The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif

C. Glenn Begley^{*†}, Peter D. Aplan[‡], Stephen M. Denning[§], Barton F. Haynes[§], Thomas A. Waldmann[†], and Ilan R. Kirsch^{*}

*Navy-Medical Oncology Branch, [†]Metabolism Branch, and [‡]Pediatric Branch, National Cancer Institute, National Naval Medical Center, Bethesda, MD 20814; and [§]Department of Medicine, Duke University, Durham, NC 27710

Contributed by Thomas A. Waldmann, October 3, 1989

ABSTRACT We have identified the human gene, SCL. We discovered this gene because of its involvement in a chromosomal translocation associated with the occurrence of a stem cell leukemia manifesting myeloid and lymphoid differentiation capabilities. Here we report the sequence of a cDNA for the normal SCL transcript, as well as for an aberrant fusion transcript produced in the leukemic cells. Although different at their 3' untranslated regions, both cDNAs predict a protein with primary amino acid sequence homology to the previously described amphipathic helix-loop-helix DNA binding and dimerization motif of the Ly1-1, myc, MyoD, immunoglobulin enhancer binding, daughterless, and achaete-scute families of genes. For these cDNAs, at least two different 5' ends are predicted, both of which retain this putative DNA binding domain and predict proteins in the range of 20-30 kDa. SCL mRNA is observed in "early" hematopoietic tissues. Taken together, these studies lead to the speculation that SCL plays a role in differentiation and/or commitment events during hematopoiesis.

The study of hematopoiesis is a prototype for the study of differentiation in general. A population of stem cells maintained throughout life remains capable of self-renewal and commitment to the lymphoid, myeloid, monocytoid, erythroid, or megakaryocytoid lineages. Numerous investigations are now focused on identifying the critical genes involved in the decisions governing the differentiation of blood-forming cells. Among the expected features of such genes might be their expression at a pivotal time in hematopoietic development. More intensive attention to such a candidate gene might be merited if it appeared to encode a protein whose features were similar to proteins already implicated as playing roles in other eukaryotic developmental systems. We describe one such candidate gene in this article.

Our approach to the question of cell-type-specific gene function is via the cloning and characterization of celltype-specific chromosomal translocations. The rationale behind these efforts is our belief that such translocations often highlight chromosomal regions of differentiated activity in the cells in which they occur (1). This premise has so far seemed particularly compelling in the translocations associated with hematopoietic malignancies (2). It was with this in mind that we undertook the cloning and characterization of a reciprocal translocation, t(1;14) (p33;q11.2), associated with the development of a stem cell leukemia in a 16-year-old male. This patient's leukemic cells and the cell line DU.528 subsequently derived from them were capable of responding to a variety of inducing agents by changing their phenotypic pattern from that of an early lymphoid cell to that of a cell of myeloid or monocytoid lineage (3-5). Our study identified a transcript unit on chromosome 1 abutting the translocation breakpoint

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

(6-8). We called this transcript SCL for stem cell leukemia. The essential facts and features of these earlier reports have been independently confirmed (9). The same probe that identified the transcript also identifies a single copy gene in other mammalian species by Southern blot analysis.

We have now cloned and sequenced both a normal SCL cDNA and a truncated form present in the leukemic stem cell line.[¶] Among hematopoietic tissues studied, SCL expression is not ubiquitous but is seen in phenotypically less mature T and myeloid cells and in cells with stem cell attributes. The cDNA sequence predicts a protein that contains a consensus sequence characteristic of other previously identified differentiation-related genes. The stem cell leukemia and derived cell line DU.528 contains a SCL gene with an altered 3' untranslated region. The expression of this altered mRNA is quantitatively increased over that of the message from the other allele.

MATERIALS AND METHODS

DNA (Southern) and RNA (Northern) blot hybridization analyses were performed as described (10, 11). cDNA libraries were constructed as described (12, 13). Specific primers, vectors, and sources for cDNA library construction are described in the text. The dideoxynucleotide chaintermination method (14) was used for DNA sequencing.

RESULTS

Sequence of a Normal SCL Gene. Initial experiments were performed to obtain clones of the normal SCL gene for nucleotide sequence determination. Tissues were examined as potential sources of RNA for construction of a cDNA library. Using probes derived from the region on chromosome 1 involved in the 1;14 translocation in DU.528, a transcript of >4 kilobases (kb) (just below the 28S RNA) was observed in bone marrow (BM), and the cell line K562, as well as the DU.528 cell line, but not in the T-cell line NL (Fig. 1). However, in the stem cell leukemia, DU.528, the predominant transcript was ≈ 2 kb (Fig. 1). The ≈ 4 -kb transcript in DU.528 appeared to be slightly smaller than that observed in other tissues.

Among the tissues examined, a particularly high level of SCL expression was noted in BM recovering from chemotherapeutic insult. A cDNA library was therefore constructed from BM mRNA obtained from a patient with Ewing sarcoma with no BM involvement. The BM was harvested during recovery from chemotherapy prior to autologous BM transplantation. The BM sample was hypercellular and normal immature myeloid elements predominated. There was no evidence of malignant cells. An oligo(dT) and random hexamer-primed cDNA library was constructed in λ ZAPII (Stratagene). In addition, an oligo(dT)-primed cDNA library

Abbreviation: BM, bone marrow.

