MOLECULAR AND CELLULAR BIOLOGY, Oct. 1992, p. 4297–4304
0270-7306/92/104297-08$02.00/0

Vol. 12, No. 10

Characterization of Spi-B, a Transcription Factor Related to the Putative Oncoprotein Spi-1/PU.1

DOMINIQUE RAY,[1] RÉMY BOSSELUT,[2] JACQUES GHYSDAEL,[2] MARIE-GENEVIÈVE MATTEI,[3] ARMAND TAVITIAN,[1] AND FRANÇOISE MOREAU-GACHELIN[1]*

Faculté de Médecine Lariboisière Saint-Louis, INSERM U-248, 10 Avenue de Verdun, 75010 Paris,[1] Laboratoire d'Oncologie Virale et Cellulaire, Section de Biologie, Institut Curie, 91405 Orsay Cédex,[2] and Hôpital d'Enfants de la Timone, INSERM U-242, 13385 Marseille,[3] France

Received 27 April 1992/Returned for modification 29 May 1992/Accepted 10 July 1992

We have cloned a human cDNA from a new gene, *spi-B*, on the basis of its homology with the DNA-binding domain of the Spi-1/PU.1 putative oncogene product. *spi-B* codes for a protein of 262 amino acids presenting 43% overall identity with Spi-1. Its highly basic carboxy-terminal region exhibits 34% sequence identity with the DNA-binding domain of the Ets-1 protein. We showed that the Spi-B protein is able to bind the purine-rich sequence (PU box) recognized by Spi-1/PU.1 and to activate transcription of a reporter plasmid containing PU boxes. Chromosome in situ hybridization allowed us to map *spi-B* to the 19q13.3-19q13.4 region of the human genome. *spi-B*, like *spi-1*, was found to be expressed in various murine and human hematopoietic cell lines except T lymphoid cell lines.

spi-1 is a putative oncogene involved in the malignant erythroblastic transformation induced by the acutely leukemogenic Friend and Rauscher spleen focus-forming viruses (25, 26). It was originally identified as a genomic locus rearranged by spleen focus-forming proviral insertion in 95% of these tumors (25, 27). The transcriptional deregulation of *spi-1* by insertional mutagenesis results in overexpression of the Spi-1 protein (unpublished data). The product of *spi-1* is identical to the PU.1 transcriptional activator (11) that recognizes specific DNA sequences characterized by a purine-rich core sequence, 5′-GAGGAA-3′ (PU box), which acts as a Spi-1/PU.1-responsive element when linked in *cis* to the herpes simplex virus thymidine kinase promoter (18). The DNA-binding domain of Spi-1/PU.1 is localized to the carboxy-terminal half of the protein (18), a domain which includes a region of 87 amino acids related to that of the Ets proteins. The *ets* gene family includes c-*ets-1* and c-*ets-2* (5, 48, 49), *erg* (37), *elk-1* and *elk-2* (36), E74 (7), *fli-1* (2), *elg* (33), *elf-1* (45), and PEA3 (51) and encodes sequence-specific DNA-binding proteins. Ets-1 and Ets-2 (4, 12, 47), Elk-1 (14, 38), Erg (42), and PEA3 (51) have been shown to act as transcriptional activators. c-*ets-1* is the cellular progenitor of the v-*ets* oncogene of E26 (20, 29), a retrovirus which induces predominantly an erythroleukemia in chicken and transforms both myeloid and erythroid cells in vitro (35). In addition to v-*ets*, the E26 genome contains the v-*myb* oncogene, and both oncogenes are expressed as a nuclear Gag-Myb-Ets fusion protein (3). While v-*ets* as a unique oncogene can transform erythroid cells in vitro, it requires the contribution of v-Myb in a fusion protein to reproduce the fully leukemogenic potential of E26 in vivo (23). *fli-1* is activated by insertional mutagenesis in 75% of erythroleukemias induced by the Friend murine leukemia virus (2). The finding that three *ets*-related genes, *spi-1*, *fli-1*, and v-*ets*, are activated in virally induced murine and avian erythroleukemias strongly argues for their involvement in transformation of cells of the erythroid lineage.

Gene transcription is controlled by a variety of *cis*-regulatory DNA sequences that are recognized by transcriptional regulatory proteins. Transcription factors generally belong to families whose members share identical or very similar DNA-binding specificities. They contain DNA-binding domains characterized by specific motifs such as zinc finger, homeodomain, or the combination of a basic region with a helix-loop-helix, leucine zipper, or both (13, 24, 44). None of these classical motifs could be identified in the DNA-binding domain of the Ets transcription factors, suggesting that the Ets domain, extending over 87 amino acids in the carboxyl-terminal part of the protein, would represent a new DNA-binding motif (17). The DNA-binding domain of Spi-1 shares 40% sequence identity with the Ets domain of Ets-1. This degree of sequence conservation is the lowest among all Ets family members characterized until now, indicating that Spi-1 is actually the most divergent member of this family.

To search for genes related to *spi-1*, we screened a Burkitt lymphoma cDNA library with a human cDNA probe corresponding to the DNA-binding region of the Spi-1 protein. In this paper, we report the characterization of a human cDNA clone from the *spi-1*-related gene designated *spi-B*. We determined the human chromosomal localization of *spi-B* and analyzed its expression pattern by Northern (RNA) blot analysis in various cell lines and murine tissues. We also showed that Spi-B is able to bind oligonucleotides containing the PU boxes and to increase transcriptional activity of a reporter plasmid harboring PU boxes.

MATERIALS AND METHODS

cDNA library construction and screening. A λgt10 cDNA library was constructed from poly(A)$^+$ RNA extracted from Raji Burkitt lymphoma cells, using an Amersham cDNA synthesis and cloning system. The library (10^5 phages) was screened at reduced hybridization stringency (at 25°C in 25% formamide–5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]–2× Denhardt's solution–0.1% sodium dodecyl sulfate [SDS]) with a ^{32}P-labeled probe derived from the *Sma*I-*Nar*I fragment of the human *spi-1* cDNA and corresponding to the DNA-binding domain (39). Washings were in 2× SSC–0.1% SDS at 60°C. Secondary cDNA library

* Corresponding author.

screening under high-stringency conditions was performed at 42°C in 50% formamide–5× SSC–2× Denhardt's solution–0.1% SDS, and washings were done in 0.1× SSC–0.1% SDS at 60°C.

Genomic library screening. The human genomic library was constructed from liver DNA in the bacteriophage EMBL4 (a generous gift of Claudine Grégory). A total of 10^6 clones were screened under high-stringency conditions, as described above for cDNA library screening, with a ^{32}P-labeled probe corresponding to the *Eco*RI-*Pst*I fragment (nucleotides 1 to 263) derived from the *spi-B* cDNA clone reported in this paper. From one positive phage, a 5-kb *Eco*RI fragment that hybridized with the 5′ *spi-B* cDNA probe was subcloned into pBluescript vector and used for sequencing.

Subcloning and cDNA sequence determination. *Eco*RI inserts and restriction fragments of cDNA clones were subcloned into a pBluescript vector. Nucleotide sequence determination was performed on both DNA strands by the dideoxy-chain termination method (41), using [^{35}S]thio-dATP and a T7 DNA polymerase sequencing system (Pharmacia).

