Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4631–4635, May 1993
Genetics

Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs

T. NAKAMURA*, H. ALDER*, Y. GU*, R. PRASAD*, O. CANAANI*, N. KAMADA†, R. P. GALE‡, B. LANGE§, W. M. CRIST¶, P. C. NOWELL‖, C. M. CROCE*, AND E. CANAANI*

*Jefferson Cancer Institute and Department of Microbiology and Immunology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA 19107; †Department of Hematology, Hiroshima University, Hiroshima 739 Japan; ‡Department of Medicine, University of California, Los Angeles, CA 90024; §Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104; ¶Department of Hematology, St. Jude Children's Research Hospital and the University of Tennessee, Memphis College of Medicine, Memphis, TN 38101; and ‖Department of Pathology and Laboratory of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Contributed by P. C. Nowell, February 26, 1993

ABSTRACT **Chromosome translocations involving band 11q23 are associated with human acute leukemias. These translocations fuse the ALL-1 gene, homolog of *Drosophila* trithorax and located at chromosome band 11q23, to genes from a variety of chromosomes. We cloned and sequenced cDNAs derived from transcripts of the AF-4 and AF-9 genes involved in the most common chromosome abnormalities, t(4:11)(q21:q23) and t(9:11)(p22:q23), respectively. Sequence analysis indicates high homology between the AF-9 gene protein product and the protein encoded by the ENL gene fused to ALL-1 in (11:19) chromosome translocations. AF-4, AF-9, and ENL proteins contain nuclear targeting sequences as well as serine-rich and proline-rich regions. Stretches abundant in basic amino acids are also present in the three proteins. These results suggest that the different proteins fused to ALL-1 polypeptide(s) provide similar functional domains.**

Specific chromosome abnormalities, in particular reciprocal translocations, are the hallmark of most human leukemias and lymphomas. Many of the genes adjacent to the breakpoints and altered by these translocations were cloned and characterized (for reviews, see refs. 1 and 2). Of those, most genes associated with acute leukemias code for transcription factors and play a role in normal differentiation processes (for review, see ref. 3). The ALL-1 gene, a recent addition to this class of genes (4, 5), has some striking features. ALL-1 is located at 11q23, a region found to be involved in a variety of reciprocal chromosome translocations associated with lymploid and myeloid acute leukemias. The reciprocal breakpoint regions include 4q21, 9p22, 1q21, 1p32, 2p21, 5q31, 6q27, 10p11, 17q25, 19p13, and Xq13, with the first two being the most common (6). The 11q23 abnormalities are associated with the vast majority of infant acute leukemias. In addition, treatment of malignant disorders with epipodophyllotoxin appears to be linked to the development of secondary leukemias with 11q23 aberrations (7). ALL-1 is rearranged in most or all of 11q23 abnormalities (8–11). The gene codes for a protein of 3972 amino acids. The protein contains three regions with strong homology to *Drosophila* trithorax protein, two of which encompass multiple zinc fingers with a possible role in DNA binding. This suggested that the human and *Drosophila* genes may be homologs. The human gene alone contains a second potential DNA binding domain with 3 "AT hook" motifs. Very recently it was possible, by further computer search, to identify a third DNA binding motif, absent from the *Drosophila* protein and homologous to a region in mammalian DNA methyltransferases (12). 11q23 translocations occur within an ≈8-kb region of ALL-1; the rearrangements separate the second and third potential DNA binding motifs from the zinc fingers of the ALL-1 protein.

As shown for t(4:11) and t(11:19), the translocations fuse ALL-1 to genes from chromosome 4 (AF-4) or chromosome 19 (ENL), respectively (4, 5). Both derivative chromosomes are transcribed into fused RNAs encoding chimeric proteins. These results suggested that all 11q23 abnormalities lead to production of chimeric ALL-1 proteins and raised the question whether the genes linked to ALL-1 provide functional common domains or whether they simply supply initiation or termination signals for translation of the disrupted ALL-1 coding region. To answer this question, we cloned and characterized the gene AF-4 from chromosome 4 and the gene AF-9 from chromosome 9. These genes fuse to ALL-1 in t(4:11) and t(9:11) abnormalities, which represent the most common translocations involving band 11q23.**

MATERIALS AND METHODS

Construction of cDNA libraries from K562 or fibroblast RNA was described (13, 14). cDNA library was prepared from KCl 22 RNA in a similar way. AF-4 cDNA clones k1.1, k1.2, k11, and k12 originated from the K562 library and the clones kcl 6, kcl 10, and kcl 12 were cloned from the KCl 22 library. AF-9 cDNA clones v4 and v7 were obtained from the fibroblasts library, and k 16 was cloned from the K562 library.

