Jpn. J. Cancer Res. 84, 1136–1144, November 1993

Characterization of the Promoter Region of the Human c-*kit* Proto-oncogene

Katsuya Yamamoto,[1,3] Arinobu Tojo,[1,2] Nobuo Aoki[3] and Masabumi Shibuya[1,4]

Department of [1]Genetics and [2]Hematology/Oncology, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108 and [3]First Department of Internal Medicine, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113

The c-*kit* proto-oncogene encodes a tyrosine kinase receptor for stem cell factor and plays a critical role in the growth and differentiation of various types of cells including hematopoietic stem cells. To investigate the mechanisms of its transcriptional regulation, we isolated the 5′ flanking region of the human c-*kit* gene and characterized its promoter activity in hematopoietic cells. Nucleotide sequence analysis revealed that the 1.2 kb 5′ flanking region lacked a typical "TATA box," but had a relatively high G+C content and four potential Sp1-binding sites. Putative binding sites for AP-2, basic helix-loop-helix proteins, Ets-domain proteins, Myb and GATA-1 were also found. Primer extension and S1 nuclease protection analyses of hematopoietic cells indicated that the major transcription start sites are 62 bp and 58 bp upstream of the translation start site. Essentially the same start sites were detected in non-hematopoietic cells such as small cell lung carcinoma and glioblastoma: this single promoter in c-*kit* is different from the multiple promoter system of c-*fms*, a c-*kit*-related gene, in which at least two promoters are differently used in hematopoietic and non-hematopoietic cells. An analysis of the c-*kit* 5′ flanking region using the bacterial chloramphenicol acetyltransferase gene (CAT assay) in human erythroleukemia HEL cells, which express the endogenous c-*kit* mRNA at high levels, showed that a region from −180 to −22 is important for the expression of the c-*kit* gene. In addition, a negative regulatory element(s) is suggested to be involved in the regulation of the c-*kit* gene expression in mammals.

Key words: Human c-*kit* gene — Promoter — Transcription start site

The c-*kit* proto-oncogene encodes a transmembrane tyrosine kinase receptor which is structurally related to the colony stimulating factor-1 receptor (CSF-1R, c-Fms) and the platelet-derived growth factor receptor (PDGFR). It belongs to class III of the receptor tyrosine kinase family.[1] Receptor kinases of this class share an immunoglobulin-like configuration in the extracellular domain and contain a hydrophilic insertion sequence within the cytoplasmic tyrosine kinase domain.

The c-*kit* gene, located on the q11-q12 region of human chromosome 4,[2] has been mapped to the dominant white spotting *W* locus on mouse chromosome 5.[3,4] Mutations at the *W* locus are known to affect various aspects of melanogenesis, gametogenesis and hematopoiesis, resulting in white hair color, sterility, macrocytic anemia and mast cell deficiency.[5] Analysis of the c-Kit proteins from a variety of *W* mutant mice has revealed that their reduced tyrosine kinase activities are correlated with the severity of the phenotype.[6,7] The c-*kit* mRNA is widely expressed in various tissues and cells, especially in erythroid cells, mast cells, melanocytes and germ cells which are all known to be affected by *W* mutations.[8,9]

The ligand for the c-Kit receptor has recently been identified as the stem cell factor (SCF), mast cell growth factor (MGF), or kit ligand (KL) and its gene has been mapped to the mouse Steel (*Sl*) locus, a mutation of which is known to cause a similar phenotype in the *W* mutant mice.[10] This novel growth factor stimulates the proliferation of mast cells and enhances the growth of primitive hematopoietic progenitor cells synergistically along with other hematopoietic growth factors, indicating that these target cells express the c-Kit receptor.

Expression of the c-*kit* mRNA was examined in various human leukemic cells. The c-*kit* mRNA was positive in erythroleukemic and megakaryocytic cell lines, but not in myeloid and lymphoid cell lines.[11,12] The c-*kit* mRNA could also be detected in the blast cells of primary acute myeloblastic leukemia but not in cells of acute lymphoblastic leukemia or in normal bone marrow cells by Northern blot analysis.[13,14] Recently, by the use of monoclonal antibodies, the c-Kit receptor has been shown to be expressed in approximately 4% of normal bone marrow mononuclear cells, including erythroid, myeloid and multipotent progenitor cells, and it is broadly distributed in erythroid, myeloid and lymphoid cell lines.[15,16] These results suggest that the human c-Kit receptor and ligand play an important role in the regulation of both normal and leukemic hematopoietic stem cell growth. However, in spite of the accumulating knowledge on the expression and function of the c-Kit receptor, information about its transcriptional regulation is still limited.[17] To investigate the mechanisms of the

[4] To whom correspondence should be addressed.

transcriptional regulation of the c-*kit* gene in hematopoietic cells as well as in non-hematopoietic cells, we have isolated and characterized the 5′ flanking region of the human c-*kit* gene.

MATERIALS AND METHODS

Cell lines HEL (human erythroleukemia) and TALL-1 (human T cell acute lymphoblastic leukemia) cell lines were grown at 37°C and 5% CO_2 in Iscove's modified Dulbecco's medium (IMDM, GIBCO) supplemented with 10% newborn calf serum (GIBCO).[18, 19] Cells were grown to a maximum concentration of 5×10^5/ml. They showed an average viability of 90% or greater as measured by trypan blue staining. These cells were used for transfection assay and RNA extraction.

Five other human cell lines, KU812 (chronic myelogeneous leukemia basophilic crisis), MEG-O1 (megakaryocytic leukemia, kindly provided by Dr. M. Ogura, Aichi Cancer Center, Nagoya), Raji (Burkitt lymphoma), A172 (glioblastoma) and ACC-LC-80 (small cell lung carcinoma, kindly provided by Dr. T. Takahashi, Aichi Cancer Center) were used for RNA extraction.

Isolation of genomic clones A cosmid genomic library of human leukocyte DNA cloned in the Lorist B vector was kindly provided by Drs. H. Sakamoto and M. Terada (National Cancer Center Research Institute, Tokyo) and used for screening. The 369 bp 5′ c-*kit* cDNA probe was synthesized from HEL mRNA by reverse-transcribed polymerase chain reaction (PCR) using sense primer (5′-TCGCAGCTACCGCGATGAGA-3′) and antisense primer (5′-GGAAAAGCTTGGCAGGATCT-3′) according to the previously published cDNA sequence (nt 8–376).[1] In total, 4.5×10^5 clones were screened by colony hybridization using the ^{32}P-labeled probe.