In vitro and in vivo expression. For in vitro expression, the 1.5-kb *spi-B* cDNA insert was subcloned into an *Eco*RI-digested pBluescript vector. RNA was generated from this plasmid in a T3 RNA polymerase reaction, using an mCAP transcription kit (Stratagene). In vitro translations were performed by using 2 μg of synthesized RNA preheated for 10 min at 65°C and 35 μl of nuclease-treated rabbit reticulocyte lysate (Promega) in a 50-μl reaction mixture. Translation reactions were carried out for 60 min at 30°C.

For in vivo expression, the 1.5-kb *spi-B* cDNA insert was subcloned into a simian virus 40 (SV40) early promoter-based expression plasmid ΔEB (6) and transfected into COS1 cells, which were maintained in Dulbecco's modified Eagle medium containing 5% fetal calf serum, using the DEAE-dextran procedure, as previously described (5). The transfected COS1 cells were lysed 48 h after transfection in 4 volumes of 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.9)–0.3 M NaCl–1 mM dithiothreitol–0.1 mM EDTA–0.5% Triton X-100–1% aprotinin–paramethylsulfonyl fluoride (100 μg/ml)–leupeptin (10 μg/ml); after centrifugation at 10,000 × *g* for 20 min, the supernatant was collected.

Gel retardation assay. Proteins expressed in vitro (3 μl of translation mix) or in vivo (1 μl of COS1 lysate) were preincubated 10 min on ice with 1.5 μg of double-stranded poly(dI-dC) (Pharmacia) and 0.4 μg of salmon sperm DNA in a buffer containing 10 mM HEPES (pH 7.4), 25 mM KCl, 1.25 mM sodium phosphate, 0.175 mM EDTA, 0.075 mM EGTA, 1 mM dithiothreitol, and 5 mM $MgCl_2$ in a final volume of 16 μl. The double-stranded oligonucleotide probe was ^{32}P labeled by filling in recessed ends with the Klenow fragment of *Escherichia coli* DNA polymerase I; 50 fmol of probe was added to each sample, and the samples were incubated for 10 min at 25°C. For competition experiments, a 200-fold molar excess of unlabeled oligonucleotide was mixed with the labeled oligonucleotide. Following incubation periods, binding reactions were immediately loaded on 5% polyacrylamide gels (29:1, acrylamide/bisacrylamide) and run in 0.25× TBE (1× TBE is 0.089 M Tris, 0.089 boric acid, and 0.0025 M EDTA). The sequences of one strand of the double-stranded oligonucleotides were TCGGGCTC GAGTCTGAAAGAGGAACTTGGTTAGC (SV40 PU) and TCGGGCTCGAGTCTTGAAAGACCAACTTGGTTAGC (SV40 PUmut).

Transfections and CAT assays. The reporter plasmid used contained four copies of an oligonucleotide encompassing the SV40 PU box inserted upstream of the herpes simplex virus thymidine kinase gene promoter of the reporter plasmid pBLCAT2 (18) (a generous gift of R. Maki). HeLa cells were grown in Dulbecco's modified Eagle medium containing 10% fetal calf serum and were transfected by the calcium phosphate coprecipitation method with expression vector (2.5 μg per 60-mm-diameter petri dish) and reporter plasmid (0.5 μg per 60-mm-diameter petri dish) as previously described (4). After 48 h, the cells were harvested and lysed, and chloramphenicol acetyltransferase (CAT) activities were determined as previously described (4). Each transfection was repeated three times.

RNA preparation and Northern blot analysis. RNAs were isolated from human and murine cell lines and from murine tissues by the guanidine thiocyanate homogenization procedure followed by ultracentrifugation through a cesium chloride cushion as previously described (9). Poly(A)$^+$ RNAs were selected by one passage over an oligo(dT)-cellulose column. The RNAs [5 μg of poly(A)$^+$ RNA or 30 μg of total RNA] were separated by electrophoresis on 1% agarose gels containing formaldehyde and blotted onto GeneScreen membranes (New England Nuclear). The *spi-B* cDNA, *spi-1* cDNA, and glyceraldehyde 3-phosphate dehydrogenase probes were ^{32}P labeled by random priming (Amersham multiprime labeling system). Hybridizations were carried out at 42°C in 50% formamide–0.05 M Tris-HCl (pH 7.5)–1 M NaCl–0.1% sodium pyrophosphate–0.1% SDS, 5× Denhardt's solution–10% dextran sulfate–denatured salmon sperm DNA (200 μg/ml). Membranes were finally washed in 0.1× SSC–0.1% SDS at 55 to 60°C.

Human gene mapping by in situ hybridization. In situ hybridization was carried out on chromosome preparations obtained from phytohemagglutinin-stimulated human lymphocytes cultured for 72 h. 5-Bromodeoxyuridine was added for the final 7 h of culture (60 μg/ml of medium) to ensure a posthybridization chromosomal banding of good quality. The *spi-B* cDNA cloned in pBluescript was tritium labeled by nick translation at a specific activity of 3.4×10^7 dpm/μg. The radiolabeled probe was hybridized to metaphase spreads at final concentration of 25 ng/ml of hybridization solution as previously described (21). After coating with nuclear track emulsion (Kodak NTB_2), the slides were exposed for 16 days at 4°C and then developed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads were first stained with buffered Giemsa solution and metaphases were photographed. R banding was then performed by the fluorochrome-photolysis-Giemsa method, and metaphases were rephotographed before analysis.

Nucleotide sequence accession number. The nucleotide sequence of the human *spi-B* cDNA reported in this paper has been deposited in the EMBL data base under accession number X66079.

RESULTS AND DISCUSSION

Characterization of *spi-B* cDNA. A λgt10 cDNA library was constructed from the Raji Burkitt's human lymphoma cell line and screened, under low-stringency conditions, with the human *spi-1* cDNA probe *Sma*I-*Nar*I (330 bp) (39), which corresponds to the DNA-binding domain of Spi-1/PU.1. Among 10^5 phage clones screened, one positive clone, which failed to hybridize under high-stringency conditions, was selected. The nucleotide sequence of this 1.1-kb-long

