includes several segments with a high concentration of Ser or Pro (15). The *AF4* protein includes a nuclear-targeting sequence and, therefore, is probably asso-

ciated with the nucleus. The *AF6* polypeptide linked to the N terminus of *ALL-1* spans segments with abundant Glu, Pro, or acidic amino acids. The polypeptide contains the GLGF motif (16) whose function is not known, embedded within a region of ≈600 amino acids with homology to the tail domain of the myosin heavy chain from *Dictyostelium discoideum*. This domain is thought to be involved in formation of filaments probably through protein–protein interactions. The GLGF motif is found in cytoplasmic or membrane-associated proteins and this suggests that *AF6* is located in the cytosol or membrane. The recently described *AF1P* gene fused to *ALL-1* in leukemic cells with the t(1;11) (p32;q23) abnormality (18) appears to be the human homologue of the murine *eps15* gene (30), which encodes a cytoplasmic protein containing an α -helical coiled-coil domain characteristic of cytoskeleton proteins and presumably is involved in protein–protein interactions. *AF1P* polypeptide within *ALL-1*–*AF1P* protein includes the α -helical coiled-coil region, as well as domains rich in Gln or acidic amino acids.

The high homology between *AF9* and *ENL* has previously prompted us to speculate (15) that the partner polypeptides are related and possibly contribute a similar function to the chimeric protein. One such possible function previously considered (10) was transcriptional activation. Domains with this activity or with transrepression activity were character-

FIG. 4. Northern blot analysis of AF17 RNA in human cell lines. Polyadenylated RNA (5–10 μ g) was analyzed on an agarose gel containing formaldehyde. RNAs were obtained from the following cell lines. Lanes: a, KCl-22; b, MV4;11; c, *ALL-1*, GM-607; d, B1; e, 380; f, PC3; g, GM-1500; h, K562; i, T93G; j, 679.

ized in a number of transcription factors and were found to be rich in particular amino acids such as Ser, Pro, Gln, acidic amino acids, Ala, or Gly (19–23). Domains rich in one or more of the amino acids are indeed found in all partner polypeptides, with the exception of the C-terminal region of AF9 linked to *ALL-1* in some of the patients with t(9;11) (15). We note though that the AF1P protein and presumably the AF6 protein are located in the cytoplasm or the membrane of the cell, and therefore, in their normal context, these proteins presumably do not play a role in transcriptional regulation. In

three of the partner polypeptides, we find the following domains known or thought to be involved in protein-protein interaction (dimerization): the leucine zipper in AF17, the α -helical coiled-coil domain of AF1P, and the region with homology to the tail domain of myosin heavy chain (which includes the GLGF motif) in AF6.

Recently, several patients with AML were identified in which the *ALL-1* gene was rearranged through partial tandem duplication, with no involvement of a partner gene (31). We think that this finding, in conjunction with the recent data about the multitude and variety of partner genes, can be best interpreted by a model in which the cleavage of the ALL-1 protein results in loss of its function and that the partner polypeptide augments this loss. The partner polypeptide can contribute to the inactivation by supplying a transcriptional repression domain in cis. Alternatively, the partner polypep-

FIG. 5. Landmarks, common motifs, and homologous sequences within the partner proteins AF4, AF9, ENL, AF6, and AF17 and within the ALL-1 protein. Arrows indicate fusion points between ALL-1 and the partner proteins. Hatched regions in AF9 and ENL indicate domains of highest homology between the two proteins. NTS, nuclear targeting sequence; LZ, leucine zipper; MTase, methyltransferase.

tide might act through its dimerization domain to form homodimers. The latter could bind more efficiently to the target DNA, thereby blocking binding of the normal ALL-1 protein presumably present in leukemic cells. It is also possible that the dimer may be able to form a more stable complex with transcription factors necessary for activity of the normal ALL-1 protein and consequently sequester these factors. Tandem duplication of the N-terminal segment of ALL-1 (31) might be sufficient for inactivation of the latter as well as for blocking of the normal protein.

Although chromosome translocations are usually associated with overexpression or activation of oncogenes, there is a recent example for a translocation that apparently involves loss of function and a dominant negative effect. Thus, in the t(15;17) chromosome translocation associated with acute promyelocytic leukemia, a major effect of the fusion protein PML-RAR is sequestering of the normal PML protein and inhibiting its organization into nuclear macromolecular organelles (32, 33).

R.P. and D.L. contributed equally to this work and should both be considered first authors. This work was supported by a grant from the National Cancer Institute (CA 39860) and by grants from the Levy Foundation at the Weizmann Institute, the Israeli Academy of Science, and Israel Cancer Research Fund.

