MOLECULAR AND CELLULAR BIOLOGY, Dec. 1990, p. 6426–6435
0270-7306/90/126426-10$02.00/0

Vol. 10, No. 12

The SCL Gene Is Formed from a Transcriptionally Complex Locus

PETER D. APLAN,[1,2] C. GLENN BEGLEY,[3] VIRGINIA BERTNESS,[1] MARIANN NUSSMEIER,[4] ANGEL EZQUERRA,[5] JOHN COLIGAN,[5] AND ILAN R. KIRSCH[1]*

Navy Medical[1] and Pediatric Oncology[2] Branches, National Cancer Institute, Bethesda, Maryland 20814; Walter and Eliza Hall Institute, Victoria, 3050, Australia[3]; Department of Pathology, UCLA Medical Center, Los Angeles, California 90024[4]; and Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20814[5]

Received 2 July 1990/Accepted 12 September 1990

We describe the structural organization of the human SCL gene, a helix-loop-helix family member which we believe plays a fundamental role in hematopoietic differentiation. The SCL locus is composed of eight exons distributed over 16 kb. SCL shows a pattern of expression quite restricted to early hematopoietic tissues, although in malignant states expression of the gene may be somewhat extended into later developmental stages. A detailed analysis of the transcript(s) arising from the SCL locus revealed that (i) the 5′ noncoding portion of the SCL transcript, which resides within a CpG island, has a complex pattern of alternative exon utilization as well as two distinct transcription initiation sites; (ii) the 5′ portions of the SCL transcript contain features that suggest a possible regulatory role for these segments; (iii) the pattern of utilization of the 5′ exons is cell lineage dependent; and (iv) all of the currently studied chromosomal aberrations that affect the SCL locus either structurally or functionally eliminate the normal 5′ transcription initiation sites. These data suggest that the SCL gene, and specifically its 5′ region, may be a target for regulatory interactions during early hematopoietic development.

Over the past year, a new gene family whose members have roles as proven or putative transcription factors has been defined on the basis of the sharing of a consensus helix-loop-helix (HLH) DNA-binding and dimerization motif within the protein structure. This family was first recognized as uniting an array of genes stretching from *Drosophila* spp. to humans, each of which was either known or presumed to play a role in the growth or differentiation of the particular system or tissue in which it was expressed (32). This family includes the *Drosophila* Achaete-scute, twist, and daughterless genes, the c-, N-, and L-*myc* genes of higher organisms, the MyoD group of muscle differentiation genes, and the immunoglobulin enhancer-binding proteins E12 and E47. A recent count identified at least 24 distinct genes as members of this HLH family (4). We have reported that the SCL gene is a full-fledged member of this family (3).

We originally identified SCL as a result of its disruption by chromosomal translocation t(1;14), associated with the occurrence of a stem cell leukemia (2). We have reported that expression of the normal SCL gene is quite restricted, appearing predominantly in cells, cell lines, and tissues of the early hematopoietic lineage (3). Experiments in progress have demonstrated that a CD34+ sorted population of human cord blood expresses SCL (V. Bier, T. Waldmann, P. D. Aplan, and I. R. Kirsch, unpublished results). The disruption of the gene in malignant cells capable of lymphoid, myeloid, monocytoid, or erythroid differentiation (19, 24), its restricted pattern of expression, and its identification as a member of a family of putative transcription factors have suggested the possibility that SCL plays a role in some facet of early hematopoietic development. We wish to determine what regulates expression of the SCL gene and what the gene regulates. As a start to this investigation, we have determined the structural organization of the human SCL gene at the level of the genomic DNA and expressed mRNA. We find that the SCL transcript unit is notable for a marked complexity of alternative exon utilization at its 5′ end. This complexity and the patterns of mRNA transcription that we now recognize are likely to be significant for understanding the regulation and lineage-specific expression of this gene.

* Corresponding author.

MATERIALS AND METHODS

DNA and RNA isolation. DNA and RNA were isolated from cell lines and tissues by the guanidium isothiocyanate method (14).

Southern and Northern (RNA) blotting. DNA samples (10 μg) were digested to completion as recommended by the supplier (Bethesda Research Laboratories), size fractionated on agarose gels, and transferred to nitrocellulose membranes by the Southern method (40). RNA samples (10 μg of total cellular RNA) were size fractionated on a 1.0% agarose-formaldehyde gel and transferred to nitrocellulose membranes (14). Hybridizations were performed with nick-translated ^{32}P-labeled probes and washed under standard conditions (2), with the highest-stringency washes consisting of 0.1% sodium dodecyl sulfate (SDS) and 0.1× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 52°C.

Genomic and cDNA cloning. The phage library used to obtain genomic clones was obtained from DU528 genomic DNA, partially digested with *Mbo*I, size fractionated, and ligated into lambda vector EMBL3 as described previously (2). The K562 cDNA library was an oligo(dT)-primed library obtained from Clontech. RNA for the human bone marrow cDNA library was isolated from a patient with Ewing's sarcoma, recently recovered from chemotherapy. Morphologic examination showed a dramatically left-shifted marrow with an myeloid/erythroid cell ratio of 6:1. RNA was purified over an oligo (dT)-cellulose column, and 5 μg, primed with oligo (dT) and random hexamers, was used to synthesize first-strand cDNA. Second-strand synthesis, *Eco*RI methylation, ligation into the *Eco*RI site of lambda ZAP2, and

packaging were done according to standard techniques (Stratagene).

Subcloning into plasmids. Fragments of the recombinant phage clone were subcloned into plasmid vectors pSP72, pGem 7Zf (Promega), and pBluescript SK (Stratagene).

Sequence analysis. Plasmids were sequenced by the dideoxy-chain termination method, using Sequenase (U.S. Biochemical Corp.) polymerase according to standard sequencing protocols and oligonucleotide primers. An IBM PS2 with PC-Gene (Intelligenetics) was used for sequence analysis and manipulation. Genomic and cDNA sequences were compared with those in the GenBank data base.

RNase protection assay. Relevant genomic or cDNA restriction fragments were subcloned into plasmids, and uniformly labeled ^{32}P antisense RNA was synthesized by using T7, SP6, or T3 RNA polymerases and Gemini riboprobe reagents (Promega). The radiolabeled antisense RNA (1×10^5 to 2.5×10^5 cpm) was hybridized to 30 μg of test RNA samples for 12 to 16 h at 50 to 65°C. The samples were then subjected to RNase A digestion and size fractionated on a 6% acrylamide–7 M urea denaturing gel (37).

Primer extension assay. Synthetic oligonucleotides were end labeled with [^{32}P]ATP to a specific activity of 10^8 cpm/μg, using T4 kinase. A 2-ng amount of this primer was annealed to 30 to 50 μg of total cellular RNA for 45 min at 65°C and then for 45 min at 45°C; it was then slowly cooled to room temperature in 250 mM KCl–10 mM Tris (pH 8.0)–1 mM EDTA (34). Reverse transcription was accomplished with 10 U of avian myeloblastosis virus (Promega) in a buffer containing 70 mM KCl, 50 mM Tris (pH 8.0), 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.25 mM deoxynucleoside triphosphates, and 50 μg of dactinomycin per ml and incubated for 5 min at 37°C, followed by 60 min at 42°C. The RNA template was then digested with RNase A (10 μg) for 15 min at 37°C, followed by phenol-chloroform extraction, ethanol precipitation, and size fractionation on a 6% acrylamide–7 M urea denaturing gel.