[¶]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M29038).

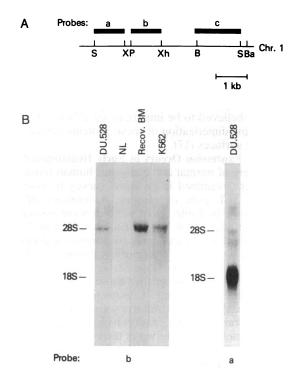


FIG. 1. (A) Map of germ-line chromosome 1 showing probes used for Northern blot analysis and for screening the cDNA libraries. B, Bgl II; Ba, BamHI; P, Pst I; X, Xba I; Xh, Xho I; S, Sst I. Not all sites for each enzyme are shown. Genomic clones were obtained from DU.528 DNA partially digested with Mbo I and cloned into EMBL-3. Chromosome 1 DNA was initially identified by using chromosome 14 probes to identify the 1;14 translocation in DU.528 as described (6). (B) Northern blot analysis of 10 μ g of total mRNA from human tissues and cell lines. Probes are shown in A. Probe a detects both the normal and aberrant messages, probe b only detects the normal message because it comes from 3' of the translocation breakpoint. Northern transfers were performed by standard techniques. Note transcript of >4 kb (below 28S rRNA) in DU.528, recovery BM, and K562. cDNA libraries from these three sources were screened to obtain the cDNA clones described below. Note also the absence of signal in NL (T-cell line) and intense 2-kb (near the level of the 18S rRNA) abnormal fusion transcript in DU.528 compared with the less intense >4-kb band. In poly(A)-selected RNA an additional band >5 kb was also seen. Hybridization to an actin probe revealed a comparable signal in all lanes (data not shown).

in $\lambda gt10$ from the CML/erythroleukemic cell line K562 (Clontech) was examined.

Twelve overlapping SCL clones were identified by using three probes from chromosome 1 (Fig. 1). One clone was from the K562 cDNA library (insert size, 2.7 kb) and 11 clones were from the recovery BM cDNA library. The inserts were overlapping and between 600 and 3000 base pairs (bp). Inserts were subcloned into the plasmid and the complete nucleotide sequence was determined in both orientations. An IBM PS2 with the PC-Gene (IntelliGenetics) program was used for data analysis and sequence comparison. GenBank and EMBL data bases were accessed via the Bionet National Computer Resource (15). A composite cDNA sequence for the SCL gene was derived from the BM clones. It and its predicted amino acid sequence are shown in Fig. 2.

The composite cDNA shown in Fig. 2 is 4195 nucleotides long. Among its distinguishing characteristics is a 3.4-kb 3' untranslated region. There were numerous stop codons in this region in all reading frames. Within this untranslated region, there was evidence of alternative splicing as manifested by a 297-bp region (nucleotides 940–1237) that was deleted in half the clones. Moreover, when the BM and K562 sequences were compared, there were several nucleotide

