MOLECULAR AND CELLULAR BIOLOGY, Aug. 1992, p. 3346–3355
0270-7306/92/083346-10$02.00/0

Vol. 12, No. 8

can, a Putative Oncogene Associated with Myeloid Leukemogenesis, May Be Activated by Fusion of Its 3′ Half to Different Genes: Characterization of the *set* Gene

MARIEKE VON LINDERN,[1] SJOZÈF VAN BAAL,[1] JOOP WIEGANT,[2] ANTON RAAP,[2] ANNE HAGEMEIJER,[1] AND GERARD GROSVELD[1]*

Department of Cell Biology and Genetics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam,[1] and Department of Cytochemistry and Cytometry, University of Leiden, 2333 AL Leiden,[2] The Netherlands

Received 10 March 1992/Accepted 11 May 1992

The translocation (6;9)(p23;q34) in acute nonlymphocytic leukemia results in the formation of a highly consistent *dek-can* fusion gene. Translocation breakpoints invariably occur in single introns of *dek* and *can*, which were named *icb*-6 and *icb*-9, respectively. In a case of acute undifferentiated leukemia, a breakpoint was detected in *icb*-9 of *can*, whereas no breakpoint could be detected in *dek*. Genomic and cDNA cloning showed that instead of *dek*, a different gene was fused to *can*, which was named *set*. *set* encodes transcripts of 2.0 and 2.7 kb that result from the use of alternative polyadenylation sites. Both transcripts contain the open reading frame for a putative SET protein with a predicted molecular mass of 32 kDa. The *set-can* fusion gene is transcribed into a 5-kb transcript that contains a single open reading frame predicting a 155-kDa chimeric SET-CAN protein. The SET sequence shows homology with the yeast nucleosome assembly protein NAP-I. The only common sequence motif of SET and DEK proteins is an acidic region. SET has a long acidic tail, of which a large part is present in the predicted SET-CAN fusion protein. The *set* gene is located on chromosome 9q34, centromeric of c-*abl*. Since a *dek-can* fusion gene is present in t(6;9) acute myeloid leukemia and a *set-can* fusion gene was found in a case of acute undifferentiated leukemia, we assume that *can* may function as an oncogene activated by fusion of its 3′ part to *dek*, *set*, or perhaps other genes.

Translocations are the best-studied nonrandom chromosomal aberrations associated with specific subtypes of leukemia. As a result of a translocation, an oncogene can be activated through alterations in regulatory DNA sequences that leave the encoded protein intact (e.g., *myc*) or through formation of a fusion gene, encoding a chimeric protein (e.g., *bcr-abl*). The t(9;22) associated with chronic myeloid leukemia, acute myeloid leukemia (AML), and acute lymphoblastic leukemia (29) results in the expression of a chimeric BCR-ABL protein with enhanced tyrosine kinase activity (16, 19, 27, 38, 45). Pendergast et al. showed that defined sequences encoded by the first exon of *bcr* interact with the SH2 domain of ABL (33). This interaction is essential for the activation of the ABL tyrosine kinase activity and for the transforming capacity of BCR-ABL. More recently, other fusion genes have been isolated. t(1;19), occurring in childhood pre-B-cell acute leukemia, fuses the *E2a* gene, encoding transcription factors E12 and E47, to a novel homeobox gene, *PBX1* (26, 32). t(15;17), strongly associated with acute promyelocytic leukemia, fuses part of the retinoic acid receptor type α gene (*RARα*) to a novel gene on chromosome 15 named *PML*, which is predicted to be a transcription factor (9, 25). *bcr-abl*, *E2A-pbx*, and *pml-RARα* seem to be highly consistent partners.

Previously we reported the cloning of t(6;9) breakpoints (43). t(6;9) is the hallmark of a specific subtype of AML characterized by a poor prognosis and a young age of onset. It is classified in the French-American-British system mostly as M2/M4 and rarely as M1 or refractive anemia with excess of blast cells (RAEB) (2, 36, 39). On chromosome 9, breakpoints take place in a specific intron, *icb*-9, of a large gene (>140 kb) named *Cain* (*can*) (43). On chromosome 6, breakpoints also occur in a single intron, *icb*-6, of a gene named *dek* (42). The result of t(6;9) is the formation of a *dek-can* fusion gene on chromosome 6p−, which is transcribed into an invariable, 5.5-kb, leukemia-specific *dek-can* mRNA (39). The fusion transcript encodes a 165-kDa chimeric protein, which derives from the in-frame fusion of *dek* and *can* open reading frames (ORFs). Sequence comparison of DEK and CAN with entries in the EMBL data base shows no homology to any known protein sequences. CAN contains several putative dimerization motifs, and the C-terminal part may function as an ancillary DNA binding domain. The DEK protein contains 43% of charged amino acids and several acidic domains (42).

Surprisingly, a breakpoint in *icb*-9 of *can* was also detected in a bone marrow sample from a patient with acute undifferentiated leukemia (AUL) and an apparently normal karyotype. No breakpoint in *dek* could be detected in this case (42). In this paper, we report the isolation and characterization of a novel gene, named *set*, that was fused to *can* in the leukemic cells of this patient. A chimeric *set-can* transcript whose sequence predicts a SET-CAN protein of 155 kDa was detected.

MATERIALS AND METHODS

Southern and Northern (RNA) blotting. Patient material and human cell lines used were described previously (43). t(9;22) hybrid cell lines used were described elsewhere (8, 15). High-molecular-weight DNA was prepared as described by Jeffreys and Flavell (23). KG1 DNA was isolated and digested in agarose blocks as described previously (44).

* Corresponding author.

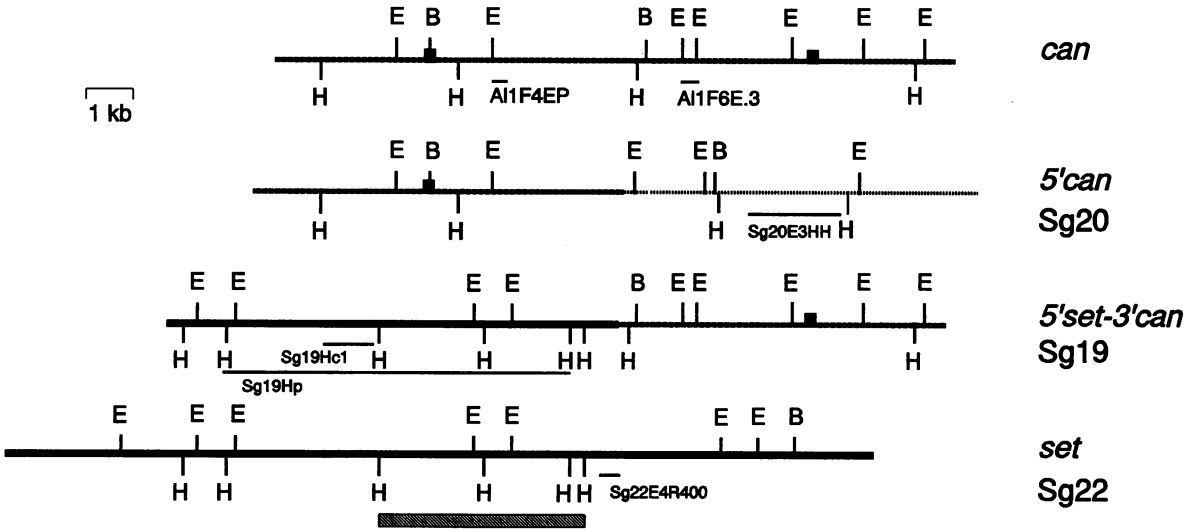

FIG. 1. Restriction maps of *can*, *set-can*, and *set*. Bold dotted lines indicate *can* sequences; solid lines indicate *set* sequences; the thin dotted line indicates sequences fused to 5′ *can*. The top line shows a region of *can* around *icb*-9; the exons flanking the intron are indicated by black boxes. The single-copy probes Al1F4EP and Al1F6E.3 are located at either side of the translocation breakpoint of patient SE. The second line represents phage Sg20, isolated with probe Al1F4EP. Sg20 contains 5′ *can* sequences and novel sequences 3′ of the translocation breakpoint. A 1.9-kb *Hin*cII-*Hin*dIII fragment, Sg20E3HH, could be used as a single-copy probe. The third line represents phage Sg19, isolated with probe Al1F6E.3. Sg19 contains 3′ *can* sequences fused to novel sequences 5′ of the translocation breakpoint that were shown to represent the *set* gene. A 1-kb *Hin*cII fragment (Sg19Hc1) was isolated as a single-copy probe. The genomic fragment Sg19Hp was used for in situ hybridization of chromosome preparations. The fourth line represents phage Sg22, containing wt *set* and isolated with probe Sg19Hc1. The striped bar indicates the region of *set* homologous to multiple copies in the genome. Sg22E4R400 is a 400-bp *Rsa*I fragment isolated as an additional single-copy probe just 5′ of the translocation breakpoint. The scale is indicated by a bar. Restriction sites: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII.