cDNA clone was determined. The deduced open reading frame encodes a 151-amino-acid polypeptide with no methionine at its NH_2 terminus, indicating that this cDNA was incomplete. To search for full-length cDNA clones, 10^5 recombinant phages from the same library were screened at high hybridization stringency with the 1.1-kb cDNA insert as a probe. We obtained 11 clones, and the nucleotide sequences of two clones containing 1.5-kb cDNA inserts were determined. Northern blot analysis of poly(A)$^+$ RNAs from Raji Burkitt's lymphoma cells, using the 1.5-kb cDNA insert as a probe, showed that these 1.5-kb cDNAs corresponded in size to the predominant mRNA observed in these cells (see Fig. 6A). Two minor 2.8- and 3.7-kb transcripts were also detected in Raji Burkitt's lymphoma cells. The nucleotide sequences of both 1.5-kb cDNA inserts revealed a 786-nucleotide-long open reading frame encoding a 262-amino-acid protein (Fig. 1). The ATG initiator codon partially conforms to the consensus sequence found in eukaryotic translation initiation sites with an adenine in position −3 (19). However, since the cDNA insert extends only 5 nucleotides upstream from the initiator codon in the nucleotide sequences of both 1.5-kb *spi-B* cDNA clones, we cannot exclude the existence of an amino-terminal portion missing from our clones. To resolve this question, we isolated the genomic region corresponding to the 5′ region of the 1.5-kb cDNA by screening a human genomic library with an *Eco*RI-*Pst*I probe containing the first 263 nucleotides of the cDNA insert. The nucleotide sequence of this genomic region revealed the presence of an in-frame stop codon 48 nucleotides upstream from the first ATG codon, suggesting that the open reading frame is entire. The two 1.5-kb cDNA clones contain a poly(A) tract starting at position 1449 in Fig. 1. However, no classical polyadenylation site could be identified within the untranslated 3′ end. The 262-amino-acid sequence, with a 29-kDa predicted molecular mass, shows 43% overall identity with the human Spi-1 amino acid sequence. Because of the identity between the amino acid sequence of the 1.5-kb cDNA clone and the Spi-1/PU.1 protein (see below), this new gene was named *spi-B*.

Amino acid sequence comparison of Spi-B with Spi-1 and other members of the Ets family. Comparison of the deduced amino acid sequence of the human Spi-B protein with sequences of the human and murine Spi-1 proteins shows 43 and 42% identity, respectively (Fig. 2). The highest similarity (67% identity) is found in a 98-amino-acid region (residues 160 to 257 [Spi-B numbering]) located in the carboxy-terminal part of the proteins. The N-terminal amino acid region (1 to 159) of Spi-B is more distantly related, with only 29% identity with the human Spi-1 protein. Analysis of the Spi-B amino acid composition reveal that there is 0.6% arginine plus lysine versus 16% aspartic plus glutamic acids in the 159 N-terminal amino acids and 24% arginine plus lysine, versus 6% aspartic plus glutamic acids in the 103 C-terminal amino acids. Therefore, two domains can be defined in Spi-B: a highly basic carboxy-terminal domain and an acidic amino-terminal domain. These features are also found in the Spi-1 protein. The basic carboxy-terminal domain of Spi-1/PU.1 is sufficient for specific DNA binding in vitro, whereas the acidic amino-terminal domain has been proposed to be implicated in transcriptional activation (18).

The basic 98-amino-acid sequence (160 to 257) of Spi-B, highly conserved (67%) with Spi-1, exhibits much less similarity (31 to 34% identity) with the Ets domain of the other members of the Ets protein family (Fig. 3). The degree of sequence conservation between the Ets domain of Spi-B and those of the other Ets family members is slightly reduced

```
                                  aatagggttggcggctgcagcgggcggcaaacagcccgcccggca      -1

ccacc ATG CTC GCC CTG GAG GCT GCA CAG CTC GAC GGG CCA CAC TTC      47
       M   L   A   L   E   A   A   Q   L   D   G   P   H   F      14

AGC TGT CTG TAC CCA GAT GGC GTC TTC TAT GAC CTG GAC AGC TGC AAG     95
 S   C   L   Y   P   D   G   V   F   Y   D   L   D   S   C   K     30

CAT TCC AGC TAC CCT GAT TCA GAG GGG GCT CCT GAC TCC CTG TGG GAC    143
 H   S   S   Y   P   D   S   E   G   A   P   D   S   L   W   D     46

TGG ACT GTG GCC CCA CCT GTC CCA GCC ACC CCC TAT GAA GCC TTC GAC    191
 W   T   V   A   P   P   V   P   A   T   P   Y   E   A   F   D     62

CCG GCA GCA GCC GCT TTT AGC CAC CCC CAG GCT GCC CAG CTC TGC TAC    239
 P   A   A   A   A   F   S   H   P   Q   A   A   Q   L   C   Y     78
                                 PstI
GAA CCC CCC ACC TAC AGC CCT GCA GGG AAC CTC GAA CTG GCC CCC AGC    287
 E   P   P   T   Y   S   P   A   G   N   L   E   L   A   P   S     94
             SmaI
CTG GAG GCC CCG GGG CCT GGC CTC CCC GCA TAC CCC ACG GAG AAC TTC    335
 L   E   A   P   G   P   G   L   P   A   Y   P   T   E   N   F    110

GCT AGC CAG ACC CTG GTT CCC CCG GCA TAT GCC CCG TAC CCC AGC CCT    383
 A   S   Q   T   L   V   P   P   A   Y   A   P   Y   P   S   P    126

GTG CTA TCA GAG GAG GAA GAC TTA CCG TTG GAC AGC CCT GCC CTG GAG    431
 V   L   S   E   E   E   D   L   P   L   D   S   P   A   L   E    142

GTC TCG GAC AGC GAG TCG GAT GAG GCC CTC GTG GCT GGC CCC GAG GGG    479
 V   S   D   S   E   S   D   E   A   L   V   A   G   P   E   G    158
     BamHI
AAG GGA TCC GAG GCA GGG ACT CGC AAG AAG CTG CGC CTG TAC CAG TTC    527
 K   G   S   E   A   G   T   R   K   K   L   R   L   Y   Q   F    174

CTG CTG GGG CTA CTG ACG CGC GGG GAC ATG CGT GAG TGC GTG TGG TGG    575
 L   L   G   L   L   T   R   G   D   M   R   E   C   V   W   W    190

GTG GAG CCA GGC GCC GGC GTC TTC CAG TTC TCC TCC AAG CAC AAG GAA    623
 V   E   P   G   A   G   V   F   Q   F   S   S   K   H   K   E    206

CTC CTG GCG CGC CGC TGG GGC CAG CAG AAG GGG AAC CGC AAG CGC ATG    671
 L   L   A   R   R   W   G   Q   Q   K   G   N   R   K   R   M    222

ACC TAC CAG AAG CTG GCG CGC GCC CTC CGA AAC TAC GCC AAG ACC GGC    719
 T   Y   Q   K   L   A   R   A   L   R   N   Y   A   K   T   G    238

GAG ATC CGC AAG GTC AAG CGC AAG CTC ACC TAC CAG TTC GAC AGC GCG    767
 E   I   R   K   V   K   R   K   L   T   Y   Q   F   D   S   A    254
         PstI
CTG CTG CCT GCA GTC CGC CGG GCC  tgagcacacccgaggctcccacctgcggagc    822
 L   L   P   A   V   R   R   A   *                                 262

cgctgggggacctcacgtcccagccaggatccccctggaagaaaaagggcgtccccacactcta  886

ggtgataggacttacgcatccccacctttggggtaaggggagtgctgccctgccataatcccc  950
          SmaI
aagcccagcccgggcctgtctgggattccccacttgtgcctggggtccctctgggatttctttg 1014
                BamHI
tcatgtacagactccctgggatcctcatgttttgggtgacaggacctatggaccactatactcg 1078
                                                          EcoRS
gggaggcagggtagcagtgcttccagagtcccaagagcttctctgggattttcttgtgatatct 1142

gattccccagtgaggcctgggacctttttaagatcgctgtgtgtctgtaaaccctgaatctcat 1206

ctggggtgggggccctgctggcaaccctgagccctgtccaaggttccctcttgtcagatctgag 1270

atttcctagttatgtctggggccctctgggagctgttatcatctcagatctcttcgcccatcta 1334

tggctgtgttgtcacatctgtcccctcatttttgagatcccccaattctctggaactattctgc 1398

tgccccttttatgtgtctggagttccccaatcacatctagggctcctcc                1448
```