For reverse transcription-polymerase chain reactions (RT-PCRs), 2 μg of RNA from patient FI were reverse transcribed in a reaction utilizing the AF-9 oligonucleotide TCCTCAGGATGTTCCAGATGT or the ALL-1 oligonucleotide GGCTCACAACAGACTTGGCAA as primers. The cDNAs were amplified with Taq 1 polymerase using the same primers together with the ALL-1 primer ACCTACTACAGGACCGCCAAG and the AF-9 primer CAGATGAAGTGGAGGATAACG, respectively. The reaction products were cloned into the SK plasmid vector, and recombinants with AF-9/ALL-1 or ALL-1/AF-9 DNA were sequenced. The AF-9/ALL-1 RNA junction of patient CO was obtained in a similar way using the ALL-1 primer CAGCGAACACACTTGGTACAG for synthesis of cDNA and the same primer together with the AF-9 primer CAACGTTACCGCCATTTGAT for PCR amplification.

RESULTS

Cloning and Sequencing AF-4 cDNA. cDNA clones containing the two reciprocal ALL-1/AF-4 RNA junctions were

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.

Abbreviations: AML, acute myeloid leukemia; ALL, acute lymphocytic leukemia; ORF, open reading frame; RT-PCR, reverse transcription polymerase chain reaction; NTS, nuclear targeting sequence.