Nylon filters were hybridized at 37°C for 48 h in a hybridizing solution containing 3×SSC, 50 m*M* Tris-HCl pH 7.5, 20 μg/ml transfer RNA, 20 μg/ml denatured salmon sperm DNA, 1 m*M* EDTA, 1×Denhardt's solution, 1% sodium dodecyl sulfate (SDS) and 50% formamide. The filters were washed in 2×SSC, 1× Denhardt's solution and 1% SDS at 37°C for 60 min followed by washing in 0.1×SSC and 0.1% SDS at 50°C for 30 min twice. Autoradiography was carried out by exposing filters on Fuji X-ray film with intensifying screens at −70°C for 24 h.

The DNAs of positive clones were purified after three or four sequential rounds of screening.

Subcloning and nucleotide sequence analysis The positive clone Cos17 was studied in detail. A 9.0 kb *Hin*dIII fragment which hybridized to the c-*kit* cDNA probe was subcloned into the pUC19 vector (Takara) for further analysis.

The 1.0 kb *Bam*HI fragment which was shown to be located immediately upstream of the exon 1 was excised and subcloned into the pUC19 vector. The nucleotide sequence of this fragment was determined by the dideoxy chain termination method using T7 DNA polymerase [Sequenase Ver. 2 (United States Biochemical Corp.)]. The upstream and downstream regions of this clone were also isolated and sequenced. Sequence-specific primers were utilized to obtain the complete sequences from both strands.

Primer extension analysis Total RNAs were prepared by the guanidine thiocyanate/cesium chloride method from various cell lines.[20] Primer extension was performed as described with some modifications.[20] A 27-mer antisense oligonucleotide (5′-CTGGACGCGAAGCAGTAGG-AGCAGAAC-3′), corresponding to nucleotide +37 to +63 relative to the translation start site,[1] was 5′ end-labeled with [γ-^{32}P]ATP (ICN) using T4 polynucleotide kinase (Takara), and 10^5 cpm of the labeled primer and 30 μg of total RNA were coprecipitated with ethanol. The pellet was dissolved in 30 μl of hybridization buffer (40 m*M* PIPES pH 6.4, 1 m*M* EDTA, 0.4 *M* NaCl and 80% formamide), denatured by heating at 85°C for 10 min and incubated at 30°C overnight. The extension reaction was performed with 100 units of Mo-MuLV reverse transcriptase (BRL) in 20 μl of RT buffer (20 m*M* Tris-HCl pH 8.3, 50 m*M* KCl, 2.5 m*M* $MgCl_2$, 10 m*M* DTT, 1 m*M* dNTP, 1 unit/μl RNasin and 50 μg/ml actinomycin D) at 42°C for 2 h. Then, 1 μl of 0.5 *M* EDTA and 1 μl of RNase A (1 mg/ml) were added with incubation at 37°C for 30 min. After phenol/chloroform extraction followed by ethanol precipitation, the extended products were analyzed on a 6% polyacrylamide sequencing gel alongside a dideoxy sequencing ladder of M13mp18 DNA.

S1 nuclease protection analysis The 971 bp genomic *Bam*HI fragment (from −992 to −22) was treated with calf intestinal phosphatase (Takara) and 5′ end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. After digestion with *Sma*I, the 159 bp *Sma*I-*Bam*HI fagment (from −180 to −22) was separated by electrophoresis on a 5% polyacrylamide gel and used as a probe.

The 5′ end-labeled probe (10^4 cpm) and 15 μg of total RNA were coprecipitated with ethanol. The pellet was dissolved in 20 μl of hybridization buffer (40 m*M* PIPES pH 6.4, 1 m*M* EDTA, 0.4 *M* NaCl and 80% formamide), denatured by heating at 90°C for 10 min and incubated at 50°C overnight. The hybridized mixture was digested with 400 units of S1 nuclease (Takara) in 200 μl of S1 nuclease buffer (250 m*M* NaCl, 50 m*M* NaOAc, 4.5 m*M* $ZnSO_4$ and 50 μg/ml carrier DNA) at 37°C for 30 min. After ethanol precipitation, the digested products were analyzed on an 8% polyacrylamide gel alongside a dideoxy sequencing ladder of M13mp18 DNA.

Plasmid construction for CAT assay CMV-LTR-CAT, kindly provided by Dr. I. S. Y. Chen (Jonsson Comprehensive Cancer Center, LA) was used as a positive control.[21] CMV-LTR-CAT was digested with *Kpn*I and *Sph*I to remove the promoter region; blunt ended with T4DNA polymease and self-ligated with T4DNA ligase. This plasmid, pUC0CAT, was used as a negative control.

The 1.0 kb *Bam*HI fragment just upstream of the c-*kit* exon 1 was subcloned into pSP73 (Promega) and named pSP*Bam*HI. The 1.0 kb 5′ to 3′ *Kpn*I-*Sph*I fragment from pSP*Bam*HI was ligated to the corresponding sites of CMV-LTR-CAT. The resultant construct containing the promoter sequence of the c-*kit* gene in the sense orientation to the bacterial chloramphenicol acetyltransferase (CAT) gene was named pUC*Bam*HICAT. To obtain the deletion constructs, pUC*Bam*HICAT was digested with 1) *Kpn*I and *Sac*II 2) *Kpn*I and *Xho*I 3) *Kpn*I and *Nae*I, blunt ended with T4DNA polymerase and self−ligated to produce 1) pUC*Sac*IICAT 2) pUC*Xho*ICAT and 3) pUC*Nae*ICAT, respectively. pUC*Bam*HICAT was also digested with *Sma*I and self-ligated to produce pUC*Sma*ICAT.

pUC*Eco*RICAT and pUC*Hin*dIIICAT were constructed as follows. The 1.4 kb *Eco*RI-*Sac*II fragment, located at −2000 to −604, was subcloned into pBluescript SK(+) (Stratagene) and named SK*Sac*II. The 1.4 kb *Cla*I-*Sac*II fragment from SK*Sac*II was ligated to the corresponding sites of pSP*Bam*HI to produce pSP*Cla*I. Then, the 2.0 kb *Eco*RV-*Sph*I fragment from pSP*Cla*I was ligated to *Sma*I- and *Sph*I-digested CMV-LTR-CAT. This was named pUC*Eco*RICAT.

The 2.0 kb *Hin*dIII-*Eco*RI fragment at the 5′ end of the clone Cos17 was subcloned into pBluescript SK(+) and named SK-HE. The 2.0 kb *Eco*RI-*Xba*I fragment from pSP*Cla*I was ligated to the corresponding sites of SK-HE to produce SK-EX. SK-EX was digested with *Cla*I, blunt ended, and digested with *Xba*I. The isolated 4.0 kb blunt-*Xba*I fragment was ligated to *Sma*I- and *Xba*I-digested CMV-LTR-CAT and named pUC*Hin*dIIICAT.

pUC*Bam*HICAT was digested with *Kpn*I and *Xho*I and then 5′-3′ delected using the ExoIII/mung bean nuclease system (Takara) to produce pUCEXCAT. pUC*Hin*dIIICAT was digested with 1) *Sph*I and *Sma*I 2) *Sph*I and *Sac*II 3) *Sph*I and *Afl*II, blunt-ended and self-ligated to produce 1) pUCSmBCAT, 2) pUCSaBCAT and 3) pUCABCAT, respectively.