Proc. Natl. Acad. Sci. USA 86 (1989) 10129

| 1 | ASTCAGASTCACTTTCTSTAAATGSTACTTAGSTAGGGGGTCOGGCTCGGTTTACAGGGGAGCTGCCOGGGCAGGCCGC |
|--------------|--|
| 81 | ATG GTG CAG CTG AGT OCT OCC GOG CTG GCT GOC GOC GOC GOC GOC GOC GCG CTG CTC TAC AGC CTC AGC CAG Met Val Gin Leu Ser Pro Pro Ala Leu Ala Ala Pro Ala Ala Pro Giy Arg Ala Leu Leu Tyr Ser Leu Ser Gin |
| 156 | COS CTG GOC TET CTC GOC AGE GOS TTC TTT GOG GAG GOS GAT GOC TET CT ATG TTC AGE AAC AAT GOA GTG PTC Lau Ala Sar Lau Gly Sar Gly The The Gly Glu PTC Amp Ala The FTC Met Thr Thr Aan Aan Arg Val |
| 231 | MG AGG AGA CCT TOC COC TAY GAG ATG GAG ATT ACT GAT GOT COC CAC ACC ANA GTT GTG GOG GOT ATC TTC ACC Lys Arg Arg Pro Ser Pro Tyr Glu Met Glu Ile Thr Ago Gly Pro His Thr Lys Val Val Arg Arg Ile The Thr |
| 306 | MC MOC COG GMG COLA TOG COG CMG CMG MAT GTG MAC GGG GCC TITT GCC GMG CTC CGC AMG CTG MTC CCC ACA CMT Man Swar Arry Glu Arry Trp Arry Glu Glu Am Val Am Gly Alla Phe Alla Glu Lau Arry Lys Lau Ile Pro Thr His |
| 381 | COC COG GAC ANG ANG TCC ACC ANG ANT GAG ATC CTC COC CTG COC ATG ANG TAT ATC AAC TTC TTG GOC ANG CTG FTO FTO AND Lys Lys Law Ser Lys Ast Glu ILe Lew Arg Lew Ala Mer Lys Tyr ILe Ant Phe Lew Ala Lys Lew |
| 456 | CTC MAT GAC CAG GAG GAG GAG GAC ACC CAG GAG GAC AAG ACT GAC AAG GAC CCT GTG GTG GAC GAT GAT GAG GAC GAT GAT GAG GAT GAT GAT GAT GAT GAT GAT |
| 531 | GAA GOT GOG GGA GOG GOG GOG GOG GOC COC GAA GAC CTC CTG CAA GAC GTG CTT TOC COC AAC TOC ACC TOC GOC GLY GLY GLY GLY GLY GLY JAA PRO PARAPA APP LAN LEAN GLA MAP VAL LAN SAF PRO ARM SAF SAF SAF CYS GLY |
| 606 | NGC TOC CTG GAT GGG GGA GGC AGC CGG GAC AGC TAC AGG GAG GAG GGC GGC GGC AGG CAC AGG GGC GGC |
| 681 | CHT COT GOC ANG CHG COT GOC GOC GAT GGA GOC GOC COT GOG NGA His Pro Ala Net Leu Pro Ala Ala Asp Gly Ala Gly Pro Arg TER |
| 726 | TOGGTUTOGGCUNCHAGHTUNGCUNGHAGGOCGTTUTINGGUTGCUGGUTGCUNGTUGGCUNGTUGGGCUNGHTUGGCOGGUTCUGANGUN |
| 825 | AGEOGRACITICAACITTICCIGEATGICICEAACITTICGEAAGCCITTIACICACCCIGGGGCCGGCITITICCIGITICCCIGIACCACITAGGACTICAGA |
| 924 | |
| | |
| 1023 | CIGCTITIGGOCIGGASTTIGGGATOCCIGTCTTTCCTAAGACCIGGGSTIGTCAGCICICATCIGAGGCATOCAGCASTCTCGCCTTGCCTTTAGC |
| 1122 | OCTOCCAMPCT03CT03G5T03OCT05T05T05CCACTTCT05T0CATNTTTATAG5TNCCCANTNGCT6CCCATTT05T0A6600CATCTTCACCAA9G0C |
| 1221 | TATGTTGATCCARCCAGCTTGCCAGATGCTGCAGAGTCACAAGCCTGGAGTGCCTTCTTCAGGGCCTGGTTGAAGAAGATGATCAGTGGACAGTCTGC |
| 1320 1419 | TCTRATGACTGGCCGGGGGGGGGGGGGGGGGGGGGGGGGG |
| | |
| 1518 | GAGCAAGAGATATTTTGGGCATGTGGGCTTCAACTCCTCGACATCACTGTTCATGCTGGGGAGTGAATGCCAGTGTGCTGATGGGCGTACGCTGGTGCT |
| 1617 | GNETNGATEGGCAGCCCCATCTGTCCATTCTCCCGCATGCTTAGAGGGATTTCTTTGCTGTAAGATGTCTGTTTGCTGATGGTCTAGGTCTAGTCCGA |
| 1716 | ATTCAGCACAAAACCTGTCCTATGAATGCTTTGCATTGGAATTTTTGCTTGACTTCAGTINTTGGTGGAATCTTTAGGGCTCAATAGGACCAGATCC |
| 1815 | ACCTUALITETAGGITATGGGAAATCOMATCAGAGAGCAGGCCTGGCTAAGACCAAACATATGGACATTCACTTAGGAGAACCTTAAACACCCCTC |
| 1914 | AGTIGIOCAGCITTIGGICAICAAGGIGGGICIGGGAGGIIGGIITTAAIGCAAIAGIGCICOCCICIGAAAGIIGEACAIGAAAITTITIGIAAAI |
| 2013 | CACATOCTTATOCTTCATCTTTTAAAGAAATAACCACTGCAAGTOCTTTTGTAAAGTGAAGAATOCTTTTGTAGAATGAACCACTGCOOCTTCATTGAT |
| 2112 | TICCIGIGICAATCCAGATCGICGGATGIGGITTTCTTAAGGIGAGCCIGICIGIGACCIGCATCTAACCCCATGGGACAAATTGCACAGAAGTCCIG |
| 2211 | TATISTICTISTICATTISTACCCTTAASTCACCCTAGCCCTCTCCCCCCTCGACGTCACAGAGAGAGA |
| 2310 | CMCMCTCATCATCATCGCCGTCTCGCAAAGCCCCCCCTTATASTTTGGGCTTCAGCCTASTGGCCTGTCCTCACCATGATGGGGCCCTAATTCAGCCA |
| 2409 | TGTACKCACAGAGAATINTGTCTGCTCCTTTCCCCTTCTTAAGTAAGGTCCAATTCTCGAGCTTGGGGCAACATTGTTCACCTTTGTAGCACTCAGG |
| 2508 | CTCTCCATTCMATTTCAGGCTCCCCMGATCATGTTTTGGTGAAAATTMGGGTTGGTTCCTTTCCAACGTTTGGAAGATCCTGTGAAGACCCCCATCTGT |
| 2607 | CINAMENTACASTORTIGCTUTAGERICINAGECUSTUGCTUCACOSTOCRUTOSCUTURGUARIGAGARIGAGARGIAGOCACASIRIOGRUGOSUGA |
| 2706 | TGGGTTTTATGAGATGGATGACHIANTITTATTAAGAACTCAAACTTCTGGCTCOCTCTTCTTTOAGACTTGGCATGTGACTCTGGCTTGGCCTATCTCC |
| 2805 | TNGGGCTATGGTGTGGACTGAATGGGATCATGAAAGTACACAGTTTTGAGAACGTAAGAACTTTTTCTTTTCCCTCAATCCTCAATCCTGCAGTGGGGT |
| 2904 | TTOGCHGOETGAGTCCAOGAOCTAGGCAGTAGGCCGGTGGCCGACTGCCCAGCATTIGGGTAATTIACATTIGTAAACOOCTTTGGCCTGAGTTAATG |
| 3003 | AGATIGTOCICATITICIOXIGATIATICIATITIGIGIGIGIGIGIGIGIGIGIGIGIGIGIGICIGICI |
| 3102 | GOCNTCATTGCTGTCTGCAGCCTTGAACTCTGGGCTCAAGCAATCCTCTCACCTCAGCCTCCCGAGTAGGGAGCACAGTGTGTGAGCCACAAGTCT |
| 3201 | GGCTAATTTTACTTTTTTTTTTTTTTTTTTTGTAGAGATGGAGTCTTGCTATATTGCOCAGGCTGGTCTTGAAGTCCTGGCTTCAGGCAATTCTCCTGCCTT |
| 3300 | TGCCTCCACAAGCACTGGGATCACAGGTGTCAGCCATTGCACCCACACTGTCTTAATTTCTATCTTGTTCCAAGGCCAGGACAGTAATAAGAA |
| 3399 | TGGAAAAGAGATAATGGGAACACTGGCACACTGTGTGAAAATGTAAATGCAACTACCCAAAACAACCTGGTAGGAAAGGGCAAGTCTTTAGGTCTTTGTAA |
| 3498 | GAACTAAAGAAGATCTGTAATTTTTTTTTTTTTTTTCACCCTCTGTACCCCATGACCTTATCCTTCCT |
| 3597 | |
| 3696 | |
| 3795 | |
| 3894 | |
| | |
| 3993 | MANGNGGASTANCCOCTOCNGSTNOCTTGACCTTTGTIK/NGOCCNGAGGCCAACACTGTGGGTGTGTGACTCTTTAGCAAAAAAAA |
| 4092 | TGATGTGTCTATATATGTGAGGATGTATGGGGAGGATTTCTA <u>AATAA</u> AGTTTTACAAAGGGAAAAAAAAAAAAAAAAAAAA |
| 4191 | |