Restriction enzyme reactions were performed as recommended by the suppliers. Southern blots were done as described by Sambrook et al. (35). RNA of mouse tissue was isolated from BCBA mice. RNA was isolated by either the guanidinium isothiocyanate (7) or LiCl (4) method. Total RNA was electrophoresed and blotted as described by Fourney et al. (12). Equal amounts of rRNA were loaded. Before loading of the samples on a denaturing gel, 5% of each sample was loaded on a nondenaturing agarose gel to estimate the amount of rRNA and to adjust the sample quantity if necessary. Southern and Northern blots were hybridized in 10% dextran (35). Northern blots of mouse tissues were hybridized with human *set* probes with 3 × SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C, and filters were washed with 0.3 × SSC at 65°C. Probes were labelled by the method of Feinberg and Vogelstein (11).

Genomic and cDNA libraries. A λEMBL3 library of bone marrow DNA of patient SE was constructed as described by Frischauf et al. (13), and 2 × 10^6 PFU was screened with the *can* probes Al1F6E.3 and AL1F4EP (Fig. 1). A human testis cDNA library in λgt11 was purchased from Clonetech (Palo Alto, Calif.).

Cloning of chimeric *set-can* cDNA. Fifty micrograms of total RNA of bone marrow cells from AUL patient SE was heat denatured, and first-strand cDNA was synthesized with avian reverse transcriptase and 100 pmol of a 20-mer primer, 5′-CCACCAGGTGATTCAGCCTC-3′, located 200 bp downstream of the translocation breakpoint in the *can* cDNA (14, 20). cDNA was size selected on a Sephacryl S-1000 column (≥800 bp). Terminal deoxynucleotidyltransferase was used to tail first-strand cDNA with deoxyadenosine. Following heat denaturation, second-strand cDNA was synthesized with Klenow polymerase and the 35-mer 5′-GTCGCGAATTCGTCGACGCGTTTTTTTTTTTTTT-3′. Part (1/20) of the double-stranded cDNA was amplified with *Taq* polymerase (Perkin Elmer Cetus) and the primers 5′-GTCGCGAATTCGTCGACGCG-3′ and 5′-TTTGAATTCG TCGACCAGATGCTGATCCCACTCC-3′. The latter primer contains *Sal*I and *Eco*RI recognition sites fused to a 20-mer sequence located 86 bp downstream of the translocation breakpoint in the *can* cDNA sequence. No polymerase chain reaction (PCR) product of a specific size was generated, but a smear of PCR fragments hybridized to *can* cDNA probe hXT37BR derived from *can* cDNA, covering sequences 3′ of the translocation breakpoint. Although this probe overlaps with the amplification primers, conditions that prevented hybridization to the primers were chosen. Blots were washed at 65°C with 0.1× SSC. DNAs larger than 800 bp were isolated from an agarose gel, reamplified, cut with *Eco*RI, and cloned into λgt10. The resulting library was screened with *can* probe hXT37BR.

Cloning of the 3′ end of the 2-kb *set* mRNA. Thirty micrograms of total RNA of bone marrow cells from AUL patient SE was heat denatured, and first-strand cDNA was synthesized with avian reverse transcriptase, using 100 pmol of the 35-mer 5′-GTCGCGAATTCGTCGACGCGTTTTT TTTTTTTTTT-3′ as a primer (14, 20). Excess primer was removed by isopropanol precipitation. Part (1/100) of the cDNA reaction was amplified by using *Taq* polymerase (Perkin Elmer Cetus) and the primers 5′-GTCGCGAAT TCGTCGACGCG-3′ and either 5′-GTTTGGGTGGGTTAG TGGGC-3′ or 5′-CCACTCAATGGGAGAATCAGC-3′. The latter primers cover nucleotides (nt) 1382 to 1401 and 1519 to 1539, respectively, of the *set* cDNA. Amplified fragments of 350 and 220 bp were sequenced by using a protocol for direct

sequencing of fragments produced by an asymmetric PCR (24).

Sequence determination and analysis. Restriction fragments of cDNA clones were subcloned in M13. Overlapping cDNA sequences were determined on both strands by dideoxy sequencing (37). Initially M13 primers were used; when no suitable restriction sites were present, a primer was generated on the basis of the available cDNA sequence. To establish intron-exon borders, genomic fragments containing the exon of interest were subcloned into M13, and a cDNA primer near the putative intron-exon border was generated to prime the sequence reaction. Sequences were analyzed with the computer program Microgenie (Beckman), and the EMBL data base was used to search for homologous sequences at both the nucleotide and amino acid levels.

In situ hybridization. Hybridization was performed as described by Arnoldus et al. (3).

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases under accession number M93651.

RESULTS

Previously it was shown that *can* probe Al1F4EP, located at the 5′ end of *icb*-9 (Fig. 1), hybridized to an aberrant fragment in bone marrow DNA of AUL patient SE (43). On a Northern blot, 3′ *can* cDNA probes hybridized to an aberrant 5-kb transcript in total RNA of bone marrow cells of patient SE (42). These data proved that a breakpoint is present in *icb*-9 of the *can* gene. However, the 5-kb aberrant transcript failed to hybridize to 5′ *dek* cDNA probes that invariably detected the leukemia-specific *dek-can* transcript of similar size in t(6;9) AML cells. This finding suggested to us that a gene other than *dek* may be fused to *can* in leukemic cells of patient SE.

Isolation of the *set* gene. To isolate and characterize DNA sequences fused to *can* in AUL patient SE, we constructed a genomic λEMBL3 phage library of high-molecular-weight DNA of leukemic cells of this patient. In total, 2×10^6 PFU was screened with *can* probes Al1F4EP and Al1F6E.3, located at either side of the translocation breakpoint (Fig. 1). Hybridizing clones (12) were analyzed by restriction enzyme mapping. Most clones contained DNA from the normal *can* allele, but clones Sg19 and Sg20 (hybridizing to Al1F6E.3 and Al1F4EP, respectively) contained *can* sequences fused to DNA with an unknown restriction enzyme pattern. Subclones of phages Sg19 and Sg20 were tested for DNA fragments not containing any repetitive sequences (not shown). A 1-kb *Hin*cII fragment from Sg19 (Sg19Hc1) and a 1.9-kb *Hin*dIII-*Hin*cII fragment from Sg20 (Sg20E3HH) (Fig. 1) could be used as single-copy probes to screen the genomic library again in order to obtain nonrearranged clones overlapping with Sg19 and Sg20. Ten phages, two hybridizing to Sg19Hc1 and eight hybridizing to Sg20E3HH, were isolated and analyzed. Phage Sg22 hybridized to probe Sg19Hc1 located at the 5′ side of the breakpoint and extended beyond the breakpoint at the 3′ end (Fig. 1). However, hybridization experiments failed to detect overlap between Sg22 and Sg20, which contains DNA juxtaposed to *can* at the 3′ end of the breakpoint. Similarly, no overlap could be detected between Sg19 and phages hybridizing to probe Sg20E3HH that extend at the 5′ end beyond the breakpoint present in Sg20 (results not shown). Therefore, it was concluded that a stretch of DNA flanking the translocation breakpoint was deleted. The size of this deletion measures at least 6 kb and may be much larger. As 3′, but not 5′, *can* cDNA probes detected an aberrant transcript in RNA of leukemic cells of patient SE, attention was focused on the newly isolated DNA fragments 5′ of the translocation breakpoint fused to 3′ *can*. We reasoned that these sequences could be part of a novel fusion gene associated with this case of AUL.