FIG. 1. Nucleotide sequence of the human *spi-B* cDNA and the 5′ flanking genomic region. The deduced amino acid sequence of the predicted open reading frame is shown under the nucleotide sequence. The genomic region 5′ upstream from the cDNA sequence is numbered negatively (−45 to −1). The in-frame stop codon preceding the open reading frame is underlined twice. The noncoding sequence is given in small letters.

compared with that of Spi-1 (35 to 41%). As no known consensus DNA-binding motif could be identified in the Ets domain, it has been proposed that this region be considered a new structural motif (17). However, the Ets domain contains three conserved tryptophan residues separated by 17, 18, or 19 amino acids. This motif resembles the structural motif identified in the DNA-binding domain of the Myb protein, in which a three-fold imperfect 52-amino-acid repeat includes three tryptophan residues separated from each other by 18 or 19 amino acids (1). In both Spi-1 and Spi-B, the first tryptophan of this putative motif is changed to a tyrosine. Such a conservative exchange between two aromatic amino acids in Myb has been shown not to affect its DNA-binding ability (40). Moreover, six basic amino acid positions (arginine or lysine) are conserved between Spi-B

FIG. 2. Alignment of Spi-B and Spi-1/PU.1 amino acid sequences. The complete amino acid sequences for human and murine Spi-1/PU.1 are compared with that of human Spi-B. Dots in the Spi-1/PU.1 sequences denote amino acid identity with Spi-B, and dashes allow optimal alignment. The Ets homology domain is underlined by dashes. Asterisks indicate the basic amino acids (arginine [R] or lysine [K]) that are conserved among Spi-B, Spi-1/PU.1, and the Ets members.

(216 to 246), Spi-1/PU.1, and all of the Ets family members. These basic residues might be involved in DNA-protein interactions, as demonstrated for clusters of basic amino acids present in the basic region/leucine zipper, basic region/helix-loop-helix, or zinc finger DNA-binding domains (13, 44).

The acidic 159 N-terminal amino acids of Spi-B contain 16% proline, versus 2% in the basic 103 C-terminal amino

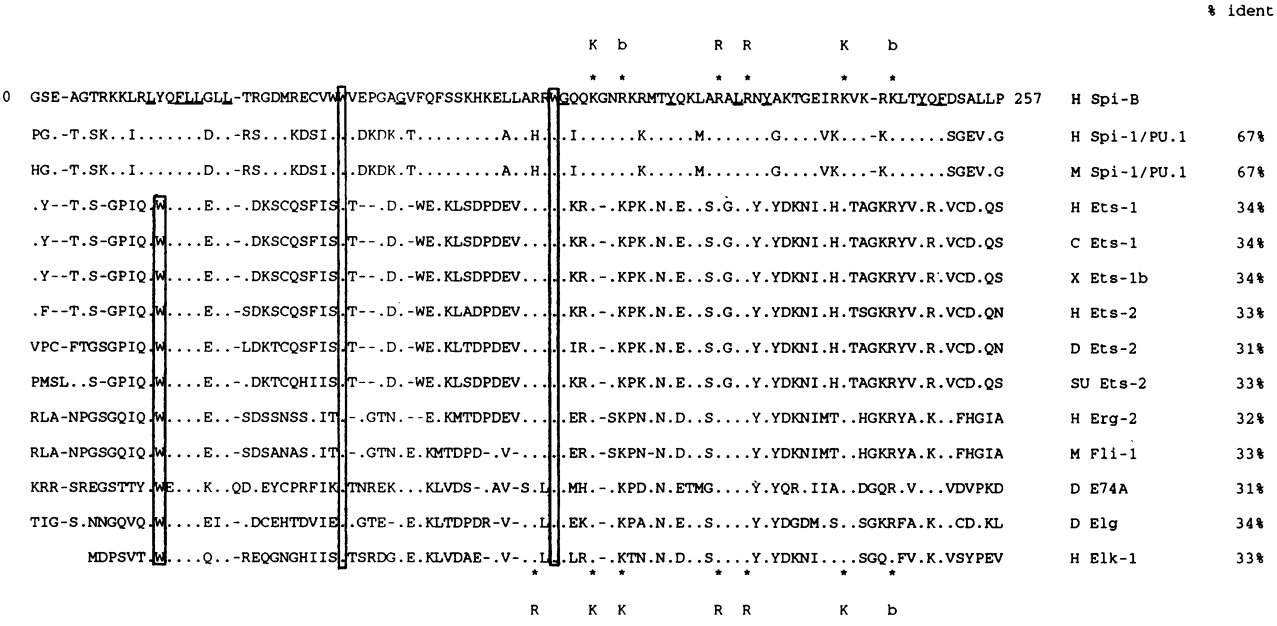

FIG. 3. Alignment of amino acid sequences of Spi-B, Spi-1/PU.1, and Ets members. The Ets domain amino acid sequences for Spi-1/PU.1 and members of the Ets family are compared with that of Spi-B. Ets sequences displayed are human (H) Ets-1 and Ets-2 (48), chicken (C) Ets-1 (49), xenopus (X) Ets-1b (43), drosophila (D) Ets-2 (32), sea urchin (SU) Ets-2 (8), human Erg2 (37), murine (M) Fli-1 (2), drosophila E74A (7), drosophila Elg (33), and human Elk-1 (36). Dots indicate identical amino acids, and dashes allow optimal alignment. The upper and lower asterisks indicate the basic amino acids that are conserved among Spi-B, Spi-1/PU.1, and the Ets members and among the Ets family members, respectively. b, basic amino acid (arginine [R] or lysine [K]). Conserved tryptophan residues are boxed. The amino acids that are underlined in the Spi-B sequence correspond to residues conserved among Spi-B, Spi-1/PU.1, and Ets proteins.

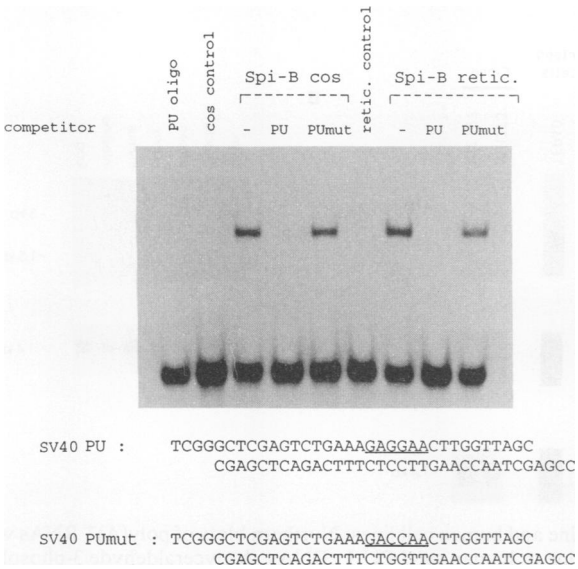

FIG. 4. Sequence-specific DNA-binding activity of Spi-B. Gel retardation assays were performed by using in vitro (Spi-B retic.) or in vivo (Spi-B cos) Spi-B products and ^{32}P-labeled SV40 PU oligonucleotide as a probe. For competition, unlabeled SV40 PU or SV40 PUmut oligonucleotide competitor was added at a 200-fold molar excess. As controls, unprogrammed reticulocyte lysate products (retic. control) and lysates of COS1 cells transfected with the control vector (cos control) were used. The underlined bases in SV40 PU and SV40 PUmut double-stranded oligonucleotide sequences indicate the core PU box and the mutated core PU box, respectively.