- Heim, S. & Mitelman, F. (1987) *Cancer Cytogenetics* (Liss, New York).
- Mitelman, F. (1991) *Catalog of Chromosome Aberrations in Cancer* (Liss, New York).
- Thirman, M. J., Gill, H. J., Burnett, R. C., Mbangkollo, D., McCabe, N. R., Kobayashi, H., Ziemer-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A. A., Chaganti, R. S. K., Larson, R. A., Le Beau, M. M., Diaz, M. O. & Rowley, J. D. (1993) *N. Engl. J. Med.* **329**, 909–914.
- Kaneko, Y., Shikano, T., Maseki, N., Sakurai, M., Takeda, T., Hiyoshi, Y., Miyama, J.-I. & Fujimoto, T. (1988) *Leukemia* **2**, 672–676.
- Pui, C.-H., Behm, F. G., Raimondi, S. C., Dodge, R. K., George, S. L., Rivera, G. K., Mirro, J., Kalwinsky, D. K., Dahl, G. V., Murphy, S. B., Crist, W. M. & Williams, D. L. (1989) *N. Engl. J. Med.* **321**, 136–142.
- Hudson, M. M., Raimondi, S. C., Behm, F. G. & Pui, C.-H. (1991) *Leukemia* **5**, 1064–1068.
- Arthur, D. C., Bloomfield, C. D., Lindquist, L. L. & Nesbit, M. E. (1982) *Blood* **59**, 96–99.
- Cimino, G., Moir, D. T., Canaani, O., Williams, K., Crist, W. M., Katzav, S., Cannizzaro, L., Lange, B., Croce, C. M. & Canaani, E. (1991) *Cancer Res.* **51**, 6712–6714.
- Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M. & Canaani, E. (1992) *Cell* **71**, 701–708.
- Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71**, 691–700.
- Ziemer-van der Poel, S., McCabe, N. R., Gill, H. J., Espinosa, R., Patel, Y., Harden, A., Rubinelli, P., Smith, S. D., LeBeau, M. M., Rowley, J. D. & Diaz, M. D. (1991) *Proc. Natl. Acad. Sci. USA* **86**, 10735–10739.
- Djabali, M., Salleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nat. Genet.* **2**, 113–118.
- Ma, Q., Alder, H., Nelson, K. K., Chatterjee, D., Gu, Y., Nakamura, T., Canaani, E., Croce, C. M., Siracusa, L. D. & Buchberg, A. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6350–6354.
- Downing, J. R., Head, D. R., Raimondi, S. C., Carroll, A. J., Curcio-Brint, A. M., Motroni, T. A., Hulshaf, M. G., Pullen, D. J. & Dimer, P. H. (1994) *Blood* **83**, 330–335.
- Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R. P., Lange, B., Crist, W. M., Nowell, P. C., Croce, C. M. & Canaani, E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4631–4635.
- Prasad, R., Gu, Y., Alder, H., Nakamura, T., Canaani, O., Saito, H., Hubner, K., Gale, R. P., Nowell, P. C., Kuriyama, K., Miyazaki, Y., Croce, C. M. & Canaani, E. (1993) *Cancer Res.* **53**, 5624–5628.
- Morrissey, J., Tkachuk, D. C., Milatovich, A., Francke, U., Link, M. & Cleary, M. (1993) *Blood* **81**, 1124–1131.
- Bernard, O. A., Mauchauffe, M., Mecucci, C., Van Den Berghe, H. & Berger, R. (1994) *Oncogene* **9**, 1039–1045.
- Licht, J. D., Grossell, M. J., Figge, J. & Hansen, U. M. (1990) *Nature (London)* **346**, 76–79.
- Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. (1991) *Cell* **67**, 377–388.
- Madden, S. L., Cook, D. M., Morris, J. F., Gashler, A., Sukhatmi, V. P. & Raucher, F. J. (1991) *Science* **253**, 1550–1553.
- Mitchel, P. J. & Tjian, R. (1989) *Science* **245**, 372–378.
- Theill, L. E., Castrillo, J.-L., Wu, D. & Karin, M. (1989) *Nature (London)* **342**, 945–948.
- Schmidt, M. H. & Berg, J. M. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 599–613.
- McKnight, S. L. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 771–795.
- Curran, T. & Vogt, P. K. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 797–831.
- Buckland, R. & Wild, F. (1992) *Nature (London)* **338**, 547.
- McCormack, K., Campanelli, J. T., Ramaswami, M., Mathew, M. K., Tanouge, M. A., Iverson, L. E. & Rudy, B. (1989) *Nature (London)* **340**, 103.
- White, M. K. & Weber, M. J. (1989) *Nature (London)* **340**, 103–104.
- Fazioli, F., Minichiello, L., Matoskova, B., Wong, W. T. & Di Fiore, P. P. (1993) *Mol. Cell. Biol.* **13**, 5814–5828.
- Schichman, S. A., Caligiuri, M. A., Gu, Y., Strout, M. P., Canaani, E., Bloomfield, C. D. & Croce, C. M. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6236–6239.
- Dyck, J. A., Maul, G. G., Miller, W. H., Don Chen, J., Kakizuka, A. & Evans, R. M. (1994) *Cell* **76**, 333–343.
- Weis, K., Ramboud, S., Lavou, C., Jansen, J., Carvalko, T., Carmo-Fonseca, M., Lamond, A. & Dejean, A. (1994) *Cell* **76**, 345–356.
- Thompson, K. A., Wang, B., Argraves, W. S., Giancotti, F. G., Schranck, D. P. & Ruoslahti, E. (1994) *Biochem. Biophys. Res. Commun.* **198**, 1143–1152.