To prove that Sg19 indeed represents the translocation breakpoint, probe Sg19Hc1 was hybridized to a Southern blot containing DNA isolated from bone marrow cells of patient SE. In addition to a large ~18-kb *Bam*HI fragment, a fragment of ~15 kb specific for the AUL sample was detected (not shown). Probe Sg22E4R400 (Fig. 1) also detected an aberrant *Bgl*II fragment in DNA of the leukemic cells from patient SE (not shown).

Isolation of *set* cDNA. To isolate cDNA sequences representing the aberrant transcript detected in RNA of patient SE, we used an anchored PCR according to the protocol for rapid amplification of cDNA ends (RACE) (14). The reaction was primed with a *can* oligonucleotide located 86 nt downstream of the translocation breakpoint in *dek-can* cDNA. Hybridization of the PCR products to a *can* cDNA probe just 3′ of the breakpoint, hXT37BR (Fig. 2), showed a smear of fragments. Therefore, size-selected DNA (≥800 bp) was isolated from a preparative agarose gel and cloned into λgt10 to generate a small selective cDNA library. Many phages hybridized to *can* cDNA probe hXT37BR, and six of these were analyzed. One clone only weakly hybridized, three clones contained wild-type (wt) *can* sequences, and two clones (SE3 [300 bp] and SE4 [700 bp]) contained 86 bp of *can* sequences linked to unknown DNA. The 5′ 500 bp of clone SE4 (an *Eco*RI-*Rsa*I fragment) were isolated and hybridized to a Southern blot containing restriction enzyme digests of the genomic phages Sg19, Sg20, and Sg22. Strong hybridization to a 5.5-kb *Eco*RI fragment in phages Sg19 and Sg22 was detected (schematically indicated in Fig. 2). On a Northern blot containing total RNA of several cell lines (K562, HeLa, Daudi, HL60, and KG1) and AUL patient SE, the SE4ER probe hybridized to two transcripts of 2.7 and 2.0 kb present in all cell lines and to additional transcripts of 5 and 6.5 kb, specific for the AUL sample (Fig. 3A). The 5-kb transcript is identical in size to the aberrant transcript detected by 3′ *can* probes in RNA of leukemic cells of this patient.

Together, the data prove that in AUL patient SE, the 3′ part of *can* is fused to a novel gene that is distinct from the previously isolated *dek* gene. This novel gene was named *set*.

To isolate full-length *set* cDNA, probe SE4ER was used to screen a λgt11 testis library. Six hybridizing clones were isolated and analyzed. The overlapping cDNAs SE9 and SE10 are shown in Fig. 2. SE10 extends most 5′, and SE9 extends most 3′. Different subfragments of clones SE9 and SE10 were hybridized to a Southern blot containing restriction enzyme digests of the genomic phages Sg19, Sg20, and Sg22. The 5′ part of SE10 hybridized to the same 5.5-kb *Eco*RI fragment that hybridized to the fusion cDNA clone SE4. Surprisingly, the 3′ part of SE10 and almost the entire clone SE9 hybridized to a 1.1-kb *Eco*RI-*Hin*dIII fragment in Sg19 that is situated 3′ of the 5.5-kb *Eco*RI fragment but 5′ of the translocation breakpoint (Fig. 2). Fine mapping of cDNA clone SE9 and the genomic fragment to which it hybridized showed that restriction maps of the two fragments are colinear. These data suggest that the *set* gene contains a large exon at its 3′ end, which is situated 5′ of the *set-can* translocation breakpoint. Since these exon sequences are

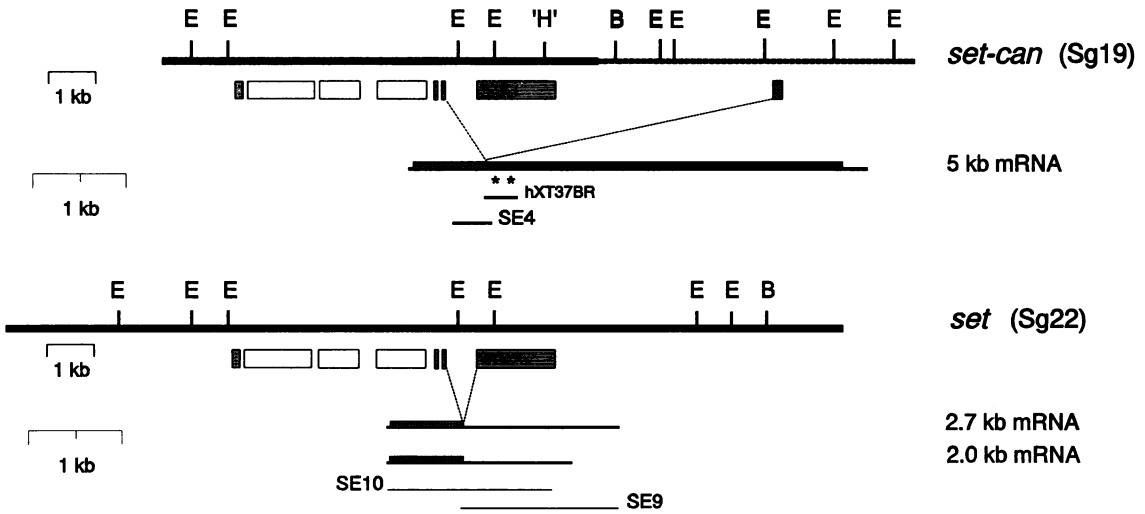

FIG. 2. The *set* gene. The top line represents part of the *set-can* gene. The second restriction map represents the wt *set* gene. The striped line indicates *can* sequences; solid lines indicate *set* sequences; open boxes below the maps indicate restriction enzyme fragments hybridizing to cDNA sequences; stippled boxes are defined *set* exons; the striped box is a defined *can* exon. The 5-kb fusion transcript is indicated below the *set-can* map. The *can* ORF is indicated by a solid bar, and the *set* ORF is indicated by a hatched bar. SE4 is the fusion cDNA clone, isolated via the RACE protocol and hybridizing to *can* cDNA probe hXT37BR. Asterisks indicate positions of the oligonucleotides used to amplify the fusion cDNAs. A stippled line indicates which *set* exon is spliced to the *can* exon 3′ of *icb*-9. The last 3′ *set* exon is alternatively spliced and not present in the 5-kb fusion transcript. The 2.0- and 2.7-kb transcripts are indicated below the *set* map; they differ only in the polyadenylation signal used. The stippled line indicates the splicing of the penultimate exon to the last exon of *set*. SE9 and SE10 are cDNA clones isolated from a λgt11 human testis library. All *Eco*RI (E) and *Bam*HI (B) sites are indicated; 'H' represents only one of several *Hin*dIII sites and is indicated for clarity.

not present in the fusion cDNAs SE3 and SE4, they must be excluded from the 5 kb *set-can* transcript by alternative splicing.

Markedly, a weak 6.5-kb transcript was detected next to the aberrant 5-kb transcript in total RNA of leukemic cells of patient SE after hybridization with the *set* cDNA probe SE4 (Fig. 3A). A transcript of the same size was also detected by the 3′ *can* cDNA probe hXT54, because the transcript migrates just a little faster than the normal 6.6-kb *can* transcript (Fig. 3B). This finding suggests that a 6.5-kb *set-can* transcript in addition to the 5-kb transcript is present in the leukemic cells. Possibly, this transcript contains part

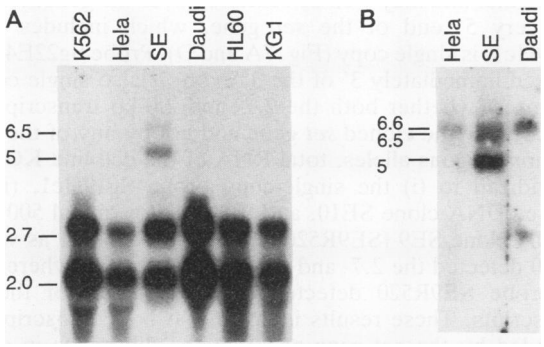

FIG. 3. Northern blots. Total RNA (20 μg) isolated from cell lines K562, HeLa, Daudi, HL60, and KG1 and from bone marrow of AUL patient SE was hybridized to the *set* cDNA probe SE10 (A). Total RNA (20 μg) isolated from cell lines HeLa and Daudi and from bone marrow of AUL patient SE was hybridized to *can* cDNA probes 3′ of the translocation breakpoint (B). Sizes of transcripts are indicated in kilobases.

of the 3′ *set* exon, spliced to *can* via a cryptic splice donor site in the 3′ *set* exon.