CAT and β-galactosidase assay Transfection was performed by electroporation.[21] Cells were resuspended in IMDM. Then 10 μg of pUC0CAT or equimolar amounts of the other CAT constructs and 5 μg of the RSV-β-galactosidase plasmid were added to 0.7 ml of the cell suspension containing 7×10^6 cells, mixed gently and incubated on ice for 10 min. Electroporation was achieved by a single pulse from a Bio-Rad Gene Pulser (Richmond) under conditions giving maximal efficiency (250 V, 960 μF), followed by incubation on ice for 10 min. After pulsing, the cells were added to 10 ml of IMDM/10% fetal calf serum in 100 mm dishes and incubated at 37°C with 5% CO_2 for 24 h.

Assays to detect CAT enzyme activity were performed as described, with some modifications.[20, 21] At 24 h after transfection, cells were collected, washed twice with cold phosphate-buffered saline and resuspended in 0.25 *M* Tris-HCl pH 7.8. Cells were lysed by four cycles of freezing-thawing, followed by heating at 65°C for 10 min to eliminate the acetyl coenzyme A-consuming activity present in many cell types. Cell lysates were incubated with 0.25 μCi/10 μl of [^{14}C]chloramphenicol (Amersham), 4 m*M* acetyl coenzyme A (Sigma) and 0.25 *M* Tris-HCl pH 7.8 in a final volume of 100 μl at 37°C for 4 h. The incubation mixture was extracted with ethyl acetate and separated on silica gel thin-layer chromatography plates in chloroform:methanol (95:5). Radioactivity was quantitated by the use of a Bio-Image Analyzer BA100 (Fuji).

Each transfection experiment was repeated four times. Futhermore, each cell lysate was simultaneously analyzed for β-galactosidase activity as described.[20] The measured CAT activity was adjusted relative to β-galactosidase activity. The results for each CAT plasmid tested represent the average and are expressed relative to the negative control, pUC0CAT.

RESULTS

Isolation of genomic clones containing the human c-*kit* 5′ flanking region A human cosmid genomic DNA library was screened for clones containing the 5′ flanking region of the c-*kit* gene by using a 369 bp probe representing the 5′ sequence of the c-*kit* cDNA. Two overlapping clones (Cos17 and Cos20) which hybridized to the c-*kit* exon 1 and one clone (Cos1) which hybridized to the c-*kit* exon 2 were isolated.[22–24] Cos17 was used for further analysis.

The length of the cellular DNA inserted in clone Cos17 was approximately 35 kb. By restriction mapping and hybridization analysis, the DNA sequence upstream of exon 1 in the c-*kit* gene in Cos17 was found to be 4.0 kb long. Since the exon 2 cDNA probe failed to hybridize to Cos17, the length of intron 1 appears to be more than 30 kb. Fig. 1 shows the restriction map of the 9.0 kb *Hin*dIII(in polylinker)-*Hin*dIII fragment located at the 5′ end of the insert DNA in Cos17. Using a genomic Southern blot analysis we confirmed that this fragment had no gross rearrangement compared to the genomic DNA (data not shown).

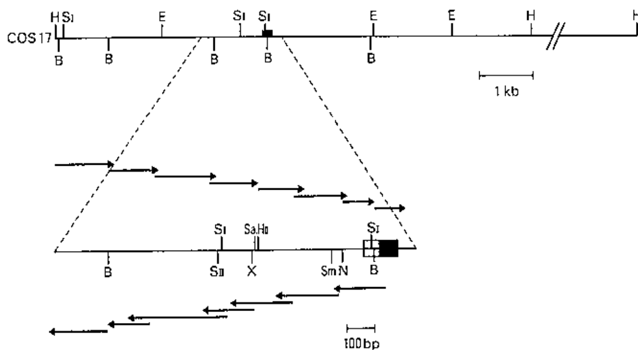

Fig. 1. Structure of the 5′ flanking region of the human c-*kit* gene. A restriction map of the 9.0 kb *Hin*dIII genomic fragment harboring exon 1 is shown at the top. This fragment is located at the 5′ end of the cosmid clone Cos17. The seven nucleotides (AGCTTAG) which constitute a part of *Hin*dIII and *Bam*HI sites at the 5′ end of this fragment are derived from Lorist B vector. An expanded map of the promoter and exon 1 region is shown at the bottom. The open box represents the 5′ untranslated region. The hatched box represents the coding region of exon 1. The arrows indicate the extent and direction of sequence analysis using specific oligonucleotide primers. Restriction sites: *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Hin*cII (HII), *Sac*I (SI), *Sac*II (SII), *Sal*I (Sa), *Sma*I (Sm) and *Nae*I (N).

```
-1215  GAGGGATGTC TTATTAGTTT ATCAGAATGC ATAGGGAAAG CTCATCTTTC
-1165  TTGGAGTAGG TCTTCCAATA AGTGTTCAAG CCCTGGGTCT TCCCTCACTA
-1115  CAATAATAAA CATACCTGCT CTACCCTCCT CACAAAGCTG TTATCGTGAG
-1065  CATGAAAAGA CGAGTCGCTG TGGGGAAGAA CTATTAACCC CTATACAGAT
-1015  ACTGTTGTTG CTTTCCGTTC AAGGATCCAC CATGAGACTG TGGGTAAGGC
-965   ACTTTCTGGG CAGTGAAATC GGGCTTAAGA GAATCATTGT ACTTTTATTT
-915   TGCTCCTGGG ACCGGCAAGG GCAGTTAAGA GCCCTAGCCC TAAAACTGTC
-865   GTCTGTTACA GAGGCAGAGA AATCATTACT GGGGTTGGCA GCGTTGGGGG
-815   GAGGGGGGAA AACGTGTATG AAAACCTGGG CTCCCAGAGC AAACTTCGCC
-765   ACGCGCCCTC CATCGCCCCT GTCTCCACAC AAGTCCCGGC AGGGGCTTCT
-715   TGGGCGGAGG CCACGCCGAC CAATAGGAAC CCACTGTGTT CCTACAGGTT
-665   ACGAAGCAGG TGGAGAAATT GAGCAGAAAC AATTAGCGAA ACCGGGCTCA
-615   GCCTTTACCG CGGTGCCAGG AGCTCCTAAC AGGCCTGGAG GGGAATGTCG
-565   GGGCTCAATT TCCTAACGCT CCCCTCCCCA TCCCCATGCC ACCTCCACGA
-515   GCAGCGGCGT CCAGCCTCCT CCCGCCCGAA CGTGCTCGAG GGGCGGGCAG
-465   TCGACCTTTA TTGTCTGGGG AGCACCTGGC AGGTGGCGGG CCCGTGCCCT
-415   AACGTGTGCG TGGTGCCCAG CTTCACAAAG CGAGCGGGCA GCACCTCCTT
-365   GGTCCGGGAA CGCCTCAGCC TGGCCGTCCA CATCCCAGGG GTGGAAAGGT
-315   GGAGAGAGAA AGGGGCTCCG GAGTCAAGAG CGGGGAGAGA GGGCGCGCGC
-265   CCTCCTCCTC CCGGCGGGCA CAGCCCCCGG GCATTAACAC GTCGAAAGAG
-215   CAGGGGCCAG ACGCCGCCGG GAAGAAGCGA GACCCGGGCG GGCGCGAGGG
-165   AGGGGAGGCG AGGAGGGGCG TGGCCGGCGC GCAGAGGGAG GGCGCTGGGA
-115   GGAGGGGCTG CTGCTCGCCG CTCGCGGCTC TGGGGGCTCG GCTTTGCCGC
-65    GCTCGCTGCA CTTGGGCGAG AGCTGGAACG TGGACCAGAG CTCGGATCCC
-15    ATCGCAGCTA CCGCGATGAG AGGCGCTCGC GGCGCCTGGG ATTTTCTCTG
                      M   R   G   A   R   G   A   W   D   F   L   C
       CGTTCTGCTC CTACTGCTTC GCGTCCAGAC AGgtgggaca cgcggctggc
        V   L   L   L   L   L   R   V   Q   T
       accccg
```