**The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L13773, L13743, and L13744).

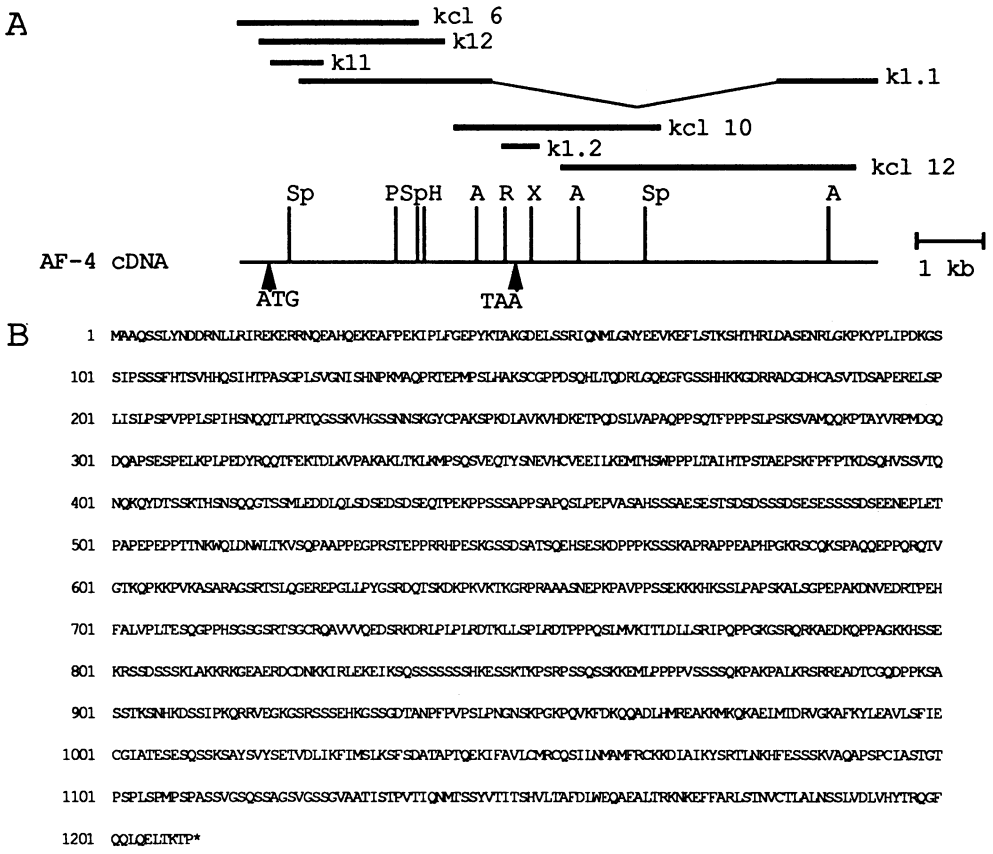

FIG. 1. Physical map of AF-4 cDNA and the predicted amino acid sequence of the protein encoded by it. (*A*) cDNA clones used for the analysis. The thin line in k1.1 represents sequences excluded from the clone, probably by modified splicing. Numbering begins at the first Met within the ORF of clone kcl 6. Sp, P, H, A, R, and X correspond to sites for the enzymes *Spe* I, *Pst* I, *Hin*dIII, *Acc* I, *Eco*RI, and *Xba* I, respectively. Complete nucleotide and amino acid sequences as well as the 5′ sequence of clone k12 were deposited in GenBank.

previously cloned from RNA of the RS4:11 cell line carrying the t(4:11) chromosome translocation (4). AF-4-specific probes obtained from these clones were used to screen cDNA libraries prepared from RNAs of the K562 (15) and KCl 22 (16) hematopoietic cell lines. Positive clones were sequenced and utilized to prepare end probes for further screening. Overlapping clones spanning most or all of the 9.5-kb AF-4 transcript (4) were obtained. Analysis of the longest cDNA composite indicated an open reading frame (ORF) initiated with a consensus ATG and coding for a protein of 1210 amino acids (Fig. 1). cDNA clone k12 diverged from cDNA clone kcl 6 at the fifth codon of the latter; 5′ of the sixth codon in

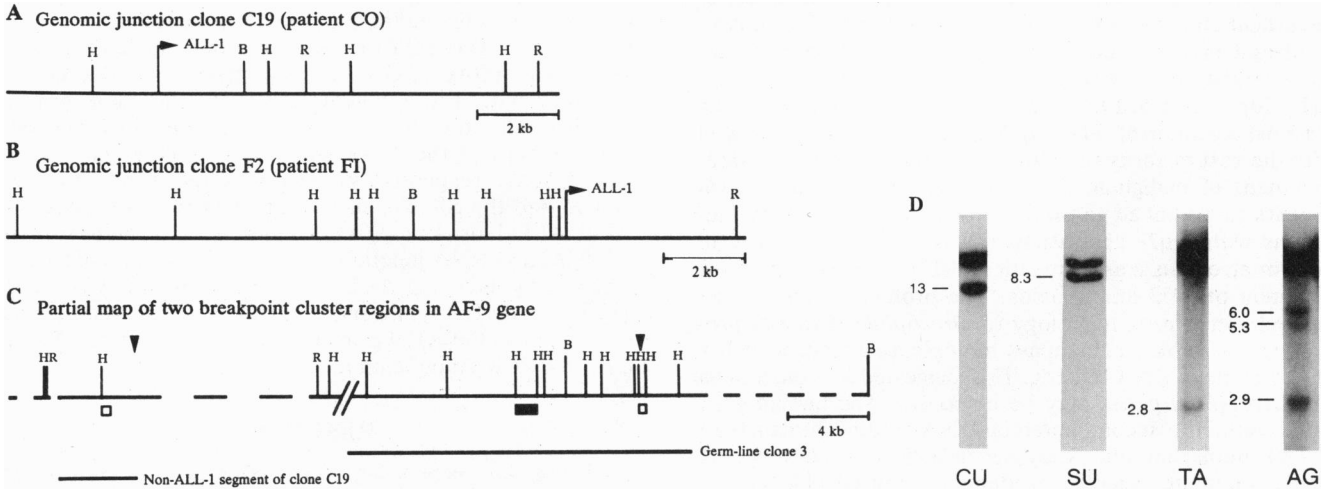

FIG. 2. AF-9/ALL-1 genomic junctions in two patients (*A*, *B*), relative orientation of the two break cluster regions in AF-9 (*C*), and rearrangement of the AF-9 gene in patients with t(9:11) abnormalities (*D*). H, B, and R correspond to sites for the *Hin*dIII, *Bam*HI, and *Eco*RI enzymes, respectively. (*C*) Regions left and right of the slash were not mapped with *Bam*HI and *Eco*RI, respectively. Arrowheads correspond to fusion points with the ALL-1 gene. Open and closed boxes correspond to positions of probes used to detect rearrangements or to clone cDNA (see text). Interrupted line represents a noncloned part of AF-9, mapped by Southern analysis of human DNA. (*D*) Numbers correspond to size in kb of rearranged fragments. DNAs of patients CU and SU were digested with *Bam*HI and *Eco*RI, respectively. DNAs of patients TA and AG were digested with *Hin*dIII.