Sequence analysis of *set* cDNA. The nucleotide sequences of SE9 and SE10 were determined. Together these clones contained 2,570 bp of cDNA sequences. An ORF of 831 nt, encoding a protein of 277 amino acids (aa) with a predicted molecular mass of 32 kDa, is present. A long 3′ untranslated region (UTR) of 1,733 nt follows the ORF (Fig. 4). This 3′ UTR contains many stop codons in all reading frames. The ORF starts immediately at the 5′ end of SE10 at position 5 of the cDNA, and the first stop codon in this ORF is found at position 837. Comparison of the sequences of the fusion cDNA SE4 and cDNA clone SE10 shows that nt 814 (aa 269) is fused in frame to the *can* ORF. Sequence determination of part of the homologous genomic clone (Sg19E5.5) showed that this position 814 in the *set* cDNA corresponds to an exon-intron border (Fig. 5D). The 3′ border of this intron was determined by sequencing an 800-bp *Eco*RI fragment of genomic phage Sg19 (Fig. 5D). A large part of the exon 3′ of this intron was sequenced from the Sg19 subclones Sg19E.8 and Sg19E3. This sequencing confirmed the suggestion that in the normal *set* gene, the exon spliced to *can* in the fusion gene is followed by a single large 3′ exon of 1,756 nt. This exon still contains 7 aa of the ORF and the entire 3′ UTR. The polyadenylation signal for the 2.7-kb transcript is found at positions 2542 to 2547, followed by the poly(A) tail at position 2562. The genomic sequence overlapping the 3′ end of the cDNA is shown in Fig. 5B.

As the ATG start codon is present at the very 5′ end of SE10, an anchored PCR was used to clone the 5′ end of the *set* transcript. Unfortunately, because of the high G+C content of the DNA, we were not successful. The 5′ cDNA sequences are contained in the 5′ end of the 5.5-kb *Eco*RI fragment of the genomic phage Sg19. The genomic nucleo-

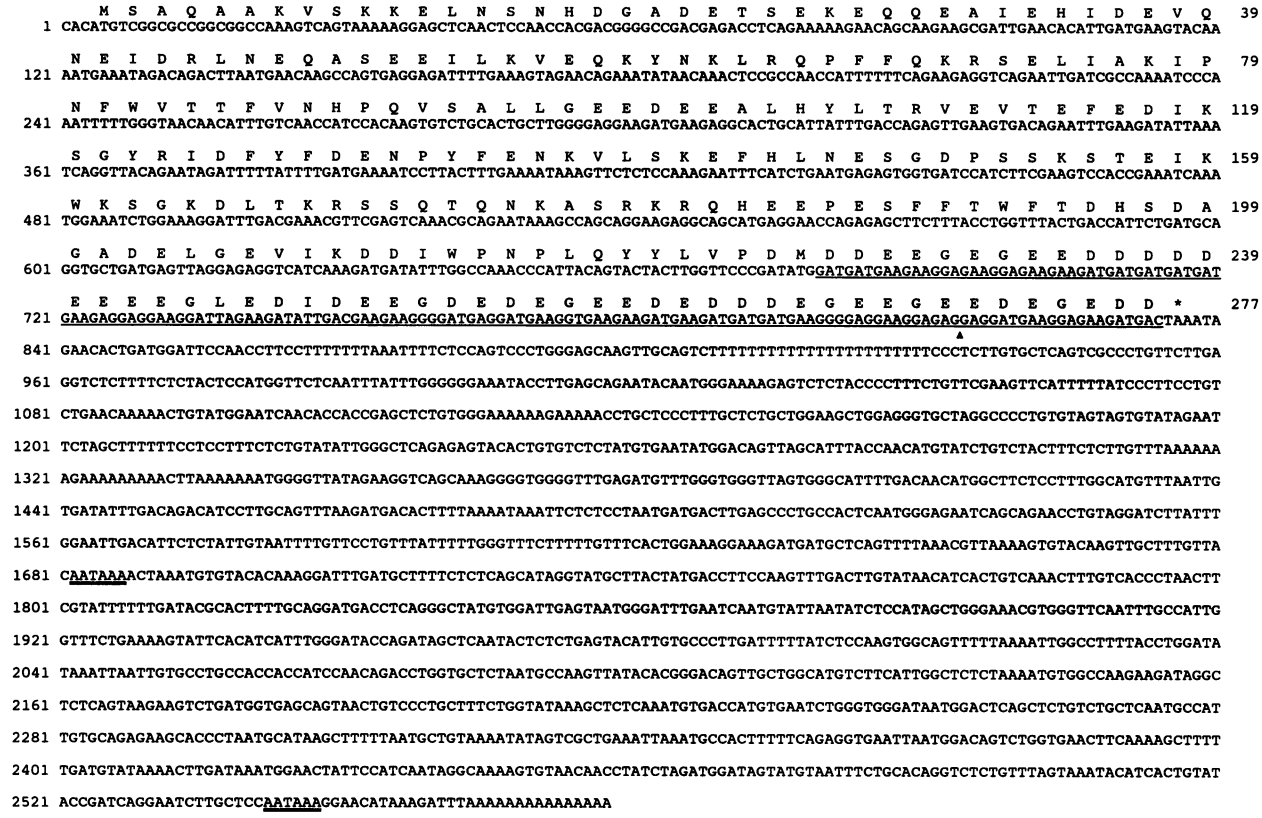

FIG. 4. cDNA and putative amino acid sequences of the SET protein. The position where *set* is fused to *can* in the *set-can* fusion protein is indicated with a triangle (nt 814). The acidic tail is underlined. The poly(A) addition sites are double underlined.

tide sequence preceding the ATG codon, which appeared to be rich in CpG, was determined (partly shown in Fig. 5A). CpG-rich areas are found to be associated with certain promoter regions (5). Therefore, it seems likely that this ATG encodes the first methionine of the putative SET protein and that the genomic region 5′ of the cloned cDNA contains the promoter sequences.

The putative SET protein contains an extremely high percentage of acidic residues, 32% (98 aa), half of which (43 aa) are present at the C terminus, forming a long acidic tail; 37 of the acidic residues are present in the chimeric SET-CAN protein. When the amino acid sequences of SET and DEK are compared, only the acidic stretches show homology. Comparison of the SET sequence with sequences in the EMBL data base revealed homology with a putative *Plasmodium falciparum* protein (31) and with a nucleosome assembly protein, NAP-I, of *Saccharomyces cerevisiae* (22) (Fig. 6). All three proteins have a large stretch of acidic residues at the C terminus. In addition, comparison of aa 38 to 221 of SET with aa 20 to 201 of the *P. falciparum* protein showed 33% (61 of 187 aa) identity and 50% (94 of 187 aa) similarity. Comparison of aa 1 to 220 of SET with aa 116 to 362 of *S. cerevisiae* NAP-I showed 24% identity and 36% similarity. However, some regions are more homologous than others; i.e., SET aa 69 to 143 showed 34% identity and 54% similarity to the corresponding amino acids in NAP-I.