Fig. 2. Nucleotide sequence of the 5′ flanking region of the human c-*kit* gene. The nucleotide sequence of the 5′ flanking region of the c-*kit* gene, approximately 1.2 kb, is shown. The adenosine in the ATG codon was designated +1. Nucleotides upstream from this site were assigned negative numbers. Intron 1 is in lowercase latters. Restriction sites are also shown. The major transcription start sites are indicated by open circles (determined by primer extension) and closed circles (determined by S1 analysis). Putative binding sites for transcription factors are underlined. Four GA-rich elements reported by Yasuda *et al.*[17] are underlined using a dotted line.

Nucleotide sequence of the 5′ flanking region of the human c-*kit* gene The nucleotide sequence of the 5′ flanking region of the c-*kit* gene, approximately 1.2 kb, is shown in Fig. 2. The 5′ flanking region had no typical "TATA box" which exists in many eukaryotic gene promoters. An AT-rich sequence "ATTAA" could be identified at −233, and "CCAAT" or "CAAT" sequence could be identified at −696, −636 and −560. However, their positions did not correspond to any of the transcription start sites and they were not essential for promoter activity (see promoter assay). The sequence between nucleotides −515 and −1, which contains the putative promoter sequence, and the 5′ untranslated region had a high G+C content of 71%, whereas the region further upstream of −515 had a G+C content of only about 50%.

An analysis of the 5′ flanking region for the presence of consensus binding sites for the eukaryotic transcription factor revealed the following: the consensus sequence for Sp1 binding (GGGCGG) was present at −714, −494, −475 and −180.[25] The consensus sequence for the binding of basic helix-loop-helix (HLH) proteins (CANNTG) was present at −659, −443, −436, and −57.[26] One site had the consensus sequence for E47 binding. The consensus sequence for AP-2 binding (T/C) C (C/G) CC (A/C) N (GCG/CGC) was found at −770, −148 and −89, except for a mismatch of 1 bp at both −148 and −89.[27] Recently Yasuda *et al.* have reported the characterization of the mouse c-*kit* promoter,[17] and pointed out that four GA-rich elements (G/A) GGAG (G/A), conserved in humans at −163, −156, −130 and −117, also resemble binding sites for PU.1 and/or other Ets-domain proteins.[28]

Other consensus motifs or related sequences found in this 1.2 kb region are Myb [(C/T) AAC (G/T) G] at −894, −863 and −589,[29] GATA-1 [(A/T) GATA (A/G)] at −1197, −1075 and −1019[30] and the core sequence of the epidermal growth factor receptor specific transcription factor (ETF) binding site consisting of only guanosine and cytidine [CCCCCGGC] at −241,[31] although the functional significance of these sequences in the expression of c-*kit* is not yet clear.

Giebel *et al.* have also reported the partial nucleotide sequence of this region up to −565 in Fig. 2.[24] In this region they pointed out a putative melanocytes-specific "upstream element," a "downstream element" and a region strikingly similar to the promoters of the tyrosinase and *brown* locus (TRP-1) genes specifically transcribed in melanocytes.

Determination of the transcription start site of the human c-*kit* gene The transcription start site of the c-*kit* gene was determined by primer extension and S1 nuclease protection analysis. For primer extension, total RNAs were annealed to the ^{32}P-labeled antisense oligonucleotide containing nucleotides from +37 to +63 at the 3′ end of the exon 1 and reverse transcriptase reaction was carried out. As shown in Fig. 3A, two major extended bands (125 and 121 nucleotides) were observed in HEL, KU812 and MEG-O1 cells. The same bands were also deteceted in non-hematopoietic cells, i.e., ACC-LC-80 (small cell lung carcinoma) and A172 ((glioblastoma). All of these cell lines expressed the c-*kit* mRNA as reported by others although A172 cells expressed it at lower levels.[11, 12, 14, 32, 33] This has been confirmed by our