Oligonucleotide synthesis. Oligonucleotides were synthesized on a DNA synthesizer (model 380B; Applied Biosystems) and used without further purification.

Anchored PCR. A modification of the anchored polymerase chain reaction (PCR) technique (18) was used to obtain full-length cDNA clones. Total cellular RNA (10 μg) from the erythroleukemia cell line HEL (29) was annealed to an SCL-specific primer (nucleotides 136 to 153 of Fig. 2), and first-strand cDNA synthesis was carried out with Moloney murine leukemia virus reverse transcriptase, using the protocol recommended by the manufacturer (Bethesda Research Laboratories), along with 2.5 μCi of [^{32}P]dCTP as a tracer. First-strand cDNA purification and tailing with dATP were done as described previously (18). PCR amplification was performed by using a 5′ oligonucleotide, 5′-GACTC GACTCGACATCGATTTTTTTTTTTTTTTT-3′, containing *Sal*I, *Cla*I, and *Xho*I sites and a nested SCL 3′ oligonucleotide (nucleotides 122 to 138 of Fig. 2). Thermal cycling was carried out for 40 cycles (45 s at 95°C, 1 min at 50°C, and 2 min at 72°C), using a Perkin-Elmer-Cetus thermal cycler. The PCR products were extracted with phenol-chloroform, digested with *Cla*I and *Pst*I, and subcloned into *Psp*72 for sequence analysis.

In vitro translation and immunoprecipitation. In vitro translation of mRNA synthesized from cDNA clones was performed with the rabbit reticulocyte lysate system (Promega) in the presence of [^{35}S]methionine. Approximately 10^5 cpm was immunoprecipitated with 10 μl of normal rabbit serum or specific antisera as previously described (21). The immune complexes were collected with protein A-agarose beads (Bethesda Research Laboratories) and run under reducing conditions on a 10% sodium dodecyl sulfate-polyacrylamide gel.

RESULTS

Genomic organization of the SCL locus. Figure 1A shows the SCL locus and its intron-exon structure. The map was created by comparing 12 overlapping genomic phage clones with 20 cDNA clones isolated from the bone marrow cDNA library and a single K562 cDNA clone. The chromosome 1 probes used to screen the libraries (1.0BR, 1.4BK, 2.1SS, 1.0XX, and 0.9PX) are shown in Fig. 1B. The anchored PCR technique was used to obtain clones that elucidated the 5′ end of the SCL gene (Fig. 1D).

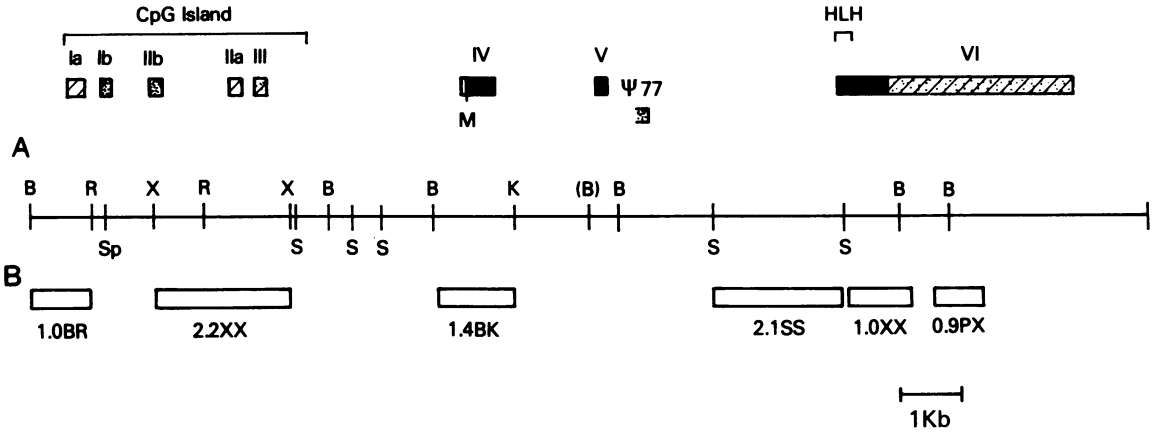
The SCL locus is composed of eight exons spanning 16 kb of genomic sequence (Fig. 1). Exon-intron boundaries were sequenced, and all contained the canonical GT splice donor and AG splice acceptor dinucleotides. The 5′ end demonstrates a complex pattern of mRNA splicing (Fig. 1D). (The sequences of these alternate exons are shown in Fig. 2.) The existence of these alternative mRNA species in several cell lines and tissue types has been confirmed by RNase protection assays (data not shown).

T-cell lines (Jurkat and HSB-2) investigated showed a splicing pattern distinct from those of other cell lines and tissues examined. While the multipotential HEL and K562 (25) cell lines initiated transcription at both exon Ia and Ib, the T-cell lines initiated transcription at exon Ib only. RNA obtained from bone marrow and peripheral mononuclear cells showed the same pattern as did the HEL and K562 cell lines. It should be noted that these alternative splicing events occur in the 5′ untranslated region (UTR), and all but one RNA species, regardless of the pattern of 5′ exon utilization, converge on exon III (Fig. 1D).

Several clones obtained from the bone marrow cDNA library spliced exon Ia to exon V. RNase protection assays using the longest of these clones as a probe revealed the existence of this mRNA species in bone marrow, peripheral mononuclear, fetal liver, K562, and HEL cells. In the positive tissues, this mRNA species consistently comprised 20 to 30% of the total SCL mRNA and predicts a truncated protein retaining the HLH motif, initiating at amino acid 176 of Fig. 2. This form initiating from exon Ia was not seen in the T-cell line Jurkat, HSB-2, or DU528.

As previously described (3), several cDNA clones had identical 297-bp deletions in their 3′ UTRs (nucleotides 1369 to 1666; Fig. 2). This deletion is felt to represent a bona fide mRNA splicing event, as it is bounded by consensus splice donor and acceptor sequences, and the mRNA species can be detected by using RNase protection probes. RNase protection experiments (data not shown) demonstrated that this species comprises <10% of the SCL mRNA.

Finally, a single clone isolated from the bone marrow cDNA library utilized an exon (designated Ψ77) distinct from the eight exons described above. This clone spliced exon Ψ77 directly to exon VI, using the same splice acceptor sequence as did the exon V-VI splice junction. However, exon Ψ77 contains an in-frame TGA termination codon upstream of the HLH motif and effectively eliminates the HLH homology region. This exon is identical to that presented by Chen et al. (10) and is likely to represent a nonfunctional transcript. An RNase protection probe derived from the cDNA showed this to be a very minor species in DU528, K562, and bone marrow RNA (data not shown).