FIG. 2. Nucleotide sequence and predicted amino acid sequence for a SCL gene. One cDNA library was constructed from BM mRNA obtained during recovery from chemotherapy. Both random hexamer priming and oligo(dT) priming were performed and the cDNA was cloned into λ ZAPII (Stratagene). Recombinant clones (10⁶) were screened and 11 overlapping inserts were obtained. All were subcloned into pBluescript and sequenced in both directions with oligonucleotide primers. A K562 cDNA library was obtained from Clontech and screened. A single 2.7-kb insert was obtained and subcloned into pGEM7Zf and sequenced in both directions. The predicted amino acid sequence begins at the second ATG (the first ATG terminates within 12 nucleotides) and continues until the first in-frame TGA (nucleotide 723).

differences in the 3' untranslated region attributable to point mutations. The cDNA clones had a typical terminal polyadenylylation signal sequence (AATAAA) and poly(A) tail, thus delineating the 3' extent of the gene.

The 5' end of the gene appears to be more complex. Among cDNA clones extending far enough 5', there was consensus of sequence beginning at nucleotide 176 of Fig. 2 and extending 3'. To the 5' side of nucleotide 176, which corresponds to an RNA splice acceptor site in the equivalent genomic DNA (data not shown), at least two different classes

... ... AGG GGC CGG GCC GCC GCC GCT CAG GAC CGG GCC ... Arg Gly Arg Ala Ala Ala Ala Gln Asp Arg Ala

TCA AAA TGG CCA CAC GCG TAC CCC CGT AGC GGA AAA ACC Ser Lys Trp Pro His Ala Tyr Pro Arg Ser Gly Lys Thr

FIG. 3. A different 5' end than that shown in Fig. 2 for a SCL cDNA. The 5' end of this class of cDNA, which diverges at nucleotide 176 from that shown in Fig. 2, is found in genomic DNA ≈ 6 kb 5' of that point of divergence.

of cDNA are found. One is represented in Fig. 2. It contains an open reading frame beginning at an ATG at position 81 extending for 642 bp to an in-frame TGA at position 723. This region is predicted to be a potential coding region with >95%certainty by the method of Fickett (16).