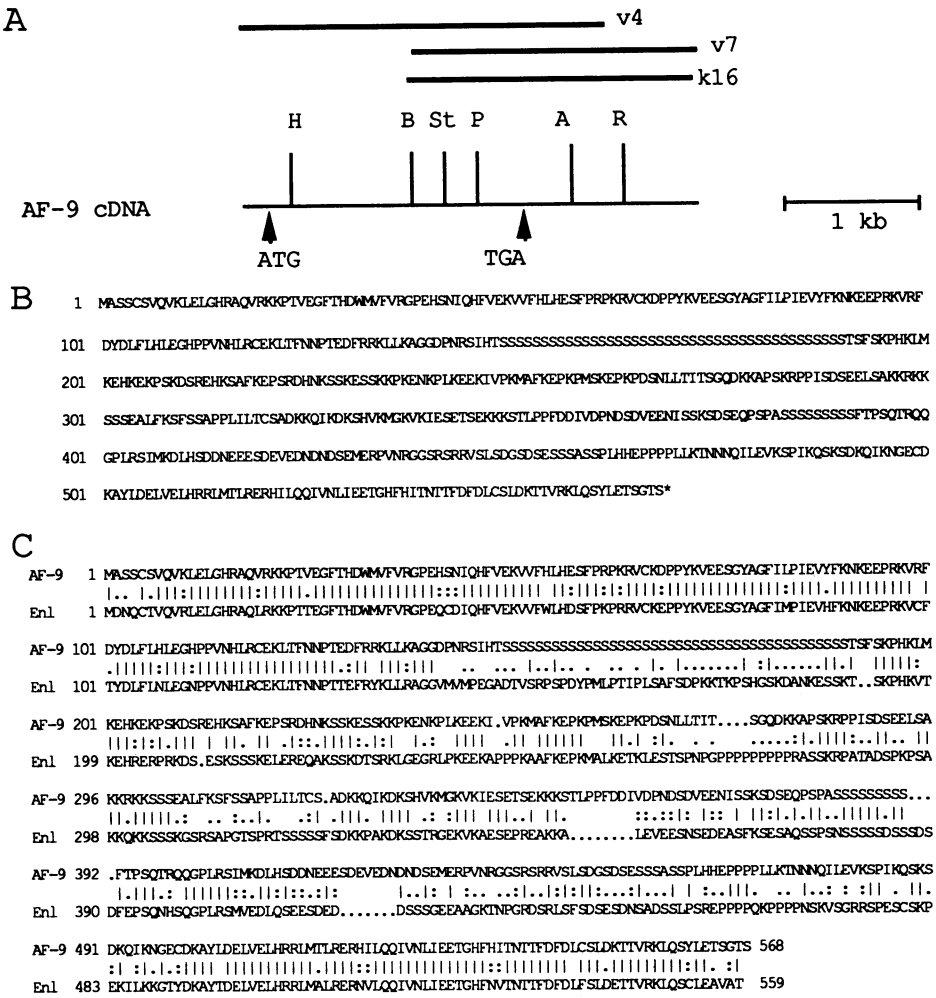

FIG. 3. Physical map of AF-9 cDNA (*A*), predicted amino acid sequence of AF-9 protein (*B*), and homology of the AF-9 and ENL proteins (*C*). H, B, St, P, A, and R in *A* correspond to sites for the enzymes *Hin*dIII, *Bam*HI, *Stu* I, *Pst* I, *Acc* I, and *Eco*RI, respectively. In *C*, bars indicate identical residues; single dots and double dots indicate first and second degree conservative differences, respectively. The sequence of ENL is from ref. 5. Nucleotide and amino acid sequences of AF-9 were deposited in GenBank.

kcl 6 the two sequences completely varied. The ORF of clone k12 started 12 codons 5′ to the divergence point. This suggests an alternative first exon for AF-4. cDNA clone k1.1 represents another RNA variant probably resulting from alternative splicing (Fig. 1); an in-frame termination codon is present in this clone immediately 3′ to the divergence point (not shown). Thus, AF-4 encodes two or more proteins varying at their termini. AF-4 contains an unusually long 3′ untranslated region of 5.3 kb. This region includes multiple AATAAA sequences located 20 nucleotides 5′ of the poly(A) as well as in several upstream positions; it also contains several stretches of T.