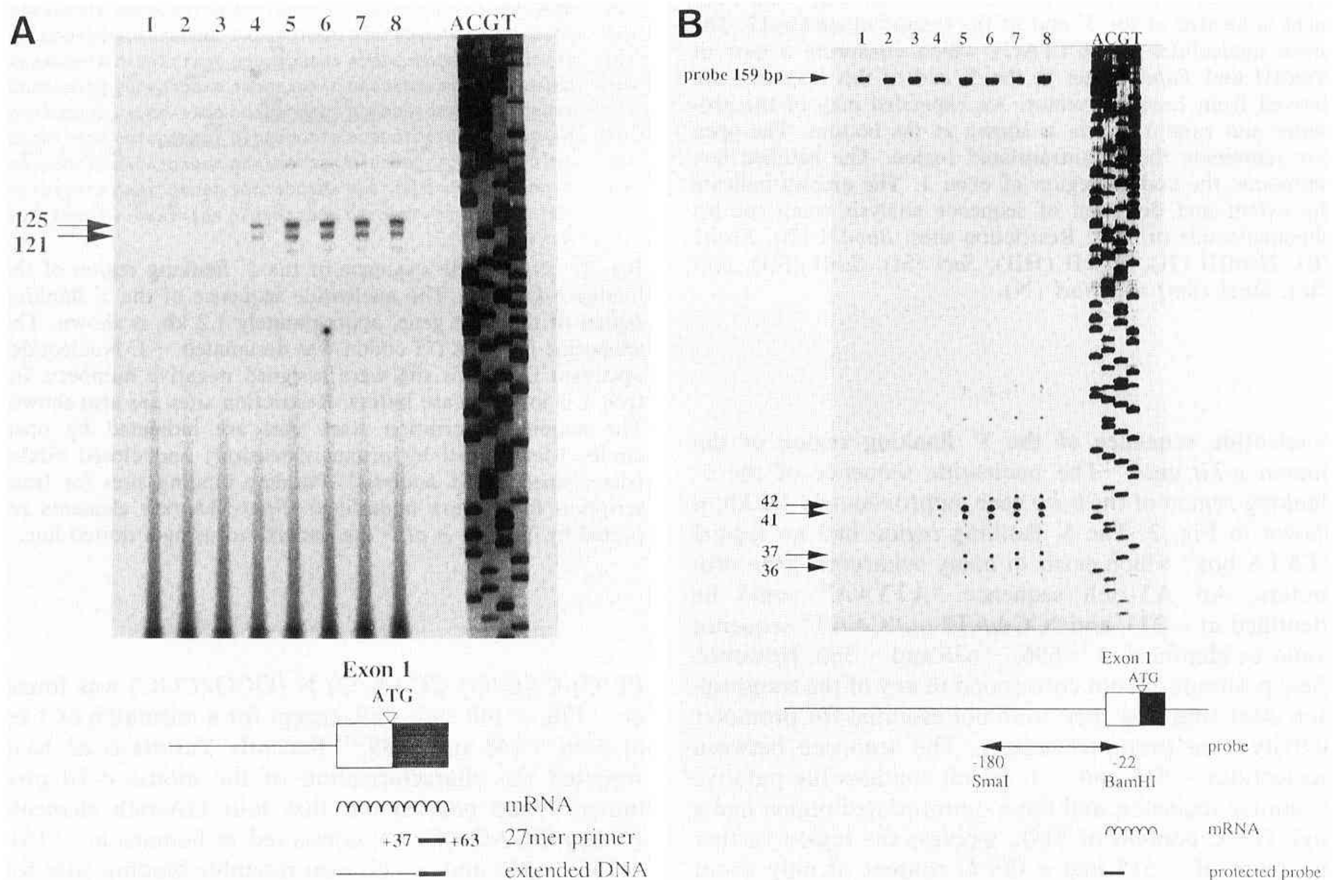

Fig. 3. Determination of the transcription start sites of the human c-*kit* mRNA. (A) Primer extension analysis was carried out by annealing ^{32}P-labeled 27-mer antisense oligonucleotide primer (nucleotide +37 to +63) to 30 μg of total RNAs followed by extension with Mo-MuLV reverse transcriptase. Two specific extended fragments (125 and 121 nucleotides) indicated by arrows were observed only in the cells which expressed the c-*kit* mRNA. Nucleotide sequence ladders of the M13mp18 DNA are shown on the right side as size markers. Lanes: 1) tRNA, 2) TALL-1, 3) Raji, 4) A172, 5) ACC-LC-80, 6) HEL, 7) KU812, 8) MEG-O1. (B) S1 nuclease protection analysis was performed by annealing a 159 bp ^{32}P-labeled DNA probe (nucleotide −180 to −22) to 15 μg of total RNAs. The DNA-RNA hybrids were digested with S1 nuclease (2000 U/ml). Four major protected fragments (42, 41, 37, 36 bp) indicated by arrows and several other faint fragments were observed in the cells which expressed the c-*kit* mRNA. Lanes: 1) tRNA, 2) TALL-1, 3) Raji, 4) A172, 5) ACC-LC-80, 6) HEL, 7) KU812, 8) MEG-O1. A schematic diagram showing the relationship of the c-*kit* 5′ genomic region to the DNA probe, the primer and the resultant fragments obtained by both analyses is shown at the bottom.

experiments (data not shown). In addition, a few faint bands upstream and downstream (142 and 117 nucleotides) were detected in a similar fashion in all these cells. No band was detected in tRNA or in cellular total RNAs obtained from TALL-1 and Raji cells which did not express the c-*kit* mRNA. These results suggest that the major transcription start sites of the c-*kit* gene are −62 and −58 relative to the translation start site.

To confirm the results obtained by primer extension, S1 nuclease protection analysis was carried out using a 159 bp ^{32}P-labeled DNA probe. This probe corresponds to the sequence from the −180 to −22 region of the genomic DNA, which includes the proposed transcription start sites. The DNA probe was hybridized with total RNAs and digested with S1 nuclease. Two major clusters of protected products (42 and 41 bp long, 37 and 36 bp long) corresponding to −63, −62, −58 and −57 bp from the translation start site were observed in all the cells which expressed the c-*kit* mRNA (Fig. 3B). A172 cells showed a similar pattern on longer exposure.

Since bands corresponding to the two major sites (−62 and −58) were detected by both primer extension and S1 nuclease protection analysis, we concluded that the major transcription start sites in the human c-*kit* gene are the cytosine and the guanine residues, 62 bp and 58 bp upstream of the translation start site. These results also showed that the major transcription start sites of the c-*kit* gene were essentially the same in both hematopoietic and non-hematopoietic cells such as small cell lung carcinoma and glioblastoma. We suggest that upstream faint bands at the −74 to −79 bp position may represent minor transcription start sites and that other bands around the major start sites are artificial products of S1 nuclease preparation, representing either overdigestion or overhanging ends, as described previously.[34)]

Functional analysis of the human c-*kit* gene promoter To examine the functional c-*kit* promoter activity and to localize the regulatory region for this promoter, the 5′ flanking genomic fragments of the c-*kit* gene were inserted upstream of the CAT reporter gene and a series of deletion mutants were constructed. Each construct was transfected by electroporation into HEL cells which express the endogenous c-*kit* mRNA. After 24 h in culture, cells were harvested and the lysates were assayed *in vitro* for the CAT activity. The transfection efficiency was adjusted for β-galactosidase activity which was expressed from the pRSV-β-gal DNA cotransfected with the c-*kit*-CAT plasmid DNAs. Structures of deletion constructs and relative promoter activity of these various constructs measured by the CAT activity assay are shown in Fig. 4A and 4B.