C

Exon	3' Splice Acceptor	5' Splice Donor
Ia		GGAAAAACgtgagcat
Ib		TTGTGAGAgtgttcat
IIb	ctggtcagTTTCGTGG	CTCTTCGGgtgagcga
IIa	ccatccagGCCCGGGA	GTTGGAGGgttcgaaa
III	cgcggcagATCGCCCA	TGCCCCAGgtgagttt
IV	gcccacagGATGACCG	CTCGGCAGgttagggg
V	tcccacagCGGGTTCT	TACTGATGgtgagtct
VI	atctgcagGTCCCCAC	
(ψ77)		CCTTTGAGgtacttg

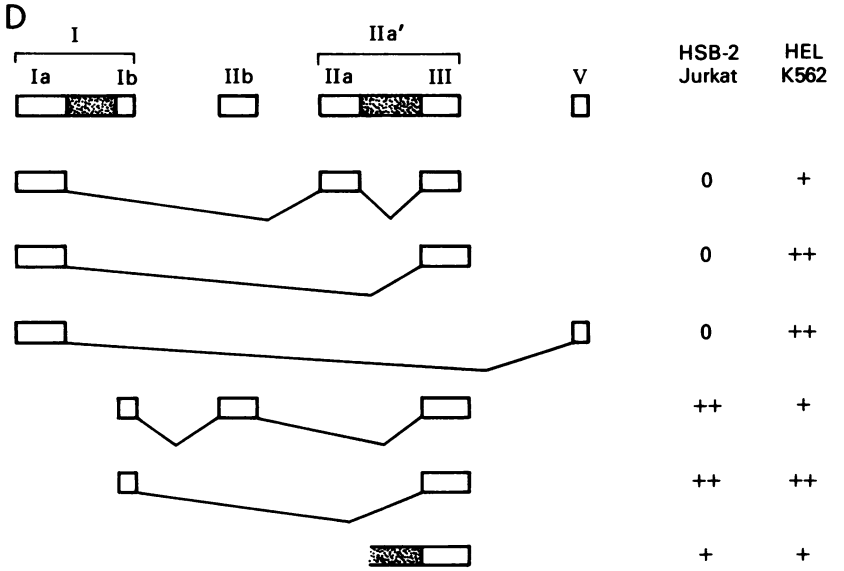


FIG. 1. Restriction map of the SCL genomic locus. Exons Ia to VI and pseudo-exon Ψ77 are shown above the restriction map. The protein-coding regions are indicated by solid boxes; the HLH homology region is positioned as shown. The alternatively spliced 5′ untranslated exons are represented by cross-hatched (Ia and IIa) and stippled (Ia and IIb) boxes indicating 5′ and 3′ UTRs. M, The initiation methionine. Restriction sites: B, *Bam*HI; K, *Kpn*I; R, *Eco*RI; S, *Sst*I; Sp, *Sph*I; X, *Xho*I. The *Bam*HI site in parentheses (B) is polymorphic. *Kpn*I, *Sph*I, and *Xho*I sites are not complete. (B) Probes used for hybridizations. (C) Nucleotide sequence at splice junctions. Exon sequences are in capital letters; intron sequences are in lowercase letters. (D) mRNA forms generated by alternative exon utilization; mRNA splicing occurs as depicted. The stippled regions between exons Ia and Ib and exons IIa and III indicate regions of contiguous genomic DNA that were protected in RNase protection assays. The protected fragments seen (data not shown) may represent additional mRNA forms that retain this intervening sequence or incompletely processed pre-mRNA. The relative abundance of each mRNA species is indicated by 0, +, or ++. Two patterns were seen, one typified by the T-cell lines Jurkat and HSB-2, and the other typified by K562 and HEL.

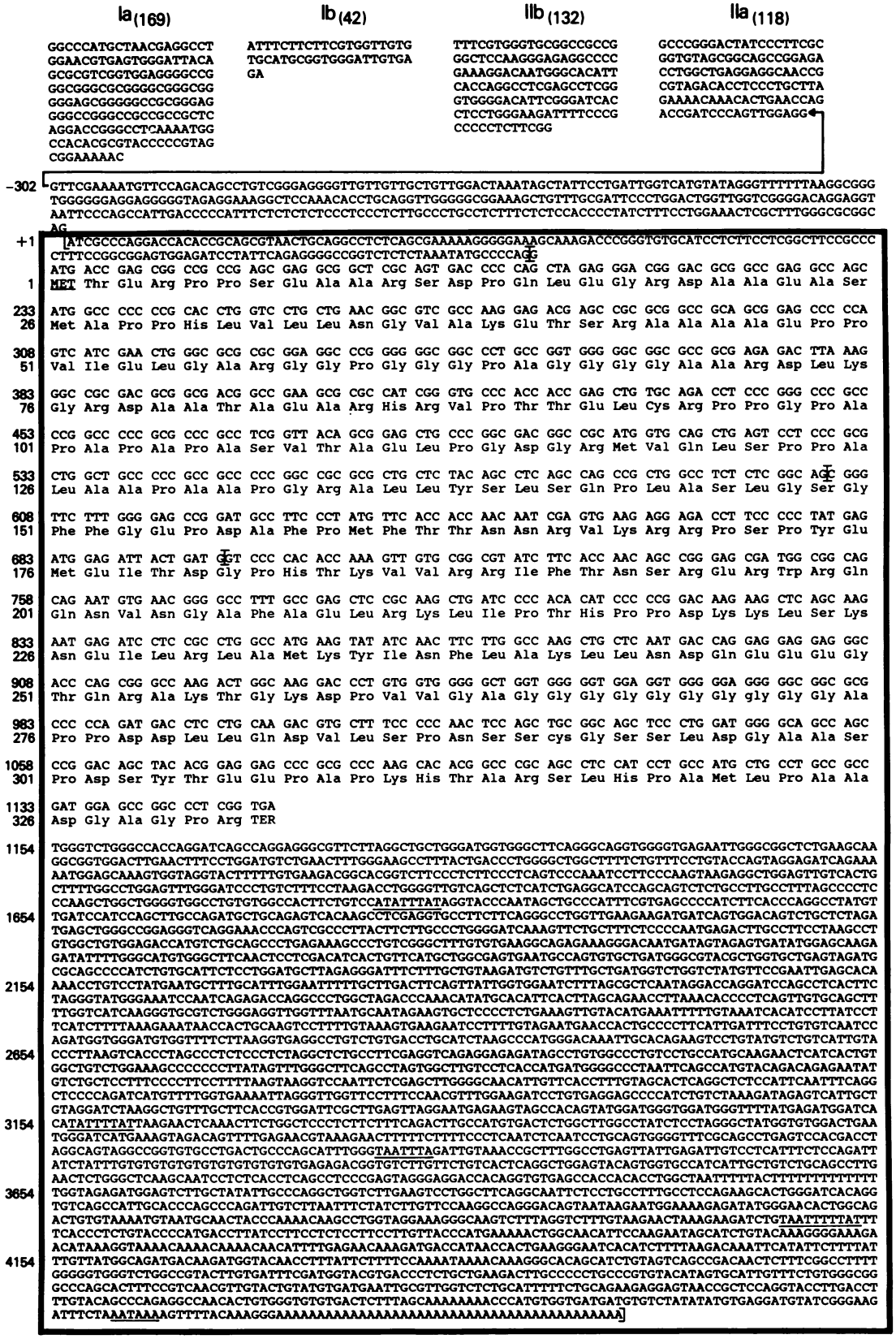


FIG. 2. Nucleotide sequences of full-length SCL cDNAs. The sequences of the alternative exons Ia, Ib, IIa, and IIB are as shown. Nucleotides -302 to -1 represent the contiguous intervening sequence between exon IIa and exon III (which begins with nucleotide +1). RNase protection and cDNA cloning experiments indicated that some mRNA forms retained this intervening sequence. Exons III to VI, which are common to all SCL mRNA species but one (see text), are outlined in the large box (nucleotides +1 to +4623). Exon junctions are indicated with brackets. Translation of the long open reading frame is as shown. A polyadenylation signal (AATAAA) is underlined, as are several sequences containing the ATTTA motif thought to destabilize mRNA (39).