Using the Swiss, Prosite, and Profilescan data bases, the complete AF-4 protein sequence was searched for homology to other proteins and for the presence of motifs. The sequence AKKRK at positions 811–815 matched the consensus nuclear targeting sequence (NTS) -(RKTA) KK (RQNTSG) K- (17). AF-4 was relatively rich in serine (16%) and proline (11%) compared to the average frequency of these amino acids (7.1% and 4.6%, respectively).

Inspection of AF-4 sequence at the fusion point to ALL-1 RNA in the RS4:11 cell line (4) indicates that three nucleotides (1959–1961) of AF-4 RNA are missing from cDNA clone 25 corresponding to ALL-1/AF-4 fused RNA; these nucleotides might have been excluded through an error in the splicing process, where an AG at positions 1960–1961 was mistaken to be the 3′ end of an intron.

Cloning of AF-9/ALL-1 Genomic Junctions. The nonavailability of cell lines with the t(9:11) abnormality made it impossible to obtain intact mRNA in amounts sufficient for preparation of a cDNA library and cloning from it fused ALL-1/AF-9 cDNA. To circumvent this problem, we first cloned (clone C19) the genomic junction fragment from the leukemic cells of patient CO with acute myeloid leukemia (AML) and t(9:11). We also cloned (clone F2) the genomic junction fragment from tumor cells of patient FI with acute lymphocytic leukemia (ALL) and t(9:11). Mapping and hybridization analysis of the non-ALL-1 segments within the two phage clones indicated no homology between them (Fig. 2 *A* and *B*). A 1-kb *Hin*dIII fragment from the non-ALL-1 region in clone F2 (closed box in Fig. 2*C*) was used to clone the corresponding normal DNA (clone 3 in Fig. 2*C*). A 0.4-kb *Hin*dIII fragment from clone 3 (open box on the right of Fig. 2*C*) and 0.4-kb *Hin*dIII–*Ava* II probe from clone C19 (open box on the left of Fig. 2*C*) hybridized to human DNA within Chinese hamster cell hybrids containing human chromosome 9 (not shown). This established that in both patients' DNAs the ALL-1 gene is linked to chromosome 9 sequences. Subsequent work (see below) showed that both sequences are included in a single gene, which we term AF-9, for ALL-1 fused gene from chromosome 9.

The same repeat-free fragments were used as probes for detecting rearrangements in DNAs from leukemic cells with t(9:11) chromosome translocations. Samples from three pa-