As shown in Fig. 3 another class of cDNA clones was also identified, identical to those in Fig. 2 3' of nucleotide 176 but distinctly different 5' of this point. The occurrence of these distinct 5' ends of the SCL gene is consistent with a process of alternative RNA splicing. Both of the sequences 5' of the point of divergence between Figs. 2 and 3 have been linked in the genome to the main body of the gene. The sequence in Fig. 2 arises within 2 kb of the point of divergence (the splice acceptor mentioned previously). The sequence in Fig. 3 is found an additional 6 kb 5' of the point of divergence. RNase A protection analyses show protection occurring within each of these sequences (data not shown). Primer extension (data not shown) using a primer 3' from the point of divergence is consistent with an mRNA of ≈4200 nucleotides, the size of the composite cDNA shown in Fig. 2, which is therefore likely to represent one form of a full-length SCL cDNA.

The SCL Gene Encodes a Potential DNA Binding and Dimerization Motif. Part of the predicted SCL gene product showed striking homology to a recently described putative DNA binding domain of a number of interesting proteins (17). These include genes important in neurogenesis, germ-layer development, and sex determination in *Drosophila*; Lyl-1, a newly described gene active in T-cell acute lymphoblastic leukemia (see below); MyoD, a gene important in myogenesis; immunoglobulin enhancer binding proteins, and three *myc* family genes. The identity between this region of the SCL gene and the analogous domain of the T8 achaete-scute gene of *Drosophila* was 53% over 58 amino acids and the region of homology extended over 120 amino acids. There was 30% identity with MyoD over 120 amino acids and 49 conservative amino acid substitutions. Amazingly, the identity between this region of the SCL gene and the analogous region of Lyl-1 was 84% over 58 amino acids with the nonidentical residues representing mostly conservative changes (Fig. 4).

The likely structure of this group of proteins has recently been described in detail (17). As with the other members of this group, the SCL gene product fits the proposed amphipathic helix-loop-helix structure. Preservation of this helixloop-helix is believed to be important for DNA binding and may also allow dimerization of these proteins through their hydrophobic surfaces (17).

SCL Gene Expression Occurs in Early Hematopoietic Tissues. A variety of normal and malignant human tissues and cell lines were examined in an initial survey to assess the spectrum of SCL gene expression. A summary of these results is shown in Table 1. All samples were assessed by Northern blot analysis of 10 μ g of total mRNA or 2 μ g of poly(A) selected RNA; examples of positive and negative results are shown in Fig. 1. In normal tissues, the highest levels of expression on a message per μ g of RNA basis were observed in fetal liver, higher than those seen in BM during recovery from chemotherapy. Control tissues from nonhematopoietic fetal tissue (extremity) were negative, as were adult liver, brain, thymus, and activated T cells. Of the malignant tissues examined, two of three myeloid leukemias were positive. One was classified as FAB M2, and one was classified as FAB M5 but it was also positive for the early T-cell marker CD7. A T-cell acute lymphoblastic leukemia sample was positive, while other B- and T-cell tumors and the epithelial tumors examined were negative. These results were also supported by examination of cell lines. Mature Band T-cell lines were negative, while K562 and two CD7⁺. CD3⁻, CD4⁻, CD8⁻ cell lines (DU.528 and HSB.2) were positive. In addition, three neuroendocrine cell lines (two medulloblastoma, one small cell lung carcinoma) were positive for SCL gene expression although other medulloblastoma and neuroblastoma cells did not show SCL transcripts. The SCL gene was therefore not expressed ubiquitously, and within normal, malignant, and cell lines derived from hematopoietic tissues was seen in early T cells, early myeloid cells, and cells with stem cell attributes.

Sequence of an Aberrant 2-kb SCL Gene. Experiments were performed to characterize the aberrant 2-kb fusion transcript in the stem cell leukemia DU.528. An oligo(dT)-primed cDNA library from DU.528 mRNA was prepared in λ gt10 and screened with a chromosome 1 probe (Fig. 1, probe). Inserts were between 500 and 2200 bp and were subcloned