Nucleotide sequence of the full-length SCL cDNA. We are now able to present the composite SCL cDNA, including the alternate exons seen in the 5′ UTR. A long open reading frame begins in exon IV, with the sequence AGGATGA, which is in good agreement with the consensus sequence for protein initiation (22, 23) and predicts a protein of 34 kDa. There is an in-frame TAA termination codon present 15 bases upstream. This cDNA sequence was compared with the murine SCL cDNA sequence (C. G. Begley et al., Proc. Natl. Acad. Sci. USA, in press). Both the ATG and the TAA termination codon are conserved in the murine SCL cDNA. The human SCL cDNA has an extensive 3′ UTR containing numerous stop codons in all reading frames, along with several TTATTTAT motifs, which have been found in mRNAs coding for cytokines (8) and are thought to destabilize cytoplasmic mRNA (38). An AATAAA polyadenylation signal is located 20 nucleotides 5′ of the poly (A) tail.

An in vitro-translated cDNA clone produces an SCL protein. A cDNA clone containing the entire SCL coding region was transcribed with T7 polymerase. The mRNA was then translated by using the rabbit reticulocyte lysate system in the presence of [^{35}S]methionine. Immunoprecipitation using antisera raised from rabbits immunized with SCL-specific peptides (Fig. 3) demonstrated bands at 37 and 41 kDa, slightly larger than the 34-kDa molecular mass of our predicted protein. Preliminary experiments with calf alkaline phosphatase-treated immunoprecipitates indicated that the size discrepancy was not due to phosphorylation.

Determination of the transcript initiation site and 5′ regulatory features. Having determined the location of the first exons, we sought to define their 5′ extent. Exons Ia and Ib, along with 562 nucleotides of upstream sequence, are presented in Fig. 4. To determine the transcript initiation site(s), an oligomer complementary to nucleotides 464 to 488 was synthesized and used in primer extension analysis. Figure 5 shows a tight cluster of bands corresponding to an mRNA initiation site at nucleotide +447 of Fig. 4. The sequence surrounding the transcript initiation site, TTCATTTC is in good agreement with the consensus start site sequence of PyPyC(A/G)PyPyPyPy (12). In addition, a TATA box (ATATA) is present 35 nucleotides upstream, beginning at nucleotide +412. To confirm these findings, we carried out RNase protection assays, using a probe derived from a 252-bp genomic *Eco*RI-*Sph*I fragment (nucleotides 221 to 473). The protected fragments of 28 and 36 nucleotides seen in Fig. 5B are consistent with the start site determined by primer extension. In an analogous fashion, we determined the 5′ extent of exon Ia to be nucleotide +1 of Fig. 4 (data not shown).

Inspection of the sequence shown in Fig. 4 reveals TATA and CCAAT boxes 5′ of exons Ia and Ib, several Sp1 recognition sequences, a potential Ap1-binding site, and a potential Eryf 1-binding site. The potential binding site for Eryf 1 (also known as NF-ε1 and GF-1), a transcription factor that is felt to be specific for megakaryocytic and erythroid genes (28, 36), is present 5′ of exon Ia but not exon Ib.

The 5′ end of the SCL gene lies in an unmethylated CpG island. Sequence analysis of exons Ia to III demonstrated this region to be highly G+C rich (Fig. 2). Furthermore, restriction endonuclease analysis of phage clones encompassing this region revealed sites for several rare restriction enzymes, such as *Bss*HII, *Not*I, and *Sac*II, which recognize G+C-rich sequences. Segments of G+C-rich DNA, located at the 5′ end of transcript units, may represent CpG islands, which are thought to be genomic regions involved in gene regulation (9, 42). These islands tend to be unmethylated in tissues expressing the gene and heavily methylated in tissues not expressing the gene. To investigate the methylation status of the G+C-rich region noted above, we digested and size fractionated genomic DNA with *Msp*I, which cleaves the sequence 5′-CCGG-3′ regardless of methylation status, and *Hpa*II, which cleaves the same sequence only when the 3′ cytosine is unmethylated. Figure 6A shows Southern blots of these genomic digests hybridized to a probe (2.2XX) from this CpG island. The small fragments generated by *Msp*I or *Hpa*II digests of K562 (an SCL-expressing cell line) genomic DNA were identical. However, when genomic DNA from HL60 (11), a cell line that does not express SCL, was

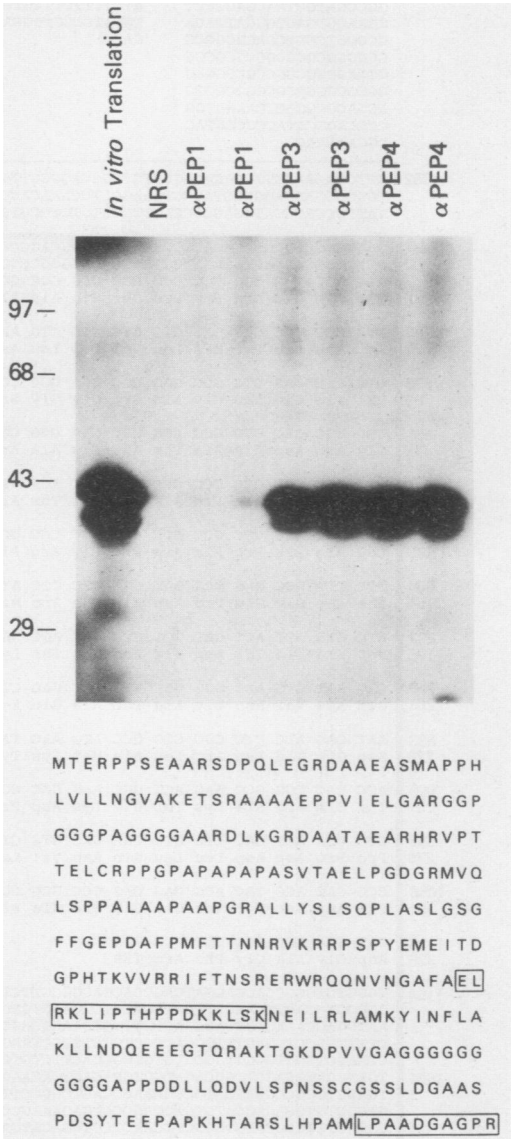
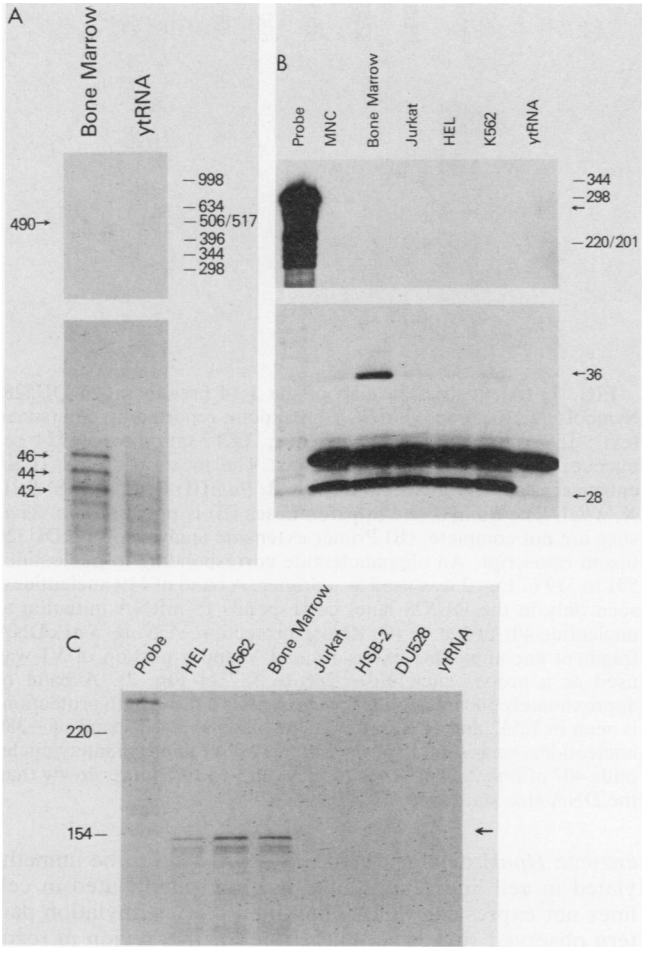