The 3′ part of the *set* gene is not single copy. Upon screening genomic phage Sg19 for fragments free of repetitive sequences, it was found that a 1-kb *Eco*RI-*Hin*dIII fragment just upstream of the *can* sequences (Fig. 2) detected a number of bands after hybridization to a Southern blot containing human DNA (Fig. 7A). This finding suggested that at least 10 cross-hybridizing copies of this sequence are present in the genome. This *Eco*RI-*Hin*dIII fragment appeared to be part of the 3′ exon of *set*. Also, cDNA probe SE4ER appeared to be multicopy. This probe is located upstream of the G+A-rich stretch of DNA encoding the acidic tail of the SET protein. On long-range mapping Southern blots with *Bss*hII-digested DNA, a comparable number of fragments, ranging from 30 to 800 kb in size, could be detected (Fig. 7B). This finding indicates that the sequences hybridizing to *set* are located at large distances from each other and may well be scattered over the genome. Only the very 5′ end of the *set* gene, which includes probe Sg19Hc1, is single copy (Fig. 7A and D). Probe Sg22E4R420, located immediately 3′ of the 3′ exon, is also single copy.

To test whether both the 2.7- and 2.0-kb transcripts are encoded by the cloned *set* gene and not by any of the other *set*-homologous alleles, total RNA of the cell line KG1 was hybridized to (i) the single-copy probe Sg19Hc1, (ii) the entire cDNA clone SE10, and (iii) the 3′-terminal 500 bp of cDNA clone SE9 (SE9R520) (Fig. 7C). Sg19Hc1 as well as SE10 detected the 2.7- and the 2.0-kb mRNAs, whereas the 3′ probe SE9R520 detected only the larger of the two transcripts. These results indicate that both transcripts are encoded by the *set* gene and that the difference in size is most likely due to alternative polyadenylation. Two putative poly(A) signals for the 2-kb mRNA are present in the 3′ UTR, at positions 1485 to 1490 and 1682 to 1687, respectively. To test whether these signals are used, the RACE protocol (14) was employed. cDNA generated with an adaptor-oligo(dT) primer was further amplified with an adaptor

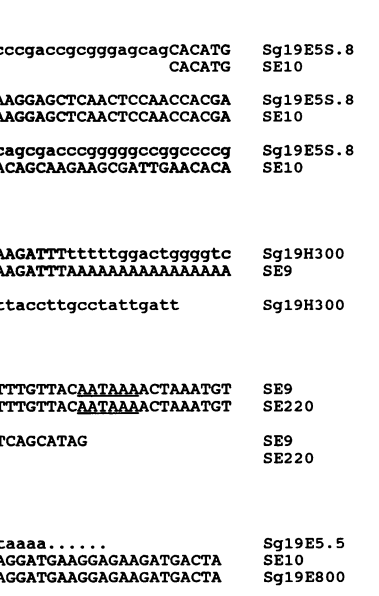

FIG. 5. Comparison of genomic and cDNA sequences of *set*. (A) A 5.5-kb *Eco*RI fragment in Sg19 contains the most 5′ sequences of *set* cDNA clone SE10 in an 800-bp *Sst*I fragment (Sg19E5S.8). (B) The end of the 3′ exon of *set* is present in cDNA clone SE9 and in a genomic 300-bp *Hin*dIII fragment derived from phage Sg19. The poly(A) signal is underlined. (C) A 2-kb mRNA is generated by alternative polyadenylation on nt 1703 of the *set* cDNA sequence. SE9 represents part of the 2.7-kb mRNA; SE220 is the amplified 3′ end of the 2-kb *set* mRNA. The poly(A) signal is underlined. (D) The intron-exon borders of the last *set* intron were sequenced to determine the position of the *set-can* fusion in the *set* cDNA sequence. The exon that is spliced to *can* is present in a genomic 5.5-kb *Eco*RI fragment derived from phage Sg19 (Sg19E5.5). The border of the 3′ exon is present in an 800-bp *Eco*RI fragment of the same phage clone (Sg19E800). Genomic sequences are aligned with the sequence of cDNA clone SE10.

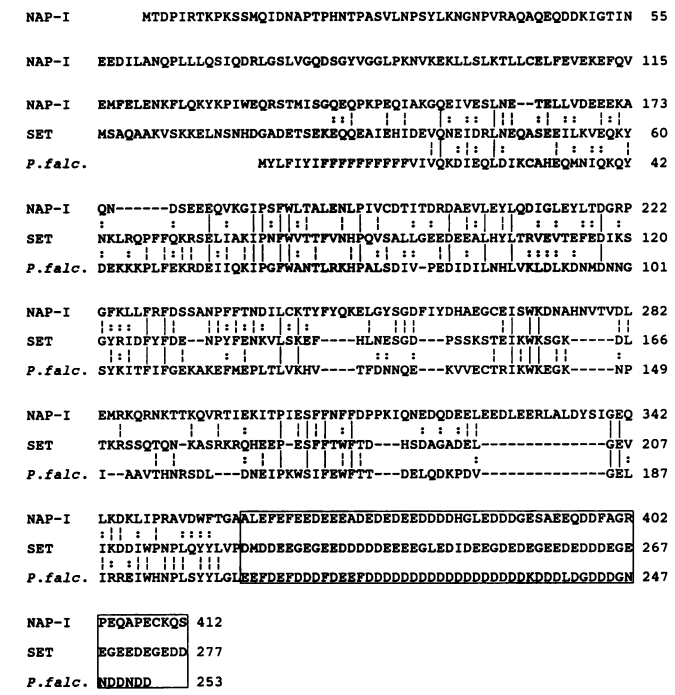

FIG. 6. Alignment of the SET protein with a putative protein of *P. falciparum* (*P.falc.*) and the NAP-I protein of *S. cerevisiae*. Identical amino acids in all three proteins are indicated with a vertical line; identical amino acids in SET and NAP-I or in SET and the *P. falciparum* protein are indicated with a broken vertical line; similar amino acids are indicated with colons; acidic C-terminal regions are boxed.

oligonucleotide and a primer located at either positions 1382 to 1401 or positions 1519 to 1539. This procedure resulted in a fragment of 350 or 220 nt, respectively. Direct sequencing of these fragments showed that the putative poly(A) signal at positions 1485 to 1490 is not used, while the poly(A) signal at positions 1682 to 1687 results in polyadenylation of the mRNA at position 1703 (Fig. 4 and 5C). Hybridization of probe SE10 to a blot containing DNA from different vertebrates (zoo blot) showed a considerable degree of conservation of the *set* gene among humans, mice, marsupials, chickens, amphibians, and fish; hybrids are stable at 65°C in 1× SSC (data not shown). Markedly, the repetitive nature of the 3′ part of *set* was seen in both human and mouse genes.

Expression of *set*. The expression pattern of *set* in different mouse tissues was analyzed by Northern blotting. Total RNA (20 μg) from bone marrow, spleen, thymus, brain, liver, kidney, testes, ovaries, placenta, and whole embryos sacrificed 10, 13, 16, and 19 days postcoitum was loaded on a denaturing agarose gel and hybridized to the human SE10 cDNA probe. As shown in Fig. 8, *set* is expressed in all adult tissues analyzed. The expression during embryogenesis is remarkable. *set* expression is relatively high in the youngest embryos (10 days old) and decreases during development. In mouse cells, not two but three or four transcripts were detected. The nature of the smaller mRNAs is not known. The band migrating just under the 18S rRNA band may be background due to compression of the background smear by the bulk of rRNA.

Chromosomal localization of the *set* gene. The karyotype of leukemic cells of patient SE appeared to be normal and gave no clue as to where in the genome *set* is located. A biotinylated genomic fragment from phage Sg19, encompassing three *Hin*dIII fragments of 3.3, 2.3, and 1.9 kb (Sg19Hp; Fig. 1), was used to determine the chromosomal localization by in situ hybridization. Although some background fluorescence was present, a clear signal was detected at the tip of the long arm of chromosome 9 (Fig. 9A and B). The oncogenes c-*abl*, involved in t(9;22), and *can*, involved in t(6;9), are also located on chromosome 9q34. Somatic cell hybrids containing the 9q+ or 22q− chromosome of t(9;22) were hybridized to probe Sg19Hc1 to determine the localization of *set* relative to c-*abl* and *can*. As shown in Fig. 9C, *set* sequences hybridized to cell lines carrying the normal chromosome 9 (17CB-10) or the 9q+ chromosome (8CB-7B and 15CB-7D) but not to cell lines carrying the 22q- chromosome (Wedy9 and WESP-2A). This finding confirmed the results of the in situ hybridization and locates the *set* gene on chromosome 9, centromeric of c-*abl*. The physical distance between *set* and *abl* is unknown.