| SCL | VRRIFTNSRERWRQQNVNGAFAELRKLIPTHPPDKKLSKNEILRLAMKY | INF <u>L</u> AKLLN |
|-----------|--|--------------------|
| CONSENSUS | R N ER R Ψ F L Ψ KΨ IL AΨ Y T L T V | Ψ L V |
| Lyl-1 | ARRVFTNSRERWRQQNVNGAFAEERKLLPTHPPDRKLSKNEVERIAMK | TGERVELLE |
| c-myc | VKRRTHWIERORRNELKRSFFALROOIPELENNEKAPKVVILKKATAY | |
| L-myc | TKRKNHNFLERKRRNDLRSRFLALRDOVPTLASCSKAPKVVILSKALEY | |
| N-myc | ERRNHNILERORNDLRSSFLTLRDHVPELVKNEKAAKVVILKKATEX | . ~ |
| MyoD | DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNORLPKVETLRNAI RY | |
| E12 | ERRVANNARERIRVRDINEARKELGRMCQLHLNSEKPQTKLLILHQAVSV | ILNLEQOVR |
| twist | NORVMANVRERORTOSLNDAFKSLOOIIPTLPSDKLSKIOTLKIATRY | |
| da | ERROANNARERTRIRDINEALKELGRMCMTHLKSDKPOTKLGILMMAVEN | IMTLEQQVR |
| AS-C T5 | VIRNARERNRVKQVNNGRSQLRQHIPAAVIADLSNGRRGIGPGANKKLSKVSTLKMAVEY | IRRLQKVLH |
| AS-C T4 | VORRNARERNRVKQVNNSFARLROHIPOSIITDLTKGG-GRGPHKKISKVDTLRIAVEY | IRSLODLVD |
| | | |
| | | |
| | | |
| | HYDROPHILIC AMPHIPATHIC AMPHIPA | THIC |

| • | DIGITITI | |
|---|----------|---------|
| | DOMATN | HELTX T |

AMPHIPATHIC HELIX II

FIG. 4. Amino acid sequence relationship between SCL and a variety of other proteins. Comparison of amino acid residues (single-letter code) of a predicted SCL gene product with regions of the achaete-scute, daughterless, and twist genes of *Drosophila* and the immunoglobulin enhancer binding protein E12, MyoD, N-myc, L-myc, c-myc, and Lyl-1. The conserved amino acid residues are boxed. The hydrophilic domain and the two predicted amphipathic helices are indicated. ψ , Hydrophobic residues. Data are from Murre *et al.* (17).

| Table 1. | Spectrum | of | expression | of SCL |
|----------|----------|----|------------|--------|
| | | | | |

| Tissue or cell line | SCL expression | Tissue or cell line | SCL expression |
|--|-------------------|---|-------------------|
| Normal tissue | | | |
| Fetal liver (10 and 12.75 weeks) | + | Fetal extremity (10 weeks) | _ |
| Recovery BM [poly(A) and total] | + | Thymus tissue (<1-yr-old child) | _ |
| Term placenta [poly(A)] | ± | Unfractionated thymocytes | - |
| Mononuclear cells (peripheral blood) | + | CD3 ⁻ , CD4 ⁻ , CD8 ⁻ thymocytes | - |
| Neutrophils | ± | Brain (hippocampus) [poly(A)] | - |
| Phytohemagglutinin stimulated peripheral blood | - | Adult liver [poly(A)] | - |
| Malignant tissue | | | |
| AML, M5, CD7 ⁺ | + | CML | |
| AML, M2 | + | pre-B-cell ALL | _ |
| T-cell ALL | + | AML, M5, CD7 ⁺ | - |
| Burkitt lymphoma [poly(A)] | _ | Mycosis fungoides [poly(A)] | _ |
| ATL | _ | SCC [poly(A)] | - |
| Neuroblastoma | _ | Medulloblastoma | - |
| Cell line | | | |
| DU.528 [poly(A) and total] | + | H929 [poly(A)] | - |
| HSB-2 | + | HL60 | - |
| K562 | + | NL [poly(A) and total] | - |
| TE671 | + | NALL-1 | _ |
| 592 [poly(A)] | + | CEM | _ |
| DAOY | + | SB | - |
| SUP-T1 [poly(A)] | _ | Hut 234 | - |

Spectrum of expression of the SCL gene. Northern blots were prepared by using $10-20 \mu g$ of total RNA or $2 \mu g$ of poly(A) RNA from normal and malignant tissues and cell lines (19–29). All tissues were obtained in accordance with the requirements of the Ethics Committee of the National Institutes of Health. Malignant tissues examined included acute myeloid leukemia (AML) FAB M2 and M5 (CD7⁺). Samples from patients with chronic myeloid leukemia (CML) (n = 1), acute lymphoblastic leukemia (ALL) (n = 4; 1 T cell, 3 pre-B cells), adult human T-cell leukemia virus type positive T-cell leukemia (ATL) (n = 1), squamous cell carcinoma of lung (SCC) (n = 1), neuroblastoma (n = 1), and medulloblastoma (n = 1) were examined. Cell lines included CD7⁺, CD3⁻, CD4⁻, and CD8⁻ cells (DU.528, HSB-2), K562 (erythroleukemia of CML origin), HL60 (promyelocytic leukemia), SUP-T1, NL, CEM (T-cell lines), H929 (plasma cell), NALL-1, SB (B cells), Hut-234 (melanoma), 592 (small cell lung cancer), TE671, and DAOY (medulloblastoma). Blots were examined with the probes shown in Fig. 1A and were interpreted as positive (+) or negative (-) relative to the examples shown in Fig. 1B. A negative result is not intended to be a claim for zero relevant mRNA, but only for a level beneath the sensitivity of this assay. In two cases (normal placenta and neutrophils) the interpretation was equivocal (±). Integrity of RNA samples was assessed by ethidium bromide staining and by hybridization with an actin probe.