FIG. 3. Immunoprecipitation of an in vitro-translated SCL protein. An SCL cDNA clone was transcribed and translated in vitro in the presence of [^{35}S]methionine. Imunoprecipitation was carried out by using antisera from rabbits immunized with the indicated peptides. Peptide 1, A nonsense peptide from the 3′ UTR of SCL; peptide 3, SCL COOH terminus (LPAADGAGPR); peptide 4, a portion of the SCL HLH region (ELRKLIPTHPPDKKLSK); NRS, normal rabbit serum. The first lane represents the untreated (no immunoprecipitation) in vitro-translated product.

```
-562                                                    ..... CCTAGCAGAT AGGCATTTAT
-542        GTCCCTATTG ACAGGGTTCT GCCAGGCCGC CGTAGAGCCC ATGGATTTGA GTTACCGGCT CTATTTTAGA
-472        AGGGGGTGAA GAGGTAGGGA AGGGCTGGGG AGCTTTCTGG GATAGGGAGA CTGCCCATTG AAATCTGGGT
-402        GTCTGGGACA CTGGAGGAAG AATAAGAAAG AAGCCCTGGG AGGTGACGAT TTTGGTATCA TGGTGGTCTT
-332        CAGAGGTAGG GACAGAGTGC TGGGGTGAGA GTCGCTGAAG AAAGGTGTCT GAAAGGTTAA GGGACTGAGG
-262        TTGTGGGAAG CTGGGGTTCC TGAGAGGCCT TGGGATTAGA GATGGGGGCT GGATCTGCTG GGTCCTAGAC
-192        AGCGCGTCCA ACGCTCCCCG CAGCGCGTTC CGGAGCCGGT CCTAGCGCCG CTCAACCAGT CCCCACTCCC

-122        TCCGGTGAAA TTGCCAAATT AAAATGAATC ATTTGGCCCA TAATGGCCGA GGCGCTTATC GGGGGCGGGC

-52         GGACCCGCGG CAGTGCCTTA TCTCTGCGGC GCACACGGCC CCGGCGCGCC TCGGCCCATG CTAACGAGGC

+19         CTGGAACGTG AGTGGGATTA CAGCGCGTCG GTGGAGGGGC CGGGCGGGCG CGGGGCGGGC GGGGGAGCGG

+89         GGGCCGCGGG AGGGGCCGGG CCGCCGCCGC TCAGGACCGG GCCTCAAAAT GGCCACACGC GTACCCCCGT

                                                                                 EcoRI
+159        AGCGGAAAAA CGTGAGCATT CTGGCCTTTT TCTAGGGGAA AAGCAACCCG CGGCCTCCCA CCGAATTCTT

+229        CATCAATTTC CTCCTCTTTT TCCCCACGTC CGCAAAAGAG GTCTTCGCTC CCTTTCCCTC GGTTCCTTTC

+299        GATGGCCGGG GGGGGCGGTG GGGGGGCATT TTCCACGGAC GCCCCCGCCC CGGCTGCCGC CGCCCCTCCC

+369        GCCTCGGAGA CTCTCTTCCT TCCTTCCCCC TTTTCCTTAC GCAATATACA GAAATGCGCG AGGCTGTGGT
                                            SphI
+439        TGGTTTTCAT TTCTTCTTCG TGGTTGTGTG CATGCGGTGG GATTGTGAGA GTGTTCATGA GAACTGTGGG
```

FIG. 4. Nucleotide sequence surrounding exons Ia and Ib. Exons Ia (nucleotides +1 to +169) and Ib (nucleotides 447 to 488) are boxed. TATA and CCAAT boxes 5′ of the initiation exons Ia and Ib are boxed. Spl recognition sequences (GGGCGG) are underlined with dashes. A potential Eryf 1-binding site (TTATCT, complement AGATAA) is underlined, and a potential Apl-binding site (TGAATCA) is overlined.



analyzed in the same fashion, a different pattern emerged. The small *Hpa*II fragments were not seen in the HL60 cells (Fig. 6), indicating that the genomic DNA from this region is resistant to *Hpa*II digestion and therefore methylated in HL60. Since mitochondrial DNA is not methylated at CpG dinucleotides (26), we demonstrated complete digestion by hybridizing the Southern blots to a mitochondrial DNA probe (Fig. 6B).

The SCL transcript in DU528 initiates aberrantly, 3′ of the CpG island. The previously reported t(1;14)(p33, q11) translocation in the DU528 cell line juxtaposes the SCL locus, with the coding region and 5′ regulatory region intact, to the T-cell receptor delta locus, resulting in a fusion message, with the 3′ UTR of SCL fused to Dδ3 (Fig. 7) (3). Figure 7B and C show primer extension and RNase protection data demonstrating that the DU528 fusion message initiates aberrantly, from within exon IV. The increased quantity of SCL message and its aberrant initiation in DU528 may be a reflection of the influence of the t(1;14) translocation that

FIG. 5. (A) Primer extension using an oligonucleotide complementary to nucleotides 464 to 488. Prominent bands at 42, 44, and 46 nucleotides are seen. In addition a faint band of 490 nucleotides is seen. This may represent primer extension from an unspliced pre-mRNA initiating from exon Ia. (B) RNase protection demonstrating mRNA initiation at exon Ib. A 252-bp genomic *Eco*RI-*Sph*I fragment (see Fig. 4) encompassing exon Ib was used to synthesize an antisense RNA probe. A 28-nucleotide protected fragment is seen in all but the control (ytRNA) lane. A band at 36 nucleotides in the bone marrow lane most likely represents a minority of transcripts initiating 8 bases further upstream. An additional faint band showing full-length protection in the HEL lane (arrow) may represent protection of an unspliced pre-mRNA. Size standards were pBR322 cleaved with *Eco*RI and *Hin*fI, along with a conventional sequencing ladder. (C) RNase protection demonstrating cell-type-specific mRNA initiation. A genomic fragment (nucleotides −9 to +151 of Fig. 4) containing exon Ia was used to synthesize an antisense RNA probe. A band of 150 nucleotides, representing full-length protection of exon Ia (nucleotides +1 to +151), is seen in HEL, K562, and bone marrow cells but not in the T-cell line HSB-2, Jurkat, or DU528.