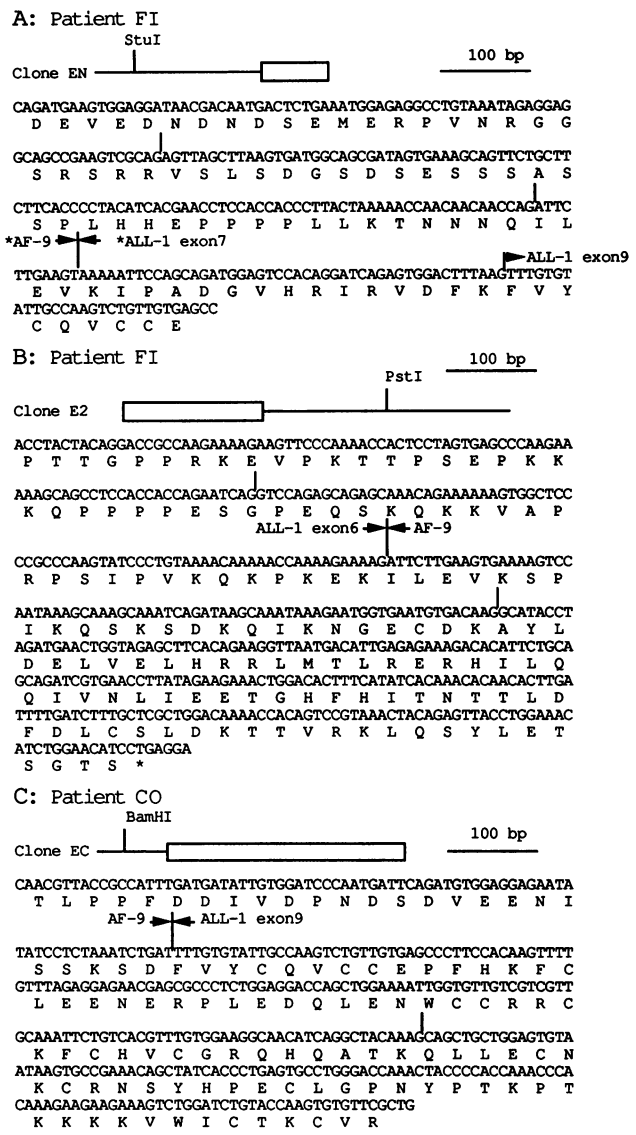

FIG. 4. Partial sequence of ALL-1/AF-9 chimeric transcripts. Boxes represent AF-9 information. Vertical bars correspond to exon–exon boundaries. The * in *A* indicates fusion in internal positions within exons (corresponding to the genomic fusion point in patient FI).

tients with ALL and from five patients with AML were studied. The 0.4-kb *Hin*dIII fragment detected rearrangement in DNA of the ALL patient CU (Fig. 2*D*). The *Hin*dIII–*Ava* II probe showed rearrangements in patients TA, SU, and AG, all with AML (Fig. 2*D*). This indicated that at least two regions in the AF-9 gene are involved in recurrent t(9:11) aberrations. Presently, it is not known whether one region is preferentially rearranged in AML and the second is rearranged in ALL; it is also not clear whether the AF-9 gene is involved in all t(9:11) abnormalities.

Characterization of Normal and Chimeric cDNAs of AF-9. We now examined repeat-free fragments from AF-9 DNA for hybridization to cDNA libraries. The 1-kb *Hin*dIII fragment (closed box in Fig. 2*C*) reacted with several overlapping cDNAs spanning 3.4 kb (Fig. 3*A*). These cDNAs reacted in Northern analysis with a major 5-kb transcript expressed in several hematopoietic cell lines (not shown). Nucleotide sequence analysis of AF-9 cDNA revealed an ORF beginning in a consensus initiation codon and coding for a protein of 568 amino acids (Fig. 3*B*). The protein encloses a nuclear targeting sequence, AKKQK, at positions 297–301. AF-9 protein is serine rich (20%) and includes a remarkable uninterrupted stretch of 42 serines at positions 149–190 [These serines are coded by the repetitive motif AGC; interestingly, a related unstable repetitive motif, CAG, was recently discovered within the coding region of the Huntington disease gene (18).] It also contains proline at a frequency of 7%, which is above the average frequency of 4.1%. Homology search showed, unexpectedly, that the predicted protein shared high similarity with the ENL protein. The latter is located on chromosome 19 and is fused to the ALL-1 gene in t(11:19) chromosome translocations (5). The two proteins show 56% identity and 68% similarity (Fig. 3*C*). The homology is highest within the 140 amino acids at the N terminus, where the proteins are 82% identical and 92% similar, and within the 67 amino acids at the C terminus, where the corresponding values are 82% and 91%.

To demonstrate chimeric ALL-1/AF-9 RNAs, we designed primers supposed to flank the RNA junction points in the two genes and used them in RT-PCRs with RNA from patient FI. Two reciprocal cDNA products were amplified (Fig. 4 *A* and *B*). Close examination of sequences at the RNA junctions showed a stretch of 11 nucleotides of AF-9 (ATTCTTGAAGT) at both RNA junctions. In an attempt to understand this, we sequenced the genomic junction in clone F2 and determined exon–intron boundaries of AF-9 exons in this region (not shown). This analysis suggested that the two derivative chromosomes of patient FI were formed by staggered breaks in the DNAs of chromosomes 9 and 11, resulting in a small overlapping AF-9 genomic DNA segment and consequently in the overlapping of 11 nucleotides of AF-9 at the RNA junction points (Fig. 4 *A* and *B*). The der(9) chromosome resulted from a break within exon 7 of ALL-1 and a break within an exon of AF-9 (11 nucleotides 3′ of the intron–exon boundary). The hybrid exon spans the fusion point in cDNA clone EN (ALL-1 exon 8 was skipped during splicing). The der(11) chromosome was due to a break in the other ALL-1 DNA strands within the intron flanked by exons 6 and 7 and to a breakage of the second AF-9 DNA strand within an intron located 5′ of the AF-9 exon mentioned above. The der(11) is transcribed into an RNA corresponding to cDNA clone E2.