DISCUSSION

In t(6;9) AML, a specific fusion of *dek* to *can* is found. All t(6;9) leukemic cells analyzed to date invariably contained a translocation breakpoint in *icb*-6 of *dek* and in *icb*-9 of *can* (39). Here we provide evidence that *can* is involved in at least two different translocation events. In leukemic cells of

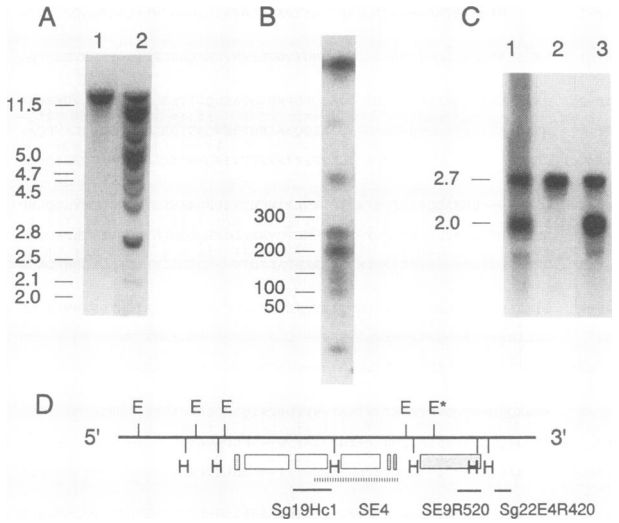

FIG. 7. (A and B) Presence of multiple copies of *set* in the genome. Southern blots containing DNA derived from human thymus, digested with *Bgl*II (A), or DNA derived of the cell line KG1, digested with *Bss*hII (B), were hybridized to probes Sg19Hc1 (A, lane 1) and SE9R520 (A, lane 2; B). Positions of size markers (in kilobases) are shown next to the blots. (C) Derivation of both *set* transcripts from the cloned *set* gene. Northern blots with RNA derived from cell line KG1 were hybridized to probes Sg19Hc1 (lanes 1), SE9R520 (lanes 2), and SE10 (lanes 3). Sizes of the transcripts are indicated in kilobases. (D) Genomic map indicating positions of the probes. Open boxes underneath the *set* restriction map indicate restriction fragments hybridizing to *set* cDNA probes. Probes Sg19Hc1, SE9R520, and Sg22E4R420 are indicated by bars. cDNA probe SE10 contained cDNA sequences encoded upstream of the *Eco*RI site (E) marked with an asterisk. The estimated position of SE4 is indicated with an interrupted line. H, *Hin*dIII.

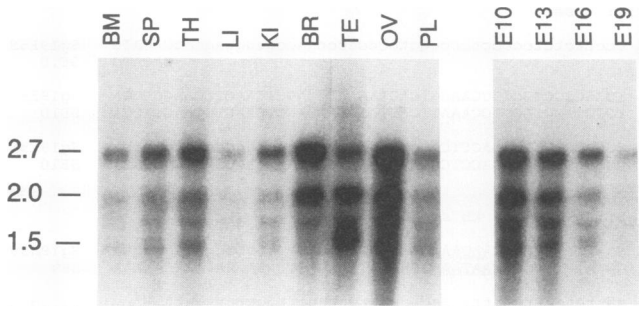

FIG. 8. Detection of *set* mRNAs in mouse tissues. A Northern blot containing total RNA derived from various tissues of BCBA mice was hybridized to the human *set* cDNA probe SE10. Lanes: BM, bone marrow; SP, spleen; TH, thymus; LI, liver; KI, kidney; BR, brain; TE, testes; OV, ovaries; PL, placenta; E10, E13, E16, and E19, embryos at 10, 13, 16, and 19 days postcoitum. Lanes E10 to E19 were exposed for a shorter length of time than were the other lanes. In the original exposure, the signal of lane E19 is comparable to the signal of lane SP.

a patient with AUL, *can* is fused to a novel gene named *set*. The *set-can* fusion gene encodes a 5-kb chimeric transcript whose nucleotide sequence predicts a 155-kDa SET-CAN fusion protein. The finding that the same part of CAN is linked by translocation to two different protein moieties at its N-terminal side suggests that the C-terminal part of CAN contains domains involved in leukemogenesis that can be activated in more than one way. Markedly, the phenotype of the leukemic cells carrying the *set-can* fusion gene was very immature in this patient (1), whereas a variable degree of differentiation into the myeloid lineage is observed in t(6;9) AML cells.

A large-scale study for involvement of *dek* and *can* in leukemia confirmed the specificity of the association of t(6;9) with the *dek-can* fusion gene. However, among the 320 cases of myelodysplastic syndrome (MDS), AUL, AML, and acute lymphoblastic leukemia studied, two leukemia samples contained a breakpoint in *icb*-9 of *can*, whereas no breakpoint in *dek* could be detected. One sample was a RAEB, and the other was a common acute lymphocytic leukemia (40). These samples were tested for a breakpoint in *set*, and the results so far have been negative (41). These results suggest that in addition to *dek* and *set*, other genes might be able to activate *can*. Although the different types of CAN fusion proteins are all associated with leukemia, the differentiation potential of the leukemic cells may be influenced by the N-terminal moiety of the fusion protein.

The predicted protein sequence of SET shows no homology with DEK apart from the fact that both proteins contain acidic regions. Many proteins containing acidic regions are located in the nucleus and may have different functions (reviewed in reference 10). Although the function of CAN and DEK is not known, domains identified in these proteins suggest a role in transcription regulation (42). Analogous to acidic domains in VP16 and GAL4 (6, 21), the acidic motifs of DEK and SET could serve as transcription activation domains. On the other hand, acidic domains are also present in proteins associated with chromatin, such as nucleolins (30) and high-mobility-group proteins (28). Functional assays are needed to determine whether the acidic domains of SET and DEK are essential for the putative transforming capacity of the DEK-CAN and SET-CAN fusion proteins.

The homology of SET with the *S. cerevisiae* nucleosome assembly protein NAP-I (22) argues that SET may be a nuclear protein. This notion is interesting since fusion of CAN to DEK results in a nuclear localization of the fusion protein, whereas CAN itself is mainly cytoplasmic. Fusion of SET to CAN could have the same effect and result in a nuclear localization of the SET-CAN fusion protein. A nuclear localization of the C-terminal part of CAN may be essential for the putative leukemogenic effect of the fusion proteins.

In addition to the acidic stretch at the N terminus, the entire SET protein and the NAP-I protein are 24% identical. SET also shows homology to a putative protein of *P. falciparum* whose function is unknown (31). Possibly, the functions of SET and the putative *P. falciparum* protein in the cell are related to the function of the nucleosome assembly protein. SET and the *P. falciparum* protein are similar in size, and their overall identity is 33%. The NAP-I protein is larger and extends at its N-terminal side.

Like *dek*, the *set* gene is expressed in all tissues of the mouse, suggesting that SET has a rather general function in the organism. Still, the expression pattern is not entirely identical to that of *dek*. Most remarkable is the high expression level in early embryos. It will be interesting to analyze by in situ hybridization whether the high expression of *set* in early embryos is found throughout the embryo or whether it is restricted to specific structures.

set is a relatively small gene of 8 kb. The most 5′ sequences of the cloned *set* cDNA are located in a CpG island that measures at least 1,000 bp in the genomic DNA.