The CAT constructs containing less than 600 bp of the 5′ c-*kit* fragment from −604 to −22 bp had definite CAT activity compared to that of the negative control

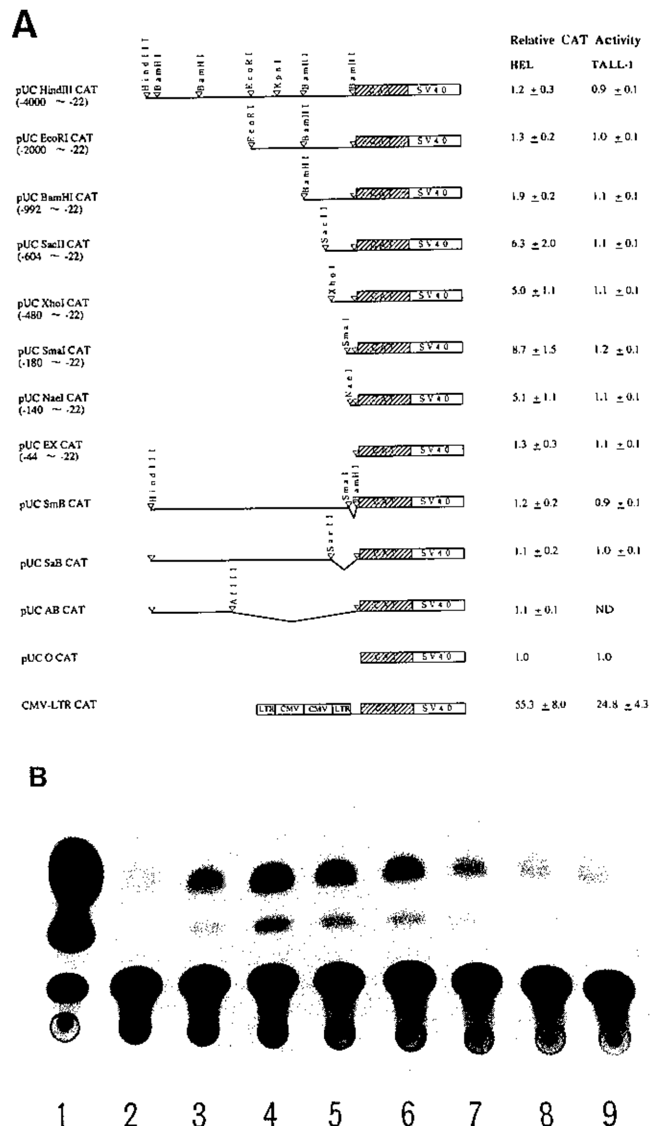

Fig. 4. (A) Construction of CAT plasmids and relative CAT activity in HEL and TALL-1 cells. Structures of deletion mutants are shown. The deletion mutants are named according to the restriction sites at the 5′ end. Relative CAT activity is indicated on the right. Cell extracts from both cells transiently transfected with each construct were assayed for their abilities to acetylate ^{14}C-labeled chloramphenicol. The measured CAT activity was adjusted according to the β-galactosidase activity. The result for each CAT plasmid expressed relative to the negative control plasmid pUC0CAT represents the average and standard error of four individual experiments. ND; not determined. (B) Autoradiogram of CAT activity in HEL cells transfected with each construct. The presence of acetylated forms demonstrates promoter activity. Lanes: 1) CMV-LTR-CAT, 2) pUC0CAT, 3) pUC*Nae*ICAT, 4) pUC*Sma*ICAT, 5) pUC*Xho*ICAT, 6) pUC*Sac*IICAT, 7) pUC*Bam*HICAT, 8) pUC*Eco*RICAT, 9) pUC*Hin*dIIICAT.

pUCOCAT. These results indicate that the 5′ flanking region of the c-*kit* gene has a transcriptional promoter activity in HEL cells despite the lack of a typical "TATA box." The pUC*Sma*ICAT, containing 159 bp *Sma*I-*Bam*HI fragment (−180 to −22), showed a higher promoter activity when compared to other c-*kit* constructs in HEL cells. This activity was about 16% of that with the positive control plasmid CMV-LTR-CAT, an efficient promoter in mammalian cells. On the other hand, the constructs containing more than 1 kb of the 5′ c-*kit* fragment and the 3′ deletion constructs showed a significant decrease in the CAT activity. The CAT activity of all constructs was hardly detectable in TALL-1 cells which did not express the c-*kit* mRNA at a detectable level.

These results indicate that a short sequence including a 120-bp stretch upstream from the transcription start site in the c-*kit* gene has a promoter activity in this transient CAT assay. Further, a negative regulatory element(s) may exist in an upstream region between −992 and −604.

DISCUSSION

In this study we isolated the 5′ flanking region of the human c-*kit* gene and characterized the transcriptional initiation sites and its promoter activity in hematopoietic cells. The 5′ flanking region lacked a typical "TATA box" and had a relatively high G+C content with four Sp1 binding sites. These features are reported to exist in the promoters of constitutively expressed "housekeeping" genes as well as certain receptor tyrosine kinase genes, including the epidermal growth factor receptor gene, the insulin receptor gene and the c-*fms* gene.[35-37]

We found two major transcription start sites, 62-bp and 58-bp upstream of the translation start site in human c-*kit* gene, by primer extension and S1 nuclease protection analyses. Additional bands detected by S1 nuclease analysis appear to be artifacts since such a microheterogeneity is frequently observed in S1 nuclease preparations, probably due to the tendency of this enzyme to produce nucleotide over-hanging ends and "end-nibble" products.[34, 38]

The localization of the transcription start sites at −62 and −58 bp is consistent with the observation reported by Giebel *et al.*[24] They suggested that the 5′-end of the human c-*kit* mRNA lies approximately between −120 and −65 bp, by using human melanocyte mRNA for reverse-transcribed (RT) PCR assay; RT-PCR with primer corresponding to −65 to −44 and primer located in exon 2 could detect amplified products, but primers further upstream failed to amplify cDNAs. In the case of the mouse c-*kit* gene, Yasuda *et al.* have recently identified the major transcription start site to be 58-bp upstream from the translation start site by primer extension and RNase protection analyses.[17] Therefore, the location of the transcription start sites appears to be highly conserved between the human and mouse c-*kit* genes.