A *Bam*HI–*Stu* I cDNA probe (Fig. 3*A*) detected some normal genomic fragments, which were also detected by the 0.4-kb *Hin*dIII–*Ava* II probe derived from the genomic junction cloned from DNA of patient CO (Fig. 2). This enabled designing primers predicted to flank the RNA fusion point of patient CO and use them in a RT-PCR to amplify AF-9/ALL-1 RNA (Fig. 4*C*). In this patient the AF-9 protein is linked at position 375 to the ALL-1 moiety, whereas in patient FI the junction point is at amino acid 477 or 481 of AF-9. In the three junctions examined, the reading frames of the two genes are joined in phase.

DISCUSSION

Perhaps the most unusual feature of 11q23 abnormalities is the multitude of chromosome partners participating in translocations with the ALL-1 locus. This promiscuity in partners for rearrangement and fusion could suggest that the only critical event in all these different translocations is the separation of one or more of the DNA binding domains in ALL-1 protein from a positive or negative regulatory element and that the proteins encoded by the partner genes solely provide initiation or termination codons.

Our sequence analysis of AF-4 and AF-9 proteins and a comparison to the sequence of the ENL protein are not consistent with such interpretation. The finding that AF-9 and ENL share extensive sequence homology indicates that the two proteins have similar biological function and that presumably they contribute an identical activity to the chi-

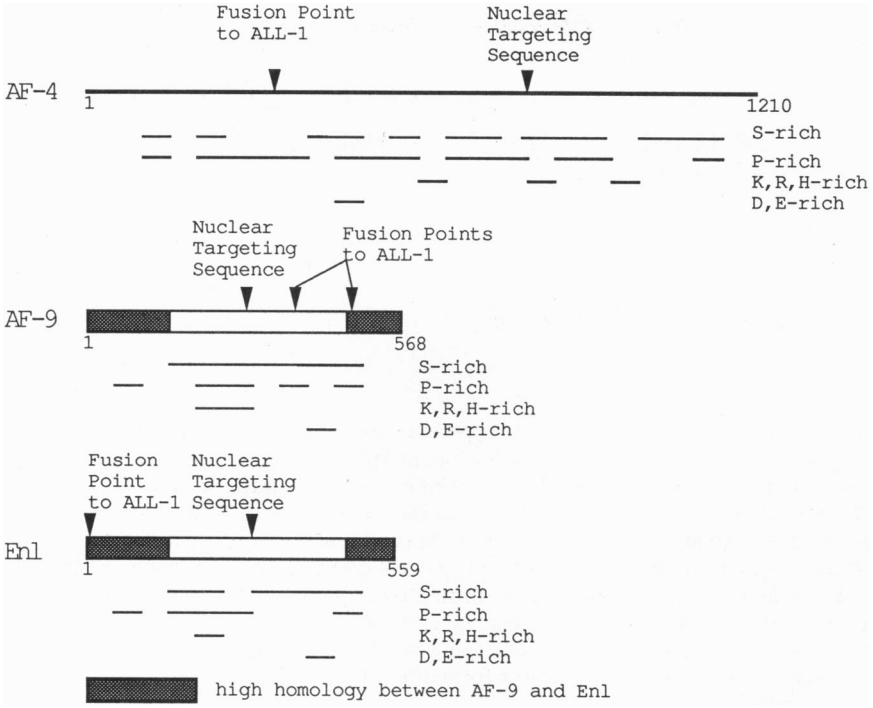

FIG. 5. Homologous sequences and common motifs in the proteins encoded by AF-4, AF-9, and ENL genes. Horizontal bars indicate stretches of 50 amino acids, in which the frequency of serine, proline, basic amino acids, or acidic amino acids is at least twice the frequency in an average protein (7.1%, 4.6%, 13.8%, and 11.7%, respectively).

meric proteins. Possibly, other genes participating in 11q23 aberrations also have sequence homology with AF-9 and ENL. Moreover, these two proteins share with AF-4 several common motifs (Fig. 5): (*i*) a NTS (suggesting that the three proteins are nuclear), (*ii*) serine-rich domains, the most prominent being an uninterrupted stretch of 42 serines in AF-9, (*iii*) stretches rich in proline or in basic amino acids reaching frequency of ≈30% in some regions. Domains with abundant prolines (19) or regions rich in serines and prolines (20) were shown to act as transcription activators, and domains rich in positively charged amino acids were found to bind DNA (for review, see ref. 19). These common structural motifs suggest that AF-4, AF-9, and ENL are involved in transcription regulation, possibly representing a new class of transcription factors. Proteins encoded by the other genes involved in 11q23 chromosome translocations might belong to this class.