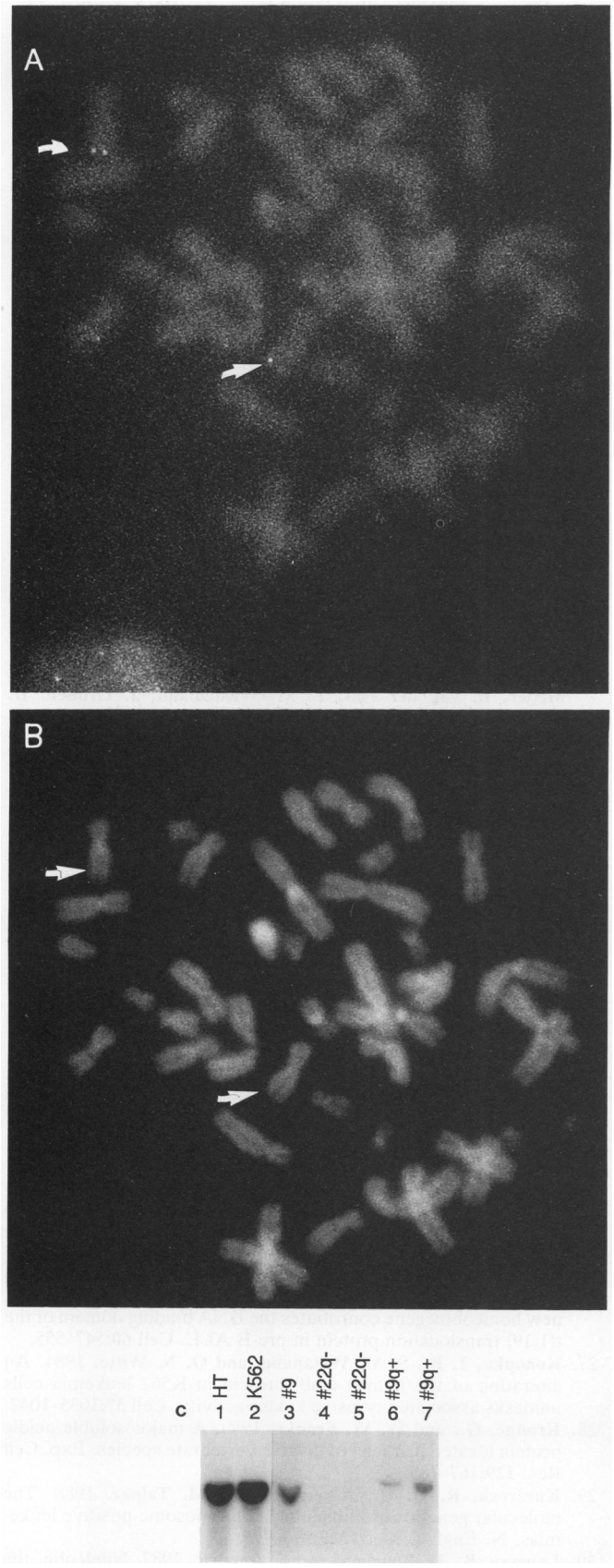

FIG. 9. Chromosomal localization of the *set* gene. (A) Fluorescence in situ hybridization of a chromosome preparation with the genomic *set* fragment Sg19Hp (Fig. 1); (B) 4',6-diamidino-2-phenylindole–actinomycin counterstaining; (C) Southern blot containing *Eco*RI-digested DNA derived from somatic cell hybrids with the segregated chromosomes of the t(9;22), hybridized to the genomic *set* probe Sg19Hc1. Total human DNA was loaded in lanes 1 (K562) and 2 (human thymus). The somatic cell hybrid 17CB-10 (lane 3) contains the entire chromosome 9, Wedy9 and WESP-2A (lanes 4 and 5) contain the 22q− chromosome, and 8CB-7B and 15CB-7D (lanes 7 and 8) contain the 9q+ chromosome.

The sizes of the mRNAs on Northern blots are estimated to be 2.7 and 2 kb. The cloned cDNA sequences represent 2.5 and 1.7 kb. It is likely that sequences at the 5′ side of the transcript are missing from the cDNA clones. Since CpG islands are strongly associated with promoter regions (5), we assume that the 5′ part of phage Sg19 contains the most 5′ *set* sequences and the *set* promoter.

In comparison with other fusion genes, the position of the breakpoint in the *set-can* fusion gene is peculiar since it is located 800 bp 3′ of the *set* gene and not in an intron. Apparently, the primary transcript is not terminated in the 4.8 kb that separate the 3′ end of *set* from the most 5′ exon of the translocated *can* gene but rather proceeds to the end of the *can* gene. Subsequently, the 3′ *set* exon must be skipped by splicing to produce an in-frame *set-can* fusion transcript. Large exons seldom occur within a gene, and it has been reported that large exons are preferentially spliced out of primary transcripts (17, 34). Markedly, next to an aberrant 5-kb transcript, a weaker 6.5-kb transcript which hybridizes to cDNA clone SE10 can be detected in total RNA of leukemic cells of patient SE. An additional transcript of similar size also hybridizes to 3′ *can* cDNA probes and is just smaller than the 6.6-kb normal *can* transcript (Fig. 3). This finding suggests that the 6.5-kb mRNA may be a *set-can* fusion transcript containing part of the 3′ *set* exon, spliced in via a cryptic splice donor site. Considering the intensities of the signals on the Northern blot, the use of the cryptic splice site is not very efficient. The presence of transcripts containing part of the 3′ *set* exon will be of no functional importance since the natural *set* stop codon is present in this mRNA. The longer fusion transcripts would encode only wt SET protein.

Hybridization of total RNA of leukemic cells of patient SE with *set* cDNA probes shows that the steady-state levels of *set* are much higher than those of *set-can*, despite the presence of >90% leukemic cells in the bone marrow sample from which RNA was isolated. As each cell contains one *set* promoter driving wt *set* expression and one *set* promoter driving transcription of the *set-can* fusion gene, additional elements regulate the relative abundance of *set* and *set-can* transcripts. (i) Transcription of the *set* gene may be upregulated by an enhancer element located 3′ of the gene, which is removed by the translocation. (ii) The presence of the last 3′ exon on the rearranged chromosome may give rise to production of wt *set* transcripts from the *set-can* fusion gene. Transcription termination and polyadenylation of normal *set* transcripts will decrease the level of *set-can* transcripts in the cell. (iii) The difference in *set* and *set-can* steady-state levels may also be due to a reduced half-life of *set-can* mRNA compared with that of *set* mRNA.

The karyotype of the leukemic cells of patient SE appeared to be normal. We show here that *set* is localized on chromosome 9q34. Thus, the chromosomal aberration is not

detected by cytogenetic means since the distance between *set* and *can* is relatively small. The chromosomal rearrangement that took place to fuse *set* to *can* may be either an insertion, an inversion, or a translocation. A deletion can be excluded because genomic phages that contained novel sequences fused to the 5′ part of *can* as well as to its 3′ part were cloned. Fluorescent in situ hybridization techniques (3) may distinguish between the other possibilities.

Only the 5′ end of *set* appeared to be single copy. The 3′ part of the *set* gene is present at least 10 times in the human genome. The *set* gene may be part of a gene family whose members have highly homologous 3′ ends and specific 5′ leaders. Alternatively, the other copies may represent nonexpressed pseudogenes lacking the 5′ end of *set*. Similarly, four duplications of the 3′ *bcr* gene are present on chromosome 22 at large distances from each other (18). These *bcr* copies are not expressed and contain conserved exons as well as conserved introns. The *bcr*-related genes are present in gorilla and chimpanzee DNAs but not in mouse DNA (18).

Northern blot and PCR analyses suggest that both the 2.0- and 2.7-kb *set* transcripts are encoded by the *set* gene identified in this report. However, it can not be excluded that *set*-homologous genes are expressed in specific cell types. The hybridization analysis indicated that the *set* gene is evolutionarily well conserved and that a multiplication of *set* occurred after the divergence of marsupials and mammals but before the divergence of primates and rodents. It is surprising that multiple copies remained so well conserved not only in the protein-coding region but also in the large 3′ UTR. It is not known whether introns are conserved as well.

ACKNOWLEDGMENTS

We thank D. Bootsma, D. Meijer, and M. Fornerod for continuous support and discussion, J. Abels and H. Adriaansen for patient material, and M. Kuit and T. de Vries Lentsch for photographic work.

This work was supported by the Dutch Cancer Foundation.