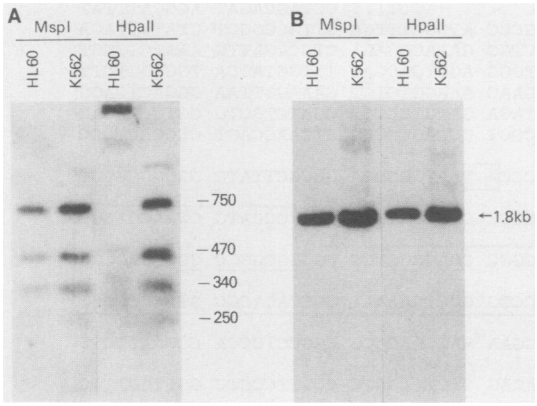


FIG. 6. (A) Genomic DNA from HL60 or K562 digested with *Msp*I or *Hpa*II and hybridized to a probe from the CpG island region (2.2XX of Fig. 1). Sizes in base pairs are indicated. (B) Same blot rehybridized to a mitochondrial DNA probe (see text).

brings the TCRδ enhancer (7, 35) to within 20 kb of the SCL exon IV.

DISCUSSION

We report here our characterization of the genomic structure of the SCL gene, with particular attention to its alternate mRNA initiation sites and processing. The SCL gene consists of eight exons distributed over 16 kb, encoding a protein with the recently described HLH motif. The 3′ portion of the gene is straightforward, with exons IV, V, and VI encoding the entire protein; there is an extensive 3′ UTR. The 5′ portion of the gene is complex, with alternate mRNA processing producing several different mRNA species. These alternate exons at the 5′ end likely account for the cluster of mRNA bands at 4.5 to 5.0 kb seen by us and others (3, 6, 10, 17). Unique 5′ UTR exons converging on a common coding sequence have been reported for several eucaryotic genes, from yeasts (yeast invertase gene) to mammals (murine α-amylase gene) (5, 13, 38). Tissue-specific, differential promoter utilization has been shown to be the cause of 5′ UTR heterogeneity in *Drosophila* alcohol dehydrogenase (5) and murine α-amylase (38) genes and is postulated to occur in the gene coding for common acute lymphoblastic leukemia antigen (13). The SCL gene would also seem to shown differential promoter utilization, with the exon Ia promoter being inactive in the T-cell lines (HSB-2, Jurkat, and DU528) analyzed. It is possible that these alternative mRNA species have different stabilities and half-lives, although this has not been proven. Furthermore, the 5′ UTR of SCL contains several ATG codons in good context for protein initiation, followed by stop codons, which may interfere with translation from the "true" initiation methionine (24), as has been shown for the proto-oncogene *lck* (27).

As mentioned above, no transcripts seen in T-cell lines initiated with exon Ia. However, cell lines K562 and HEL, both of which have the potential for erythroid and megakaryocytic differentiation (29, 43), do initiate transcripts at exon Ia. As previously noted, there is a consensus binding sequence (AGATAA) for Eryf 1 5′ of exon Ia but not 5′ of exon Ib. Therefore, it is possible that SCL transcripts initiating at exon Ia are mediated in part through Eryf 1.

A CpG island encompassing exons Ia to III was detected, and experiments with the methylation-sensitive restriction

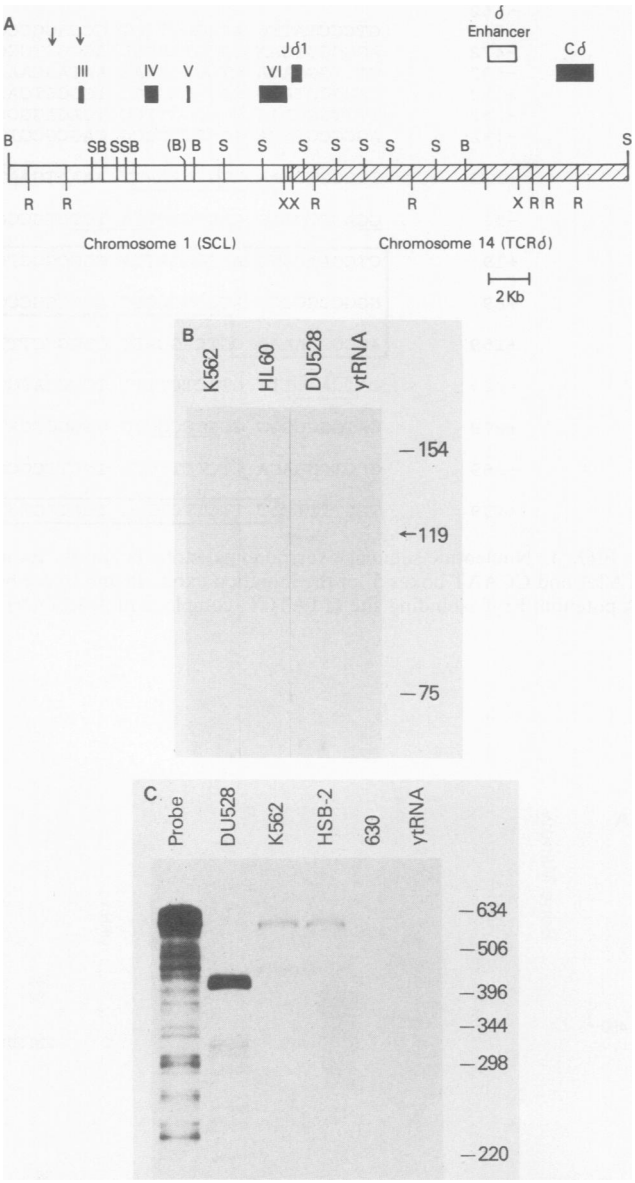


FIG. 7. (A) Restriction map of the 1;14 breakpoint in DU528. Symbols: ↓, location of an SCL breakpoint reported by others (see text); ▭, chromosome 1 sequence; ▨, chromosome 14 sequence; ■, SCL and TCRδ exons. The location of the TCRδ enhancer is shown. Restriction sites: B, *Bam*HI; R, *Eco*RI; S, *Sst*I; X, *Xba*I. The *Bam*HI site in parentheses (B) is polymorphic. *Xba*I sites are not complete. (B) Primer extension analysis of the DU528 fusion transcript. An oligonucleotide corresponding to nucleotides 501 to 519 of Fig. 2 was used as a primer. A band of 119 nucleotides, seen only in the DU528 lane, corresponds to mRNA initiation at nucleotide 401 of Fig. 2. (C) RNase protection. A *Not*I- *Sst*I cDNA fragment encompassing exons IV and V and a portion of VI was used as a probe (nucleotides 286 to 787 of Fig. 2). A band of approximately 500 nucleotides, representing a full-length protection, is seen in K562 and HSB2; the DU528 lane shows a band of ~380 nucleotides, representing protection only 3′ of approximately nucleotide 407 of Fig. 2. (The RNA probes run 5 to 10% more slowly than the DNA size standards [37].)