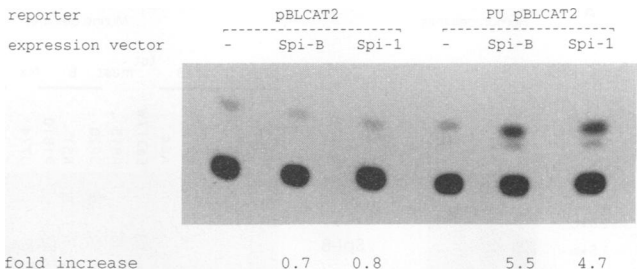

FIG. 5. Transactivation by Spi-B. HeLa cells were cotransfected with a Spi-B, Spi-1/PU.1, or control (indicated by a dash) expression vector and with PU pBLCAT2 (containing four PU boxes) or control pBLCAT2 reporter plasmid. CAT assays were performed on cell lysates 48 h after transfection. The chromatogram represents a typical transfection. Fold increase indicates the level of CAT activation due to the presence of either the Spi-B or Spi-1 expression vector.

acids. Although Spi-B and Spi-1 present only 29% sequence identity in their N-terminal halves, they are both characterized by a high frequency of proline and acidic residues, a feature frequently reported in activating regions of eukaryotic transcriptional activators (24, 34).

DNA-binding activity of Spi-B. To determine whether Spi-B is a DNA-binding protein, we studied its ability to bind the purine-rich DNA sequence (PU box) that is recognized by Spi-1. We performed gel retardation assays, using as a probe the PU oligonucleotide from the SV40 enhancer (SV40 PU), which contains a single PU box (5′-GAGGAA-3′) (18). The Spi-B protein used in these experiments was derived either from lysates of COS1 cells transfected with a Spi-B expression plasmid or from reticulocyte lysates programmed with *spi-B* RNAs transcribed in vitro from the 1.5-kb *spi-B* cDNA. Figure 4 illustrates a gel retardation assay in which the radiolabeled SV40 PU oligonucleotide was incubated with the Spi-B protein produced in vivo or in vitro. The same retarded complex was observed between the SV40 PU oligonucleotide and Spi-B, irrespective of its source. To confirm the specificity of this protein-DNA complex, competition experiments were performed with use of a 200-fold molar excess of either nonlabeled SV40 PU oligonucleotide or a nonlabeled mutated version of this oligonucleotide (SV40 PUmut) in which the central two guanines of the PU box were changed to cytosines. It was previously determined, by methylation interference analysis, that the two guanines of the central core 5′-GGAA-3′ are contact points of Spi-1/PU.1 (18). Competition was observed with the unlabeled oligonucleotide SV40 PU, whereas the unlabeled mutated oligonucleotide SV40 PUmut failed to compete. These results show that Spi-B can bind specifically an oligonucleotide which contains the purine-rich sequence 5′-GAGGAA-3′ and that the central two guanines of the PU box are indispensable for binding of the protein. The recognition core sequence 5′-GGAA/T-3′ is present in the Ets-binding sites described so far for Ets-1 (10, 50), Erg (42), Elf-1 (45), E74 (46), and PEA3 (51) and appears to be essential for Ets proteins, as for Spi-B and Spi-1, to bind DNA. Further experiments of gel retardation assay using oligonucleotides with various residues around the central core sequence 5′-GGAA-3′ might enable determination of the DNA-binding specificities of both Spi-1 and Spi-B proteins.

Transcriptional activation by Spi-B. To investigate the possible function of Spi-B as a transcriptional regulator, we cotransfected HeLa cells with a Spi-B expression plasmid and the pBLCAT2 reporter plasmid in which four copies of a SV40 PU oligonucleotide was inserted immediately upstream of the thymidine kinase promoter (pBLCAT2 PU). As controls, we used the expression plasmid without a cDNA insert and the pBLCAT2 reporter plasmid. Spi-B was found to stimulate the CAT activity of the PU pBLCAT2 reporter plasmid four- to fivefold compared with the level of activity in the absence of Spi-B (Fig. 5). When the Spi-1 expression plasmid was cotransfected with the PU pBLCAT2 reporter plasmid, the CAT activity was increased three- to fourfold, as previously described (18). We conclude from these experiments that Spi-B is able to stimulate transcription of a CAT reporter plasmid harboring PU boxes, demonstrating that *spi-B* encodes a sequence-specific transcriptional activator. Although the naturally occurring DNA-binding site of Spi-B in vivo remains to be identified, our results show that Spi-B and Spi-1/PU.1 are able to bind a common element and to transactivate a reporter plasmid through this element. However, sequences flanking the PU box in vivo can modulate the transcriptional regulator activities of Spi-B and Spi-1 either by changing their affinities for their binding sites or by interacting with factors cooperating positively or negatively. Indeed, it has been recently reported that the protein-protein interaction of Spi-1/PU.1 with the factor NF-EM5 within the immunoglobulin κ 3′ enhancer is dependent on the binding of PU.1 to the PU box and NF-EM5 to an adjacent site (31).

Expression pattern of the *spi-B* gene. *spi-B* gene expression was assessed by Northern blot analysis of poly(A)$^+$ RNAs from various murine and human cell lines. As shown in Fig. 6A, three *spi-B* transcripts (1.5, 2.8, and 3.7 kb) were