The c-*kit* gene is known to be expressed in several normal tissues including bone marrow and in some malignant cells.[8, 9, 11-14] Here we have shown that the transcription start sites in the c-*kit* gene were essentially the same in hemetopoietic and non-hematopoietic cells such as small cell lung carcinoma and glioblastoma. In contrast, the c-*fms* (or CSF-1 receptor) gene, which shares a similar genomic intron/exon organization to the c-*kit* gene,[22-24] has been reported to carry two separate promoters for hematopoietic cells and non-hematopoietic cells.[37, 39] The c-*fms* transcripts of monocytes originate from the sequences immediately upstream of the initiation codon (exon 2), while the placental c-*fms* transcripts include a non-coding exon (exon 1). Therefore, the molecular mechanisms of transcriptional regulation in these two structually related genes, c-*kit* and c-*fms* could be different from each other. The fact that the transcription start sites of the mouse c-*kit* gene are identical in both mast cells and cerebellar tissue also suggests that the location of the transcription start sites in the c-*kit* gene is likely to be highly similar in different types of cells. Further studies using other types of cells which express the c-*kit* gene are necessary to confirm this hypothesis.

Functional promoter analysis revealed that the region extending from −180 to −22, relative to the translation start site, is important for the promoter activity of the c-*kit* gene in HEL cells. This region shows an especially high G+C content, i.e., about 80%, and carries putative binding sites for Sp1, AP-2, HLH proteins and Ets-domain proteins. These transcription factors are likely to play a role in the regulation of the c-*kit* gene. Among these factors, Ets-related proteins, such as PU.1 and Spi-B are expressed in hematopoietic cells including totipotent, erythroid and mast cells.[40] Thus, Ets-related proteins might contribute to the c-*kit* expression in hematopoietic cells. The sequence in this region is highly homologous (75%) between human and mouse and all of the binding motifs described above are conserved.[17] The corresponding sequence in the mouse c-*kit* gene also showed a maximal promoter activity among several constructs in HEL cells. These results suggest the physiological significance of the region spanning −180 and −22 in the regulation of the transcription of c-*kit*.

In addition, we showed the possible existence of a negative regulatory element(s) between −992 and −604. However, the 5′ flanking sequence of the mouse c-*kit* gene corresponding to −992 and −604 bp in humans was shown to have very weak, if any, negative regulatory activity in HEL cells.[17] Thus, it is not clear

yet whether this negative regulation usually occurs in the c-*kit* gene in different types of cells and in different species of animals.

Recently Tono *et al.* reported the analysis of W^{sh}/W^{sh} mutant mice.[41] These mice have a normal number of erythrocytes and germ cells, but they show a remarkable depletion of mast cells and have a white hair color. The lesions in the W^{sh}/W^{sh} mice are supposed to be due to mutation(s) in the regulatory region for the c-*kit* expression, because they possess a normal c-*kit* coding region. Thus, this phenotype suggests that the expression of the c-*kit* gene is regulated in a cell-type specific manner, although the same transcription start sites are used. It would be interesting to see what kinds of mutation(s) occur in the regulatory regions in the c-*kit* gene.

ACKNOWLEDGMENTS

We thank Dr. S. Hayashi (Kumamoto University) for helpful discussions, Dr. T. Takahashi (Aichi Cancer Center) for RNA of ACC-LC-80 cells, Dr. M. Ogura (Aichi Cancer Center) for MEG-O1 cells and Dr. I. S. Y. Chen (Jonsson Comprehensive Cancer Center) for CMV-LTR-CAT. We are also grateful to Drs. M. Terada and H. Sakamoto (National Cancer Center) for providing a human cosmid library and Mrs. Lata Seetharam for reading the manuscript. This work was supported by a Grant-in-Aid for Special Project Research on Cancer Bioscience (04253204) from the Ministry of Education, Science and Culture of Japan and a research grant from the Foundation for Promotion of Cancer Research in Japan.

(Received June 25, 1993/Accepted August 5, 1993)

REFERENCES

1) Yarden, Y., Kuang, W. J., Yang-Feng, T., Coussens, L., Munemitsu, S., Dull, T. J., Chen, E., Schlessinger, J., Francke, U. and Ullrich, A. Human proto-oncogene c-*kit*: a new cell surface receptor tyrosine kinase for an unidentified ligand. *EMBO J.*, **6**, 3341–3351 (1987).
2) d'Auriol, L., Mattei, M. G., Andre, C. and Galibert, F. Localization of the human c-*kit* protooncogene on the q11-q12 region of chromosome 4. *Hum. Genet.*, **78**, 374–376 (1988).
3) Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. and Bernstein, A. The proto-oncogene c-*kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse *W* locus. *Nature*, **335**, 88–89 (1988).
4) Geissler, E. N., Ryan, M. A. and Housman, D. E. The dominant-white spotting (*W*) locus of the mouse encodes the c-*kit* proto-oncogene. *Cell*, **55**, 185–192 (1988).
5) Russell, E. S. Hereditary anemias of the mouse: a review for geneticists. *Adv. Genet.*, **20**, 357–459 (1979).
6) Nocka, K., Tan, J. C., Chiu, E., Chu, T. Y., Ray, P., Traktman, P. and Besmer, P. Molecular bases of dominant negative and loss of function mutations at the murine c-*kit*/white spotting locus: W^{37}, W^{V}, W^{41} and *W*. *EMBO J.*, **9**, 1805–1813 (1990).
7) Reith, A. D., Rottapel, R., Giddens, E., Brady, C., Forrester, L. and Bernstein, A. *W* mutant mice with mild or severe developmental defects contain distinct point mutations in the kinase domain of the c-*kit* receptor. *Genes Dev.*, *4*, 390–400 (1990).
8) Nocka, K., Majumder, S., Chabot, B., Ray, P., Cervone, M., Bernstein, A. and Besmer, P. Expression of c-*kit* gene products in known cellular targets of *W* mutations in normal and *W* mutant mice—evidence for an impaired c-*kit* kinase in mutant mice. *Genes Dev.*, **3**, 816–826 (1989).
9) Keshet, E., Lyman, S. D., Williams, D. E., Anderson, D. M., Jenkins, N. A., Copeland, N. G. and Parada, L. F. Embryonic RNA expression patterns of the c-*kit* receptor and its cognate ligand suggest multiple functional roles in mouse development. *EMBO J.*, **10**, 2425–2435 (1991).
10) Witte, O. N. Steel locus defines new multipotent growth factor. *Cell*, **63**, 5–6 (1990).
11) Andre, C., d'Auriol, L., Lacombe, C., Gisselbrecht, S. and Galibert, F. c-*kit* mRNA expression in human and murine hematopoietic cell lines. *Oncogene*, **4**, 1047–1049 (1989).
12) Sekido, Y., Obata, Y., Ueda, R., Hida, T., Suyama, M., Shimokata, K., Ariyoshi, Y. and Takahashi, T. Preferential expression of c-*kit* protooncogene transcripts in small cell lung cancer. *Cancer Res.*, **51**, 2416–2419 (1991).
13) Wang, C., Curtis, J. E., Geissler, E. N., McCulloch, E. A. and Minden, M. D. The expression of the proto-oncogene c-*kit* in the blast cells of acute myeloblastic leukemia. *Leukemia*, **3**, 699–702 (1989).
14) Ikeda, H., Kanakura, Y., Tamaki, T., Kuriu, A., Katayama, H., Ishikawa, J., Kanayama, Y., Yonezawa, T., Tarui, S. and Griffin, J. D. Expression and functional role of the proto-oncogene c-*kit* in acute myeloblastic leukemia cells. *Blood*, **78**, 2962–2968 (1991).
15) Ashman, L. K., Cambareri, A. C., Bikto, L., Levinsky, R. J. and Juttner, C. A. Expression of the YB5.B8 antigen (c-*kit* proto-oncogene product) in normal human bone marrow. *Blood*, **78**, 30–37 (1991).
16) Broudy, V. C., Lin, N., Zsebo, K. M., Birkett, N. C., Smith, K. A., Bernstein, I. D. and Papayannopoulou, T. Isolation and characterization of a monoclonal antibody that recognizes the human c-*kit* receptor. *Blood*, **79**, 338–346 (1992).
17) Yasuda, H., Galli, S. J. and Geissler, E. N. Cloning and functional analysis of the mouse c-*kit* promoter. *Biochem. Biophys. Res. Commun.*, **191**, 893–901 (1993).
18) Martin, P. and Papayannopoulou, T. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science*, **216**, 1233–1235