into plasmid for determination of nucleic acid sequence. A total of 14 clones were obtained and 6 were sequenced in both directions. The composite nucleotide sequence and predicted amino acid sequence of the aberrant 2-kb SCL message from the stem cell leukemia DU.528 are available on request. The cDNA at the point of fusion between SCL and T-cell receptor Δ diversity (TCR D δ) is shown in Fig. 5. The 297-bp region (between nucleotides 940 and 1237 in Fig. 2) of putative alternative splicing as noted previously for the normal BM cDNAs was deleted as determined by restriction endonucle-

ase map analysis in 6/14 clones and was present in 8/14 clones; 325 nucleotides beyond this region, the nucleotide sequence was of chromosome 14 origin. This sequence included the diversity (D) $\Delta 3$ gene and its flanking 3' genomic signal sequences and was identical to the previously described genomic sequence at the site of the chromosomal translocation. The sequences from chromosome 14 extended for an additional 293 bp. All clones had a poly(A) tail immediately beyond the genomic sequence AATACA, which served as a polyadenylylation signal.

FIG. 5. Nucleotide sequence at the point of fusion between SCL and T-cell receptor D δ . An oligo(dT)-primed cDNA library was constructed from DU.528 mRNA and cloned into λ gt10. Recombinant clones (10⁶) were examined and 14 clones were identified by probe A (see Fig. 1*A*). Six inserts were subcloned into pGEM7ZF and sequenced in both directions using oligonucleotide primers and dideoxynucleotide sequencing reactions. A portion of the normal chromosome 1 3' untranslated sequences was replaced (at nucleotide 1458 of Fig. 2) by the genomic D Δ 3 sequences from chromosome 14 (boxed) and the flanking 3' heptamer and nonamer signal sequences (boxed). The genomic chromosome 14 sequences (lowercase letters) continued for 293 bp and terminated with a poly(A) tail just beyond the genomic sequence AATACA (underlined). The cDNA sequence of the site of the translocation was identical to the genomic sequence and included 4 nucleotides of N-region diversity (capital letters, dashed box) present in the genomic sequence (6).

A second form of the SCL gene product was predicted based on analysis of a single DU.528 cDNA clone. A deletion of 100 nucleotides in the coding region of this clone resulted in a frameshift so that the TGA at position 723 ceased to be a termination codon and a larger protein with a different C terminus was generated. This larger form of the SCL protein nevertheless retained intact and unaltered the previously described DNA binding and dimerization motif.

Thus, the chromosomal translocation in the human stem cell leukemia served to disrupt the stem cell leukemia, gene and, as a result, a fusion transcript between sequences on chromosome 1 and chromosome 14 was generated. However, the translocation event into the 3' untranslated region preserved intact the putative SCL coding sequence.

DISCUSSION

Characterization of chromosomal translocations has allowed identification of genes critical to normal growth and differentiation. We speculated that in the human stem cell leukemia, DU.528 the translocation event at 1p33 might highlight a gene of relevance to this cell's interesting multipotential phenotype. In fact, as demonstrated in this report, the translocation occurred directly into the body of the SCL gene and generated a fusion transcript between sequence on chromosomes 1 and 14. The translocation involving the SCL gene in this case disrupted the 3' untranslated region, leaving the coding sequence and therefore the protein product preserved. In DU.528, a transcript is generated from both the allele involved in the translocation into the D $\Delta 3$ gene segment and the other allele. The \approx 4-kb transcript (which by analysis with probes 5' and 3' of the translocation could only come from the SCL chromosome 1 allele not involved in the 1:14 translocation) is smaller than the transcript observed in other tissues. In this regard, it is noteworthy that the second chromosome 1 in DU.528 is also karyotypically abnormal in the region 1p33. It is possible that both SCL alleles in the DU.528 cell line have been altered by gross chromosomal rearrangements. The level of expression of the abnormal 2-kb fusion transcript is, however, as much as 20-fold greater by densitometric analysis than the larger transcript.

The predicted SCL gene product shows an intriguing homology to other DNA binding proteins with conservation of a likely amphipathic helix-loop-helix DNA binding and dimerization motif (17). The other proteins included in this group appear to play a critical role in differentiation and/or commitment of specific tissues. Recently, another gene, Lyl-1, was described (30). It was discovered because of its presence at the site of a translocation breakpoint in the malignant cells of a patient with T-cell acute lymphoblastic leukemia. It is located on a different chromosome (chromosome 19) than SCL and is transcribed into a different sized RNA expressed in T cells. Yet, within its predicted helixloop-helix DNA binding region, it demonstrates remarkable similarity to SCL. Its discoverers speculate on the role of Lyl-1 in neoplastic transformation. Its analogous method of discovery, involvement with the T-cell receptor locus, expression in T cells, and striking similarity to SCL over a limited expanse of relevant protein domain leads to the speculation that these two genes may relate to each other in some cell-type-specific fashion. Other members of this family have recently been shown to be capable of structural interaction and heterodimer formation, which alters the DNA binding capability of each (18). In summary, the restricted pattern of expression of the SCL gene, its structural involvement in a stem cell leukemia, and its identification as a member of a family of genes sharing a DNA binding and dimerization motif strongly suggests that this gene may be important in hematopoietic differentiation or oncogenesis.