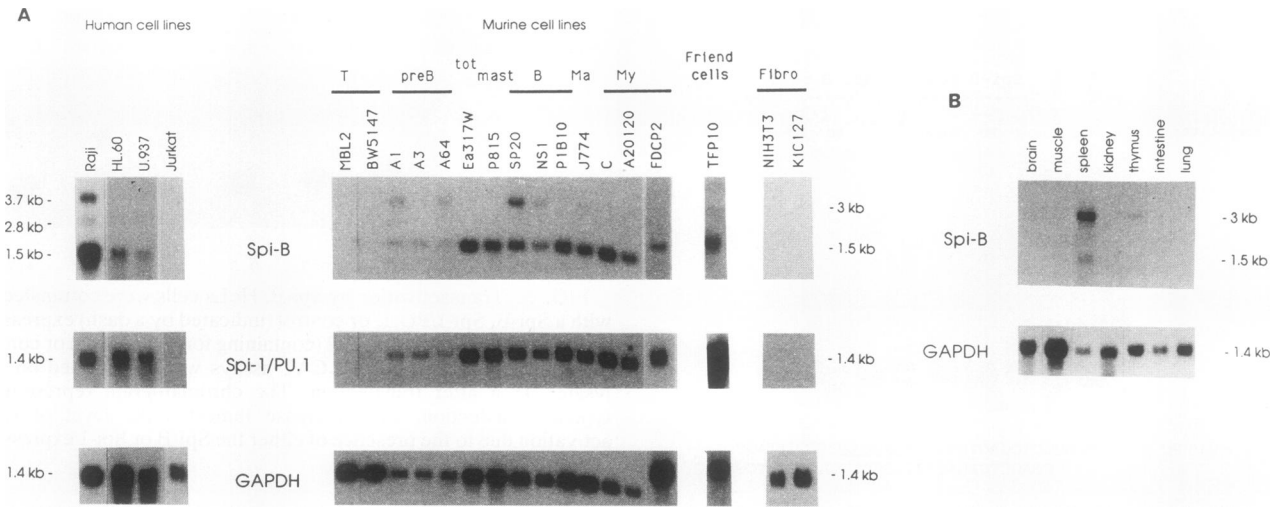

FIG. 6. (A) Northern blot analysis of *spi-B* and *spi-1*/PU.1 mRNAs in murine and human cell lines. Northern blots of poly(A)$^+$ RNAs were probed with the coding regions of the human 1.5-kb *spi-B* cDNA and the murine or human *spi-1*/PU.1 cDNAs. A glyceraldehyde 3-phosphate dehydrogenase (GAPDH) probe was used as a control for the amount of RNA applied. T, lymphoma cell line; preB, pre-B lymphoma cell line; tot, totipotent cell line; mast, mastocytoma cell line; B, myeloma cell line; My, myeloid leukemia cell line; Fibro, fibroblastic cell line (B) Northern blot analysis of *spi-B* mRNAs in murine tissues. Northern blots of total RNAs were probed with the coding region of the 1.5-kb *spi-B* cDNA.

observed in human cells and two (1.5 and 3 kb) were found in murine cells. The 1.5-kb mRNA appeared largely predominant in most cell lines examined. In murine cells, *spi-B* mRNAs were found in pre-B lymphoma cell lines (A1, A3, and A64), in the totipotent cell line Ea317W, in myeloma cell lines (Sp2o, NS1, and P1B10), in the mastocytoma cell line P815, in the macrophagelike cell line J774, in myeloid leukemia cell lines (C and A20120), and in the erythroid Friend cell line (TFP10) but were detected neither in T lymphoma cell lines (MBL2 and BW157) nor in fibroblastic cell lines (NIH 3T3 and KiC 127). In human cells, *spi-B* mRNAs were observed in the Raji Burkitt lymphoma cell line, in the promyelocytic leukemia cell line HL60, and in the monocytic cell line U937. No *spi-B* expression was detected in the Jurkat T lymphoma cell line. Hybridization of the same Northern blots with a *spi-1* cDNA probe revealed that *spi-1* and *spi-B* expression patterns are similar. Therefore, both *spi-B* and *spi-1* were found to be expressed in all of the hematopoietic cell lines examined except in the T lymphoid cell lines, suggesting that these genes would not be involved in the T-cell proliferation/differentiation process. However, *spi-B* expression patterns varied qualitatively and quantitatively according to the cell line analyzed. While the 1.5-kb transcript was generally present, the murine 3-kb transcript was not consistently observed in a cell line type. Moreover, we detected the 3.7- and 2.8-kb human transcripts only in the Raji Burkitt lymphoma cell line. In total RNAs from mouse adult tissues, *spi-B* transcripts were found at a high level in spleen and at a very low level in thymus (Fig. 6B). No *spi-B* expression was detected in all other tissues examined (brain, muscle, kidney, lung, and intestine). The 3-kb transcript was the major species expressed in spleen and thymus.

Our results indicate that expression of *spi-1*/PU.1 is more general than previously anticipated (18) and, in particular, is detected in most hematopoietic lineages except in the T-cell lines examined. In murine adult tissues, *spi-B* expression is found predominantly in spleen and is faintly detectable in thymus. The apparent discrepancy between the detection of *spi-B* expression in thymus and its lack of expression in T-cell lines could be accounted for by the restriction of its expression either to a minor T-cell subpopulation in the thymus or to non-T cells present in the thymus. Until now we have had no data to clarify this question. *spi-B* gene is transcribed in various murine and human hematopoietic cell lines either as a unique 1.5-kb transcript or as multiple mRNAs. The 1.5-kb transcript is predominant in all in vitro-tested cell lines, whereas it appears as the minor species in vivo in murine splenic and thymus tissues. For the longer mRNAs (3 kb in murine cells and 2.8 and 3.7 kb in human cells), no correlation can be established between their presence, their distribution, the hematopoietic cell phenotype, and the degree of cell maturation (B and pre-B cells). These transcripts, heterogeneous in size among human and murine species, could be generated by alternative splicing or by differential initiation or termination of transcription. Preliminary analysis of a longer *spi-B* cDNA clone suggests that the larger mRNAs may differ from the 1.5-kb mRNA by a longer 3′ untranslated region (data not shown). As the putative function of the 3′ untranslated sequences in mRNAs could be a modulation of their translation efficiency and their turnover rate in vivo (15), the variation in the relative abundance of the different *spi-B* transcripts might reflect a potential regulation of the *spi-B* translation in the cell lines and in the tissues. We are currently analyzing cDNA clones corresponding to the larger *spi-B* mRNAs to determine whether they derive from different transcription units of *spi-B* or whether they result from alternative processing of a common precursor RNA.

Chromosomal localization of the human *spi-B* gene. In the 200 metaphase cells examined after in situ hybridization, there were 437 silver grains associated with chromosomes, and 58 of these (13.6%) were located on chromosome 19; the distribution of grains on this chromosome was not random; 75.8% (44 of 58) of them mapped to the q13.3-q13.4 region of chromosome 19 long arm, with a maximum in the 19q13.4 band. These results allow us to map *spi-B* to the 19q13.3-19q13.4 region of the human genome (Fig. 7). This location is outside the translocation breakpoints that have been identi-