Inspection of the position of the elements discussed above in relation to the fusion point(s) with the ALL-1 protein (Fig. 5) shows that the NTS of AF-4 is linked to the N terminus of ALL-1 containing the AT hooks, whereas AF-4 domains rich in serine, proline, or basic amino acids are fused to both reciprocal products of ALL-1 cleavage. In patient FI with t(9:11), the NTS and most of AF-9 domains rich in specific amino acids are linked to the C terminus of ALL-1, which contains the zinc fingers. In leukemia cells with t(11:19) all landmarks observed in the ENL protein will be linked to the N terminus of ALL-1; this may suggest that N-ALL-1/ENL-C is the oncogenic product of the t(11:19) abnormality. The opposite distribution of the common elements in AF-9 fusion products in patients such as FI raises the possibility that in these cases N-AF-9/ALL-1-C is the oncogenic species. Determination of which one (or both) of the fusion products of 11q23 translocations induces malignancy should be resolved by biological assays in cells in culture and in transgenic mice. Transcription assays utilizing elements of AF-4, AF-9, and ENL should help in understanding the normal function of these elements as well as their role in the fused proteins.

We thank Dr. G. Cimino for his contributions to the initial stage of this project and Dr. M.-L. Chu for the fibroblast cDNA library. This study was supported by an Outstanding Investigator Grant (CA39860) to C.M.C. from the National Cancer Institute.

1. Croce, C. M. (1987) *Cell* **49**, 155–156.
2. Solomon, E., Borrow, J. & Goddard, A. D. (1991) *Science* **254**, 1153–1160.
3. Rabbitts, T. H. (1991) *Cell* **67**, 641–644.
4. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, M. & Canaani, E. (1992) *Cell* **71**, 701–708.
5. Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71**, 691–700.
6. Heim, S. & Mitelman, F. (1987) *Cancer Cytogenetics* (Liss, New York).
7. Pui, C. H., Ribeiro, R. C., Hancock, M. L., Rivera, G. K., Evans, W. E., Raimondi, S. C., Head, D. R., Behm, F. G., Mahmoud, M. H., Sandlund, J. T. & Crist, W. M. (1991) *N. Engl. J. Med.* **325**, 1682–1687.
8. Cimino, G., Moir, D. T., Canaani, O., Williams, K., Crist, W. M., Katzav, S., Cannizzaro, L., Lange, B., Nowell, P. C., Croce, C. M. & Canaani, E. (1991) *Cancer Res.* **51**, 6712–6714.
9. Ziemin-van der Poel, S., McCabe, N. R., Gill, H. J., Espinose, R., III, Patel, Y., Harden, A., Rubinelli, P., Smith, S. O., LeBeau, M. M., Rowley, J. D. & Diaz, M. O. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10735–10739.
10. Cimino, G., Nakamura, T., Gu, Y., Canaani, O., Prasad, R., Crist, W., Carrol, A., Baer, M., Bloomfield, C. D., Nowell, P. C., Croce, C. M. & Canaani, E. (1992) *Cancer Res.* **52**, 3811–3813.
11. Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nature Genet.* **2**, 113–118.
12. 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*, in press.
13. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, E. (1985) *Nature (London)* **315**, 550–554.
14. Chu, M. L., Zhang, R. Z., Pan, T. C., Stokes, D., Conway, D., Kuo, J. H., Glanville, R., Mayer, U., Mann, K., Deutzmann, R. & Timp, R. (1990) *EMBO J.* **9**, 985–993.
15. Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* **45**, 321–324.
16. Kubonishi, I. & Miyoshi, I. (1983) *Int. J. Cell Cloning* **1**, 105–117.
17. Gomez-Marquez, J. & Segade, F. (1988) *FEBS Lett.* **226**, 217–219.
18. The Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971–983.
19. Mitchell, P. J. & Tjian, R. (1989) *Science* **245**, 372–378.
20. Theill, L. E., Castrillo, j.-L., Wu, D. & Karin, M. (1989) *Nature (London)* **342**, 945–948.