REFERENCES

1. **Adriaansen, H. J., P. W. C. Soeting, I. L. M. Wolvers-Tettero, and J. J. M. Van Dongen.** 1991. Immunoglobulin and T cell receptor gene rearrangements in acute non-lymphocytic leukemia. Leukemia **5**:744–751.
2. **Adriaansen, H. J., J. J. M. van Dongen, H. Hooijkaas, K. Hahlen, M. B. van 't Veer, B. Lowenberg, and A. Hagemeijer.** 1988. Translocation (6;9) may be associated with a specific TdT-positive immunological phenotype in ANLL. Leukemia **2**:136–140.
3. **Arnoldus, E. P. J., J. Wiegant, I. A. Noordermeer, J. W. Wessels, G. C. Beverstock, G. C. Grosveld, M. van der Ploeg, and A. K. Raap.** 1990. Detection of the Philadelphia chromosome in interphase nuclei. Cytogenet. Cell Genet. **54**:108–111.
4. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
5. **Bird, A. P.** 1986. CpG islands and the function of DNA methylation. Nature (London) **321**:209–213.
6. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43**:729–736.
7. **Chirgwin, J. M., A. E. Przybyla, J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
8. **de Klein, A., A. Geurts van Kessel, G. Grosveld, C. R. Bartram, A. Hagemeijer, D. Bootsma, N. K. Spurr, N. Heisterkamp, J. Groffen, and J. R. Stephenson.** 1982. A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukemia. Nature (London) **300**:765–767.
9. **De Thé, H., C. Lavau, A. Marchio, C. Chomienne, L. Degos, and A. Dejean.** 1991. The PML-RARA fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. Cell **66**:675–684.
10. **Earnshaw, W. C.** 1987. Anionic regions in nuclear proteins. J. Cell Biol. **105**:1479–1482.
11. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
12. **Fourney, R. M., J. Miyakoshi, R. S. Day III, and M. C. Paterson.** 1988. Northern blotting: efficient RNA staining and transfer. Focus **10**:5–6.
13. **Frischauf, A., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
14. **Frohman, M. A.** 1990. Rapid amplification of cDNA ends (RACE): user-friendly cDNA cloning. Amplifications **5**:11–15.
15. **Geurts van Kessel, A. H. M., P. A. T. Tetteroo, A. E. G. K. von dem Borne, A. Hagemeijer, and D. Bootsma.** 1983. Expression of human myeloid-associated surface antigens in human-mouse myeloid cell hybrids. Proc. Natl. Acad. Sci. USA **80**:3748–3752.
16. **Grosveld, G., T. Verwoerd, T. van Agthoven, A. de Klein, K. L. Ramachandran, N. Heisterkamp, K. Stam, and J. Groffen.** 1986. The chronic myelocytic cell line K562 contains a breakpoint in *bcr* and produces a chimeric *bcr*/c-*abl* transcript. Mol. Cell. Biol. **6**:607–16.
17. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16**:9893–9908.
18. **Heisterkamp, N., and J. Groffen.** 1988. Duplication of the *bcr* and gamma-glutamyl transpeptidase genes. Nucleic Acids Res. **16**:8045–8056.
19. **Hermans, A., N. Heisterkamp, M. von Linden, S. van Baal, D. Meijer, D. van der Plas, L. M. Wiedemann, J. Groffen, D. Bootsma, and G. Grosveld.** 1987. Unique fusion of bcr and c-abl genes in Philadelphia chromosome positive acute lymphoblastic leukemia. Cell **51**:33–40.
20. **Hermans, A., L. Selleri, J. Gow, L. Wiedeman, and G. Grosveld.** 1989. Molecular analysis of the Philadelphia translocation in myelogenous and acute lymphoblastic leukemia. Cancer Cells **7**:21–26.
21. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46**:885–894.
22. **Ishimi, I., and A. Kikuchi.** 1991. Identification and molecular cloning of a yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. J. Biol. Chem. **266**:7025–7029.
23. **Jeffreys, A., and R. A. Flavell.** 1977. A physical map of the DNA regions flanking the rabbit B-globin gene. Cell **12**:429–439.
24. **Kadowaki, T., H. Kadowaki, and S. I. Tayler.** 1990. A nonsense mutation causing decreased levels of insulin receptor mRNA: detection by a simplified technique for direct sequencing of genomic DNA amplified by the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **87**:658–662.
25. **Kakizuka, A., W. H. Miller, K. Umesono, R. P. Warrell, S. R. Frankel, V. V. V. S. Murty, E. Dmitrovsky, and R. M. Evans.** 1991. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RARA with a novel putative transcription factor, PML. Cell **66**:663–674.
26. **Kamps, M. P., C. Murre, X. H. Sun, and D. Baltimore.** 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell **60**:547–555.
27. **Konopka, J. B., S. M. Watanabe, and O. N. Witte.** 1984. An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. Cell **37**:1035–1042.
28. **Krohne, G., and W. W. Franke.** 1980. A major soluble acidic protein located in nuclei of diverse vertebrate species. Exp. Cell Res. **129**:167–189.
29. **Kurzrock, R., J. U. Gutterman, and M. Talpaz.** 1988. The molecular genetics of Philadelphia chromosome-positive leukemias. N. Engl. J. Med. **319**:990–998.
30. **Lapeyre, B., H. Bourbon, and F. Amalric.** 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual

protein structure revealed by the nucleotide sequence. Proc. Natl. Acad. Sci. USA **84**:1472–1476.

31. **Lenstra, R., L. d'Auriol, B. Andrieu, J. Le Bras, and F. Galibert.** 1987. Cloning and sequencing of Plasmodium falciparum DNA fragments containing repetitive regions potentially coding for histidine-rich proteins: identification of two overlapping reading frames. Biochem. Biophys. Res. Commun. **146**:368–377.
32. **Nourse, J., J. D. Mellentin, N. Galili, J. Wilkinson, E. Stanbridge, S. D. Smith, and M. L. Cleary.** 1990. Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell **60**:535–545.
33. **Pendergast, A. M., A. J. Muller, M. H. Havlik, Y. Maru, and O. N. Witte.** 1991. BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. Cell **66**: 161–171.
34. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**:84–94.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Sandberg, A. A., R. Morgan, J. A. McCallister, M. B. Kaiser, and F. Hecht.** 1983. Acute myeloblastic leukemia (AML) with t(6;9) (p23;q34): a specific subgroup of AML? Cancer Genet. Cytogenet. **10**:139–142.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Shtivelman, E., B. Lifshitz, R. P. Gale, and E. Canaani.** 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature (London) **315**:550–554.
39. **Soekarman, D., M. von Lindern, S. Daenen, B. de Jong, C. Fonatsch, B. Heinze, C. R. Bartram, A. Hagemeijer, and G. Grosveld.** The translocation (6;9)(p23;q34) shows consistent rearrangement of two genes and defines a myeloproliferative disorder with specific clinical features. Blood, in press.
40. **Soekarman, D., M. von Lindern, D. van der Plas, L. Selleri, C. R. Bartram, P. Martiat, D. Culligan, R. A. Padua, K. P. Hasper-Voogt, A. Hagemeijer, and G. Grosveld.** DEK-CAN rearrangement is restricted to translocation (6;9)(p23;q34). Leukemia, in press.
41. **von Lindern, M., and C. R. Bartram.** Unpublished data.
42. **von Lindern, M., M. Fornerod, S. van Baal, M. Jaegle, T. de Wit, A. Buijs, and G. Grosveld.** 1992. The translocation t(6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, *dek* and *can*, and the expression of a chimeric, leukemia-specific *dek-can* mRNA. Mol. Cell. Biol. **12**:1687–1697.
43. **von Lindern, M., A. Poustka, H. Lerach, and G. Grosveld.** 1990. The (6;9) chromosome translocation, associated with a specific subtype of acute nonlymphocytic leukemia, leads to aberrant transcription of a target gene on 9q34. Mol. Cell. Biol. **10**:4016–4026.
44. **von Lindern, M., T. van Agthoven, A. Hagemeijer, H. Adriaansen, and G. Grosveld.** 1989. The human pim-1 gene is not directly activated by the translocation (6;9) in acute nonlymphocytic leukemia. Oncogene **4**:75–79.
45. **Walker, L. C., T. S. Ganesan, S. Dhut, B. Gibbons, T. A. Lister, J. Rothbard, and B. D. Young.** 1987. Novel chimaeric protein expressed in Philadelphia positive acute lymphoblastic leukaemia. Nature (London) **329**:851–853.