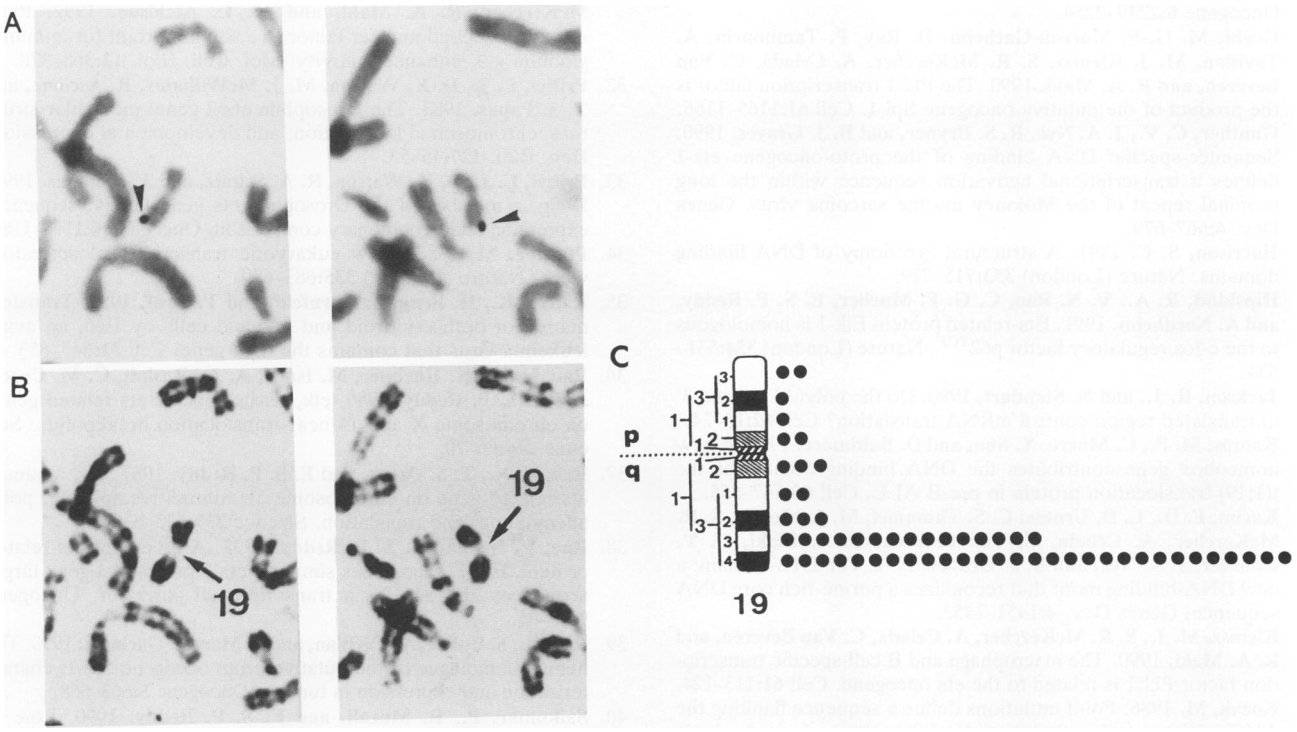

FIG. 7. (A and B) Two partial human metaphases showing the specific site of hybridization to chromosome 19. (A) Arrowheads indicate silver grains on Giemsa-stained chromosomes after autoradiography. (B) Chromosomes with silver grains were subsequently identified by R banding (fluorochrome-photolysis-Giemsa technique). (C) Idiogram of the human G-banded chromosome 19 illustrating the distribution of labeled sites for the *spi-B* cDNA probe.

fied on the chromosome 19 in pre-B-cell acute lymphoblastic leukemias [t(1;19) (q23;p13.3)] (16, 28), T-cell acute lymphoblastic leukemias [t(7;19) (q35;p13)] (22), and B-cell chronic lymphocytic leukemias [t(14;19) (q32;q13.1)] (30). Therefore, a possible involvement of *spi-B* in a human neoplastic process cannot be deduced from its chromosomal position.

In conclusion, we have characterized *spi-B*, a gene related to the putative oncogene *spi-1*/PU.1. On the basis of a high homology in their Ets domains, Spi-B and Spi-1 constitute an Ets subfamily that is actually the most divergent from Ets-1. Both are transcriptional activators which exhibit similar sequence-specific DNA binding and display a comparable expression in non-T hematopoietic cell lineages. However, the low sequence identity in their N-terminal regions, which are proposed to be implicated in transcriptional activation, suggests that Spi-B and Spi-1 may have different functional specificities in hematopoiesis.

ACKNOWLEDGMENTS

We thank Pascale Meneceur for technical assistance, Sylvie Gisselbrecht for providing the Northern blot of RNAs from murine hematopoietic cell lines, Richard A. Maki for the gift of the pBLCAT2 reporter plasmids, and Claudine Grégory for providing the human genomic library. We also thank Pierre Tambourin for helpful discussions and Bernard Boursin for photographic artwork.

This work was supported by the Institut National de la Santé et de la Recherche Médicale and a grant from the Association pour la Recherche sur le Cancer.

REFERENCES

1. **Anton, I. A., and J. Frampton.** 1988. Tryptophans in myb proteins. Nature (London) **336**:719.
2. **Ben-David, Y., E. B. Giddens, K. Letwin, and A. Bernstein.** 1991. Erythroleukemia induction by Friend murine leukemia virus: insertional activation of a new member of the ets gene family, Fli-1, closely linked to c-ets-1. Genes Dev. **5**:908–918.
3. **Bister, K., M. Nunn, C. Moscovici, B. Perbal, M. A. Baluda, and P. H. Duesberg.** 1982. Acute leukemia viruses E26 and avian myeloblastosis virus have related transformation-specific RNA sequences but different genetic structures, gene products, and oncogenic properties. Proc. Natl. Acad. Sci. USA **79**:3677–3681.
4. **Bosselut, R., J. F. Duvall, A. Gégonne, M. Bailly, A. Hémar, J. Brady, and J. Ghysdael.** 1990. The product of the c-ets-1 proto-oncogene and the related Ets2 protein act as transcriptional activators of the long terminal repeat of human T cell leukemia virus HTLV-1. EMBO J. **9**:3137–3144.
5. **Boulukos, K. E., P. Pognonec, A. Begue, F. Galibert, J. C. Gesquiere, D. Stéhelin, and J. Ghysdael.** 1988. Identification in chickens of an evolutionary conserved cellular ets-2 gene (c-ets-2) encoding nuclear proteins related to the products of the c-ets proto-oncogene. EMBO J. **7**:697–705.
6. **Boulukos, K. E., P. Pognonec, B. Rabault, A. Begue, and J. Ghysdael.** 1989. Definition of an Ets1 protein domain required for nuclear localization in cells and DNA-binding activity in vitro. Mol. Cell. Biol. **9**:5718–5721.
7. **Burtis, K. C., C. S. Thummel, C. W. Jones, F. D. Karim, and D. S. Hogness.** 1990. The drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell **61**:85–99.
8. **Chen, Z. Q., N. C. Kan, L. J. Pribyl, J. A. Lautenberger, E. Moudrianakis, and T. S. Papas.** 1988. Molecular cloning of the ets proto-oncogene of the sea urchin and analysis of its developmental expression. Dev. Biol. **125**:432–440.
9. **Chirgwin, J., A. Przybyla, R. McDonald, and W. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
10. **Fischer, R. J., G. Mavrothalassitis, A. Kondoh, and T. S. Papas.** 1991. High-affinity DNA-protein interactions of the cellular Ets1 protein: the determination of the ETS binding motif.