enzyme *Hpa*II demonstrated the CpG island to be unmethylated in cell lines expressing SCL and methylated in cell lines not expressing SCL. The differential methylation pattern observed suggests a likely role for this region in regu-

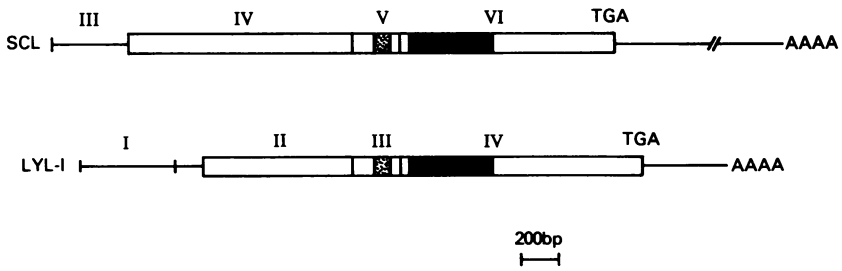


FIG. 8. Comparison of SCL and lyl-1 cDNAs. Symbols: —, 5′ and 3′ UTRs; □, protein-coding sequences; ■, the conserved HLH region; ▨, the additional upstream homology region (see text); |, exon boundaries.

lating SCL expression. This CpG island is disrupted by three recently published t(1;14) translocations in T-cell leukemia patients involving SCL and the TCRδ locus (6, 10). Two of the three patients showed an identical breakpoint at a cryptic heptamer within exon III, translocating all sequences 5′ of exon III to chromosome 14. The third patient had a breakpoint approximately 1 kb 5′ of exon III, which would also serve to translocate the normal initiation exons Ia and Ib to chromosome 14. In the DU528 t(1;14) translocation, the 3′ UTR of exon VI is disrupted by the translocation, and an SCL-TCRδ fusion message is produced. The DU528 fusion transcript also initiates aberrantly, at a site within exon IV (Fig. 7). The fact that the DU528 fusion transcript initiates aberrantly implies that normal SCL 5′ regulation is abrogated, possibly due to the influence of the TCRδ enhancer, now located 20 kb downstream of exon IV. Therefore, in a manner functionally similar but structurally distinct, the DU528 t(1;14) translocation would also seem to lose the normal SCL 5′ regulatory sequences contained within the CpG island.

Protein bands of 37 and 41 kDa were detected by immunoprecipitation with SCL-specific antisera. The migration of these bands on sodium dodecyl sulfate-polyacrylamide gels indicates a size slightly larger than the predicted size of 34 kDa; a similar phenomenon has been observed with c-, N-, and L-*myc* (reference 16 and references therein). In addition, we saw two distinct bands. These same two bands are seen when extracts from COS cells transfected with an SCL construct are immunoprecipitated (unpublished results). The lower band may simply represent a protein degradation product; however, it is interesting that an in-frame ATG in a good context for protein initiation (GCCAAGCATGG) is present 25 amino acids (2.7 kDa) downstream of the initiation methionine. Protein initiation from this ATG may account for the lower band, which is approximately 4 kDa smaller than the upper band. In addition to the HLH motif, SCL contains a proline-rich (12 of 44 amino acids; residues 89 to 132 of Figure 2) region near its amino terminus, similar to that seen in the activation domain of the transcription factors NF-1, AP-2, Jun, Oct-2, and SRF (31).

Part of the protein encoded by SCL shows a striking homology to part of the lyl-1 protein (30). Two regions of homology are noted; the first is the classic HLH motif, where SCL and lyl-1 have 84% identify over 58 amino acids, and the second region is a 12-amino-acid sequence located 10 amino acids upstream of the HLH motif (Fig. 8). Figure 8 shows the exon structures of SCL and lyl-1 to be quite similar. This conservation of nucleotide sequence and exon organization is suggestive of a common ancestral gene for SCL and lyl-1. There are several recent reports describing systems of structurally related HLH proteins regulating cell-type-specific differentiation, the most well-characterized being the MyoD, myf 5, myogenin, and Id gene products regulating muscle differentiation (4, 15). It is possible that SCL and lyl-1 are part of an analogous hematopoietic developmental system.

Several cell lines and tissues (but no T-cell lines) demonstrated an mRNA form that spliced exon Ia to exon V, predicting a truncated SCL protein initiating from amino acid residue 176 in Fig. 2, the first methionine encountered in a good context for protein initiation. This predicted protein would retain the HLH motif but initiate only 9 amino acids upstream of the HLH motif and therefore not contain a second basic domain conserved between SCL and lyl-1 (Fig. 8). Furthermore, this truncated protein would no longer contain the proline-rich region seen in the activation domain of some transcription factors. Loss of these regions may alter the function and functional interaction(s) of the SCL protein.

All four SCL translocations thus far characterized involve the TCRδ locus, which is thought to be the earliest TCR locus to undergo VDJ recombination in the course of normal lymphoid development (41). The proposed mechanism of these translocations involves a recombinase-mediated process (2, 6, 10) with cryptic heptamers and nonamers on chromosome 1. The fact that three of these translocations disrupt the 5′ unmethylated CpG island suggests that this region was in an accessible configuration, and therefore more susceptible to translocation, in an immature, possibly multipotential hematopoietic cell. This view is consistent with the premise (20) that chromosomal translocations do not occur randomly but instead occur preponderantly between regions of the genome that are transcriptionally active in the particular cell type that suffer the translocation.

In summary, we have presented data characterizing the SCL transcript unit, including cell-type-specific transcript initiation, alternate 5′ exon utilization, a methylation-sensitive CpG island, and a protein product that can be immunoprecipitated by antisera raised to SCL-specific peptides. Disruption of the gene by chromosomal translocation into the TCRδ locus, as has been shown to occur in T-cell leukemia patients, eliminates the normal sites of mRNA transcript initiation. It is possible that this change in the 5′ regulatory control of this gene contributes to the malignant phenotype observed in these patients.

ACKNOWLEDGMENTS

We thank Joseph Fedorko for oligonucleotide synthesis, Michael Birrer for assistance with COS cell transfections, and Marilyn Fourcault for her excellent work in preparation of the manuscript.

LITERATURE CITED

1. **Adams, R. A., A. Flowers, and B. J. Davis.** 1968. Direct implantation and serial transplantation of human acute lymphoblastic leukemia in hamsters, SB-2. Cancer Res. **28**:1121–1125.