Note Added in Proof. We now believe that the sequence shown in Fig. 2 may not be a full-length cDNA. We have characterized an additional clone that extends further 5' and potentially encodes an additional 90 amino acids.

We wish to thank Dr. M. Israel for providing samples of TE671 and DAOY mRNA for this analysis. We wish to thank Ms. Marilyn Fourcault and Ms. Gail Gray for excellent assistance in the preparation of this manuscript. This work was supported in part by a grant from the National Health and Medical Research Council, Canberra, Australia, to C.G.B.

- 1. Kirsch, I. R., Brown, J. A., Lawrence, J., Korsmeyer, S. J. & Morton, C. C. (1985) Cancer Genet. Cytogenet. 18, 159–171. Nowell, P. C. & Croce, C. M. (1988) FASEB J. 2, 3054–3060.
- 2.
- Hershfield, M. S., Kurtzberg, J., Harden, E., Moore, J. O., 3. Whang-Peng, J. & Haynes, B. F. (1984) Proc. Natl. Acad. Sci. USA 81, 253-257.
- Kurtzberg, J., Bigner, S. H. & Hershfield, M. S. (1985) J. Exp. 4. Med. 162, 1561-1578.
- 5. Kurtzberg, J., Waldmann, T. A., Davey, M. P., Bigner, S. H., Moore, J. O., Hershfield, M. S. & Haynes, B. F. (1989) Blood 73, 381-390.
- Begley, C. G., Aplan, P. D., Davey, M. P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M., Haynes, B. P., Cohen, D. I., Waldmann, T. A. & Kirsch, I. R. (1989) Proc. Natl. Acad. Sci. USA 86, 2031-2035.
- 7. Begley, C. G., Aplan, P. D., Davey, M. P., deVillartay, J.-P., Cohen, D. I., Waldmann, T. A. & Kirsch, I. R. (1989) J. Exp. Med. 170, 339-342.
- Begley, C. G., Aplan, P. D., Waldmann, T. A. & Kirsch, I. R. 8. (1989) UCLA Symp. Mol. Cell. Biol. New Ser. 120, in press.
- 9. Finger, L. R., Kagan, J., Christopher, G., Kurtzberg, J., Hershfield, M. S., Nowell, P. C. & Croce, C. M. (1989) Proc. Natl. Acad. Sci. USA 86, 5039-5043.
- Southern, E. M. (1975) J. Mol. Biol. 98, 503-517. 10.
- Denny, C. T., Hollis, G. F., Magrath, I. T. & Kirsch, I. R. 11. (1985) Mol. Cell. Biol. 5, 3199-3207.
- 12. Okayama, H. & Berg, P. (1982) Mol. Cell. Biol. 2, 161-170.
- Gubler, V. & Hoffman, B. J. (1983) Gene 25, 263-269. 13.
- 14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. 15 USA 85, 2444-2448.
- Fickett, J. W. (1982) Nucleic Acids Res. 10, 5303-5318. 16.
- 17. Murre, C., McCaw, P. S. & Baltimore, D. (1989) Cell 56, 777-783.
- 18 Murre, C., McCaw, P., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Haushka, S. D., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) Cell 58, 537-544.
- Lozzio, C. B. & Lozzio, B. B. (1975) Blood 45, 321-334. 19.
- Jacobsen, P. F., Jenkyn, D. J. & Papadimitriou, J. M. (1985) J. 20. Neuropathol. Exp. 44,472-485.
- 21. Carney, D. N., Gazdar, A. F., Bepler, G. & Guccion, J. G. (1985) Cancer Res. 45, 2913-2923.
- Hecht, F., Morgan, R., Hecht, B. K. M. & Smith, S. D. (1984) 22. Science 226, 1445-1447.
- 23. Smith, S. D., McFall, P., Morgan, R., Link, M., Hecht, F., Cleary, M. & Sklar, J. (1989) Blood 73, 2182-2187.
- 24. Gazdar, A. F., Oie, H. K., Kirsch, I. R. & Hollis, G. F. (1986) Blood 67, 1542-1549.
- 25. Adams, R. A., Flowers, A. & Davis, B. J. (1968) Cancer Res. 28, 1121-1125
- 26. Foley, G. E., Lazarus, H., Farber, S., Uzman, B., Boone, B. & McCarthy, R. (1965) Cancer 18, 522-529.
- Adams, R. A. (1967) Cancer Res. 27, 2479-2482.
- Collins, S. J., Gallo, R. C. & Gallagher, R. E. (1977) Nature 28. (London) 270, 347-349.
- 29 McAllister, R. M., Isaacs, H., Rongey, R., Peer, M., Au, W., Soukup, S. W. & Gardner, M. B. (1977) Int. J. Cancer 20, 206-262.
- 30. Mellentin, J. D., Smith, S. D. & Cleary, M. L. (1989) Cell 58, 77-83.