(1982).
19) Hiraki, S., Miyoshi, I., Kubonishi, I., Matsuda, Y., Nakayama, T., Kishimoto, H., Machida, K., Masuji, H. and Kimura, I. Establishment of a T-cell line from human lymphosarcoma. *Gann,* **69,** 115–118 (1978).
20) Sambrook, J., Fritsch, E. F. and Maniatis, T. "Molecular Cloning: A Laboratory Manual," 2nd Ed. (1989). Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
21) Cann, A. J., Koyanagi, Y. and Chen, I. S. Y. High efficiency transfection of primary human lymphocytes and studies of gene expression. *Oncogene,* **3,** 123–128 (1988).
22) Andre, C., Martin, E., Cornu, F., Hu, W. X., Wang, X. P. and Galibert, F. Genomic organization of the human c-*kit* gene: evolution of the receptor tyrosine kinase subclass III. *Oncogene,* **7,** 685–691 (1992).
23) Vandenbark, G. R., deCastro, C. M., Taylor, H., Dew-Knight, S. and Kaufman, R. E. Cloning and structural analysis of the human c-*kit* gene. *Oncogene,* **7,** 1259–1266 (1992).
24) Giebel, L. B., Strunk, K. M. Holmes, S. A. and Spritz, R. A. Organization and nucleotide sequence of the human *KIT* (mast/stem cell growth factor receptor) proto-oncogene. *Oncogene,* **7,** 2207–2217 (1992).
25) Dynan, W. S. and Tjian, R. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature,* **316,** 774–778 (1985).
26) Blackwell, T. K. and Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science,* **250,** 1104–1110 (1990).
27) Imagawa, M., Chiu, R. and Karin, M. Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell,* **51,** 251–260 (1987).
28) Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., Maki, R. A., Gunther, C. V., Nye, J. A. and Graves, B. J. The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Dev.,* **4,** 1451–1453 (1990).
29) Biedenkapp, H., Borgmeyer, U., Sippel, A. E. and Klempnauer, K. H. Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature,* **335,** 835–837 (1988).
30) Orkin, S. H. GATA-binding transcription factor in hematopoietic cells. *Blood,* **80,** 575–581 (1992).
31) Kageyama, R., Merlino, G. T. and Pastan, I. A transcription factor active on the epidermal growth factor receptor gene. *Proc. Natl. Acad. Sci. USA,* **85,** 5016–5020 (1988).
32) Hibi, K., Takahashi, T., Sekido, Y., Ueda, R., Hida, T., Ariyoshi, Y., Takagi, H. and Takahashi, T. Coexpression of the stem cell factor and the c-*kit* genes in small-cell lung cancer. *Oncogene,* **6,** 2291–2296 (1991).
33) Lerner, N. B., Nocka, K. H., Cole, S. R., Qiu, F., Strife, A., Ashman, L. K. and Besmer, P. Monoclonal antibody YB5. B8 identifies the human c-*kit* protein product. *Blood,* **77,** 1876–1883 (1991).
34) Green, M. R. and Roeder, R. G. Definition of a novel promoter for the major adenovirus-associated virus mRNA. *Cell,* **22,** 231–242 (1980).
35) Ishii, S., Xu, Y. H., Stratton, R. H., Roe, B. A., Merlino, G. T. and Pastan, I. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. *Proc. Natl. Acad. Sci. USA,* **82,** 4920–4924 (1985).
36) Araki, E., Shimada, F., Uzawa, H., Mori, M. and Ebina, Y. Characterization of the promoter region of the human insulin receptor gene. *J. Biol. Chem.,* **262,** 16186–16191 (1987).
37) Roberts, W. M., Shapiro, L. H., Ashmun, R. A. and Look, A. T. Transcription of the human colony-stimulating factor-1 receptor gene is regulated by separate tissue-specific promoters. *Blood,* **79,** 586–593 (1992).
38) Itoh, F., Ishizaka, Y., Tahira, T., Yamamoto, M., Miya, A., Imai, K., Yachi, A., Takai, S., Sugimura, T. and Nagao, M. Identification and analysis of the *ret* proto-oncogene promoter region in neuroblastoma cell lines and medullary thyroid carcinomas from MEN2A patients. *Oncogene,* **7,** 1201–1206 (1992).
39) Visvader, J. and Verma, I. M. Differential transcription of exon 1 of the human c-*fms* gene in placental trophoblasts and monocytes. *Mol. Cell. Biol.,* **9,** 1336–1341 (1989).
40) Ray, D., Bosselut, R., Ghysdael, J., Mattei, M. G., Tavitian, A. and Moreau-gachelin, F. Characterization of Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. *Mol. Cell. Biol.,* **12,** 4297–4304 (1992).
41) Tono, T., Tsujimura, T., Koshimizu, U., Kasugai, T., Adachi, S., Isozaki, K., Nishikawa, S-I., Morimoto, M., Nishimune, Y., Nomura, S. and Kitamura, Y. c-*kit* gene was not transcribed in cultured mast cells of mast cell-deficient W^{sh}/W^{sh} mice that have a normal number of erythrocytes and a normal c-*kit* coding region. *Blood,* **80,** 1448–1453 (1992).