2. **Begley, C. G., P. D. Aplan, M. P. Davey, K. Nakahara, K. Tchorz, J. Kurtzberg, M. Hershfield, B. F. Haynes, D. I. Cohen, T. A. Waldmann, and I. R. Kirsch.** 1989. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc. Natl. Acad. Sci. USA **86**:2031–2035.
3. **Begley, C. G., P. D. Aplan, S. M. Denning, B. F. Haynes, T. A. Waldmann, and I. R. Kirsch.** 1989. The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA **86**:10128–10132.
4. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
5. **Benyajati, C., N. Spoerel, H. Haymerle, M. Ashburner.** 1983. The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5′ end in different developmental stages. Cell **33**:125–133.
6. **Bernard, O., P. Guglielmi, P. Jonveaux, D. Cherif, S. Gisselbrecht, M. Mauchauffe, R. Berger, C.-J. Larsen, and D. Mathieu-Mahul.** 1990. Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. Genes Chrom. Cancer **1**:194–208.
7. **Bories, J. C., P. Loiseau, L. d'Auriol, C. Gantier, A. Bensussan, L. Degos, and F. Sigaux.** 1990. Regulation of translocation of the human T-cell antigen receptor delta chain gene. J. Exp. Med. **171**:75–83.
8. **Caput, D., B. Beutler, K. Hartog, R. Thayer, S. Brown-Shimer, and A. Cerami.** 1986. Identification of a common nucleotide sequence in the 3′-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA **83**:1670–1674.
9. **Cedar, H.** 1988. DNA methylation and gene activity. Cell **53**:3–4.
10. **Chen, Q., J.-T. Cheng, L.-H. Tsai, N. Schneider, G. Buchanan, A. Carroll, W. Crist, B. Ozanne, M. J. Siciliano, and R. Baer.** 1990. The tal gene undergoes chromosome translocation in T-cell leukemia and potentially encodes a helix-loop-helix protein. EMBO J. **9**:415–424.
11. **Collins, S. J., R. C. Gallo, and R. C. Gallagher.** 1977. Continuous growth and differentiation of human myeloid leukemic cells in suspension culture. Nature (London) **270**:347–349.
12. **Corden, J., B. Wasylyk, P. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein coding genes. Science **209**:1406–1413.
13. **D'Adamio, L., M. A. Shipp, E. L. Masteller, and E. L. Reinherz.** 1989. Organization of the gene encoding common acute lymphoblastic leukemia antigen (neutral endopeptidase 24.11): multiple mini-exons and separate 5′ untranslated regions. Proc. Natl. Acad. Sci. USA **86**:7103–7107.
14. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology. Elsevier Science Publishing Co., New York.
15. **Davis, R. L., P.-F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell **60**:733–746.
16. **DeGreve, J., J. Battey, J. Fedorko, M. Birrer, G. Evan, F. Kaye, E. Sausville, and J. Minna.** 1988. The human L-myc gene encodes multiple nuclear phosphoprotein from alternatively processed mRNAs. Mol. Cell. Biol. **8**:4381–4388.
17. **Finger, L. R., J. Kagan, G. Christopher, J. Kurtzberg, M. S. Hershfield, P. C. Nowell, and C. M. Croce.** 1989. Involvement of the TCL5 gene on human chromosome 1 in T-cell leukemia and melanoma. Proc. Natl. Acad. Sci. USA **86**:5039–5043.
18. **Frohman, M. A., M. K. Duyh, G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.
19. **Hershfield, M., J. Kurtzberg, F. Harden, J. Moore, J. Whang-Peng, B. F. Haynes.** 1984. Conversion of a stem-cell leukemia from lymphoid to a myeloid phenotype induced by the adenosine deaminase inhibitor 2′-deoxycoformycin. Proc. Natl. Acad. Sci. USA **81**:253–257.
20. **Kirsch, I. R., J. A. Brown, J. Lawrence, S. J. Korsmeyer, and C. C. Morton.** 1985. Translocations that highlight chromosomal regions of differentiated activity. Cancer Genet. Cytogenet. **18**:159–171.
21. **Koning, F., W. M. Yokoyama, W. L. Maloy, G. Stingl, T. J. McConnell, D. I. Cohen, E. M. Shevach, and J. E. Coligan.** 1988. Expression of C gamma 4 T-cell receptors and lack of isotype exclusion by dendritic epidermal T-cell lines. J. Immunol. **141**:2057–2062.
22. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
23. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8132.
24. **Kurtzberg, J., S. Bigner, and M. Hershfield.** 1985. Establishment of the DU.528 human lymphohemopoietic stem cell line. J. Exp. Med. **162**:1561–1578.
25. **Lozzio, C. B., and B. B. Lozzio.** 1975. Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome. Blood **45**:321–328.
26. **Markowitz, S., G. Krystal, A.-M. Lebacq-Verheyden, J. Way, E. A. Sausville, and J. Battey.** 1988. Transcriptional activation and DNase I hypersensitive sites are associated with selective expression of the gastrin-releasing peptide gene. J. Clin. Invest. **82**:808–815.
27. **Marth, J. D., R. W. Overell, K. E. Meier, E. G. Krebs, and R. M. Perlmutter.** 1988. Translational activation of the lck proto-oncogene. Nature (London) **332**:171–173.
28. **Martin, D. I. K., L. I. Zon, G. Mutter, and S. H. Orkin.** 1990. Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature (London) **344**:444–447.
29. **Martin, P., and T. Papyannopoulou.** 1982. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science **216**:1233–1235.
30. **Mellentin, J. D., S. D. Smith, and M. L. Cleary.** 1989. Lyl-1, a novel gene altered by chromosomal translocation in T-cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell **58**:77–83.
31. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
32. **Murre, C., P. Schonleber McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56**:777–783.
33. **Murre, C., P. Schonleber McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58**:537–544.
34. **Ntambi, J. M., S. A. Buhrow, K. H. Kaestner, R. J. Christy, E. Sibley, T. J. Kelly, Jr., and M. D. Lane.** 1988. Differentiation-induced gene expression in 3T3-L1 preadipocytes. J. Biol. Chem. **263**:17291–17300.
35. **Redondo, J. M., S. Hata, C. Brocklehurst, and M. S. Krangel.** 1990. A T-cell-specific transcriptional enhancer within the human T-cell receptor delta-locus. Science **247**:1225–1228.
36. **Romeo, P.-H., M.-H. Prandini, V. Joulin, V. Mignotte, M. Prenant, W. Vainchenker, G. Marquerie, and G. Vzan.** 1990. Megakaryocytic and erythrocytic lineages share specific transcription factors. Nature (London) **344**:447–449.
37. **Sambrook, J., E. F. Fritsch, T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Schibler, V., O. Hagenbuchle, P. K. Wellaver, and A. C. Pittet.** 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene Amy-1^a in the parotid gland and the liver. Cell **33**:501–508.
39. **Shaw, G., and R. Kamen.** 1986. A conserved AU sequence from

the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell **46**:659–667.
40. **Southern, E. M.** 1975. Detection of specific DNA sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
41. **Strominger, J. L.** 1989. Developmental biology of T-cell receptors. Science **244**:943–950.
42. **Tazi, J., and A. Bird.** 1990. Alternative chromatin structure at CpG islands. Cell **60**:909–920.
43. **Tetteroo, P. A. T., F. Massaro, A. Molder, R. Schreuder, and A. E. G. von dem Borne.** 1984. Megakaryoblastic differentiation of proerythroblastic K562-cell line cells. Leuk. Res. **8**:197–204.