Oncogene 6:2249–2254.
11. **Goebl, M. G.-F. Moreau-Gachelin, D. Ray, P. Tambourin, A. Tavitian, M. J. Klemsz, S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki.** 1990. The PU.1 transcription factor is the product of the putative oncogene Spi-1. Cell **61**:1165–1166.
12. **Gunther, C. V., J. A. Nye, R. S. Bryner, and B. J. Graves.** 1990. Sequence-specific DNA binding of the proto-oncogene ets-1 defines a transcriptional activation sequence within the long terminal repeat of the Moloney murine sarcoma virus. Genes Dev. **4**:667–679.
13. **Harrison, S. C.** 1991. A structural taxonomy of DNA binding domains. Nature (London) **353**:715–719.
14. **Hipskind, R. A., V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim.** 1991. Ets-related protein Elk-1 is homologous to the c-fos regulatory factor $p62^{TCF}$. Nature (London) **354**:531–534.
15. **Jackson, R. J., and N. Standart.** 1990. Do the poly(A) tail and 3′ untranslated region control mRNA translation? Cell **62**:15–24.
16. **Kamps, M. P., C. Murre, X. 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.
17. **Karim, F. D., L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKercher, A. Celada, C. Van Beveren, R. A. Maki, C. V. Gunther, J. A. Nye, and B. J. Graves.** 1991. The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev. **4**:1451–1453.
18. **Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki.** 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. Cell **61**:113–124.
19. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
20. **Leprince, D., A. Gégonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagrou, and D. Stéhelin.** 1983. A putative second cell-derived oncogene of the avian leukemia retrovirus E26. Nature (London) **306**:395–397.
21. **Mattei, M. G., N. Philip, E. Passage, J. P. Moisan, J. L. Mandel, and J. F. Mattei.** 1985. DNA probe localization at 18p11.3 band in situ hybridization and identification of a small supernumerary chromosome. Hum. Genet. **69**:268–271.
22. **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.
23. **Metz, T., and T. Graf.** 1991. Fusion of the nuclear oncoproteins v-Myb and v-Ets is required for the leukemogenicity of E26 virus. Cell **66**:95–105.
24. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
25. **Moreau-Gachelin, F., D. Ray, N. J. de Both, M. J. M. van der Feltz, P. Tambourin, and A. Tavitian.** 1989. Spi-1 oncogene activation in Rauscher and Friend murine virus-induced acute erythroleukemias. Leukemia **4**:20–23.
26. **Moreau-Gachelin, F., D. Ray, M.-G. Mattei, P. Tambourin, and A. Tavitian.** 1989. The putative oncogene Spi-1: murine chromosomal localization and transcriptional activation in murine acute erythroleukemias. Oncogene **4**:1449–1456.
27. **Moreau-Gachelin, F., A. Tavitian, and P. Tambourin.** 1988. Spi-1 is a putative oncogene in virally induced murine erythroleukemias. Nature (London) **331**:277–280.
28. **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.
29. **Nunn, M. F., P. H. Seeburg, C. Moscovici, and P. H. Duesberg.** 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature (London) **306**:391–395.
30. **Ohno, H., G. Takimoto, and T. W. McKeithan.** 1990. The candidate proto-oncogene bcl-3 is related to genes implicated in cell lineage determination and cell cycle control. Cell **60**:991–997.
31. **Pongubala, J. M. R., S. Nagulapalli, M. J. Klemsz, S. R. McKercher, R. A. Maki, and M. L. Atchison.** 1992. PU.1 recruits a second nuclear factor to a site important for immunoglobulin κ 3′ enhancer activity. Mol. Cell. Biol. **12**:368–378.
32. **Pribyl, L. J., D. K. Watson, M. J. McWilliams, R. Ascione, and T. S. Papas.** 1988. The Drosophila ets-2 gene: molecular structure, chromosomal localization, and developmental expression. Dev. Biol. **127**:45–53.
33. **Pribyl, L. J., D. K. Watson, R. A. Schulz, and T. S. Papas.** 1991. D-elg, a member of the Drosophila ets gene family: sequence, expression and evolutionary comparison. Oncogene **6**:1175–1183.
34. **Ptashne, M.** 1988. How eukaryotic transcriptional activators work. Nature (London) **335**:683–689.
35. **Radke, K., H. Beug, S. Kornfeld, and T. Graf.** 1982. Transformation of both erythroid and myeloid cells by E26, an avian leukemia virus that contains the myb gene. Cell **31**:643–653.
36. **Rao, V. N., K. Huebner, M. Isobe, A. Ar-Rushdi, C. M. Croce, and E. S. P. Reddy.** 1989. elk, tissue-specific ets-related genes on chromosome X and 14 near translocation breakpoints. Science **244**:66–70.
37. **Rao, V. N., T. S. Papas, and E. S. P. Reddy.** 1987. erg, a human ets-related gene on chromosome 21: alternative splicing, polyadenylation, and translation. Science **237**:635–639.
38. **Rao, V. N., and E. S. P. Reddy.** 1992. A divergent ets-related protein, Elk-1, recognizes similar c-ets-1 proto-oncogene target sequences and acts as a transcriptional activator. Oncogene **7**:60–65.
39. **Ray, D., S. Culine, A. Tavitian, and F. Moreau-Gachelin.** 1990. The human homologue of the putative proto-oncogene Spi-1: characterization and expression in tumors. Oncogene **5**:663–668.
40. **Saikumar, P., R. Murali, and E. S. P. Reddy.** 1990. Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. Proc. Natl. Acad. Sci. USA **87**:8452–8456.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
42. **Shyam, E., P. Reddy, and V. N. Rao.** 1991. erg, an ets-related gene, codes for sequence-specific transcriptional activators. Oncogene **6**:2285–2289.
43. **Stiegler, P., C. M. Wolff, J. Hirtzlin, F. Senan, M. Baltzinger, D. Meyer, J. Ghysdael, D. Stéhelin, N. Befort, and P. Remy.** 1990. Characterization of Xenopus laevis cDNA clones of the c-ets-1 proto-oncogene. Nucleic Acids Res. **18**:5298.
44. **Struhl, K.** 1989. Helix-turn-helix, zinc finger, and leucine zipper motifs for eukaryotic transcriptional regulatory proteins. Trends Biochem. Sci. **14**:137–140.
45. **Thompson, C. B., C. Y. Wang, I. C. Ho, P. R. Bohjanen, B. Petryniak, C. H. June, S. Miesfeldt, L. Zhang, G. J. Nabel, B. Karpinski, and J. M. Leiden.** 1992. cis-acting sequences required for inducible interleukin-2 enhancer function bind a novel ets-related protein, Elf-1. Mol. Cell. Biol. **12**:1043–1053.
46. **Urness, L. D., and C. S. Thummel.** 1990. Molecular interactions within the ecdysone regulatory hierarchy: DNA binding properties of the Drosophila ecdysone-inducible E74A protein. Cell **63**:47–61.
47. **Wasylyk, B., C. Wasylyk, P. Flores, A. Begue, D. Leprince, and D. Stéhelin.** 1990. The c-ets proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. Nature (London) **346**:191–193.
48. **Watson, D. K., M. J. McWilliams, P. Lapis, J. A. Lautenberger, C. W. Schweinfest, and T. S. Papas.** 1988. Mammalian ets-1 and ets-2 genes encode highly conserved proteins. Proc. Natl. Acad. Sci. USA **85**:7862–7866.
49. **Watson, D. K., M. J. McWilliams, and T. S. Papas.** 1988. Molecular organization of the chicken ets locus. Virology **164**: 99–105.
50. **Woods, D. B., J. Ghysdael, and M. J. Owen.** 1992. Identification of nucleotide preferences in DNA sequences recognized specifically by c-Ets-1 protein. Nucleic Acids Res. **4**:699–704.
51. **Xin, J.-H., A. Cowie, P. Lachance, and J. A. Hassel.** 1992. Molecular cloning and characterization of PEA3, a new member of the ets oncogene family that is differentially expressed in mouse embryonic cells. Genes Dev. **6**:481–496.