THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 22, Issue of June 3, pp. 15861–15867, 1994
Printed in U.S.A.

Identification of a Tripartite Basal Promoter Which Regulates Human Terminal Deoxynucleotidyl Transferase Gene Expression*

(Received for publication, December 9, 1993, and in revised form, March 16, 1994)

Dipa Bhaumik‡, Baoli Yang, Theoni Trangas§, Jeffrey S. Bartlett, Mary Sue Coleman, and David H. Sorscher¶

From the Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-7260

In order to locate the promoter region of the human terminal deoxynucleotidyl transferase gene, serially truncated segments of the 5′-flanking region of the gene were cloned into a chloramphenicol acetyltransferase reporter vector. Transient transfection analyses of the terminal transferase-reporter gene constructs identified the basal promoter region within −34 to +40 base pairs relative to the transcription start site. Three promoter elements were defined in this region. The primary element is within 34 base pairs upstream of the transcription start site. The CAP site is 62 base pairs upstream of the translation start site. The secondary element involves sequences around the transcription start site. The third is located 25 base pairs downstream from the initiation site (+25 to +40). This tripartite basal promoter was not tissue specific; similar patterns of promoter activity were observed in terminal transferase expressing and non-expressing cells. Transfection analyses also indicated the presence of negative regulatory elements upstream of the basal promoter region, and these elements were preferentially active in cells expressing terminal transferase.

Immunoglobulin (Ig) and T-cell receptor gene assembly is attained through gene rearrangement of a large number of separately coded variable (V), diversity (D), and joining (J) gene segments (Ikuta *et al.*, 1992). It is well known that the diversity of the immune repertoire is greatly increased by the addition of nucleotides (N regions) at the D/J and V/DJ junctions during Ig and T-cell receptor gene rearrangements (Desiderio *et al.*, 1984; Schatz and Baltimore, 1988). Terminal deoxynucleotidyl transferase (TdT),[1] a template-independent DNA polymerase, that randomly polymerizes deoxyribonucleoside triphosphates onto initiator DNAs (Bollum, 1974) has been shown to be the only enzyme involved in N region addition (Komori *et al.*, 1993; Gilfillan *et al.*, 1993). Although we understand the biochemistry and role of this enzyme, little is known about the expression of the TdT gene. TdT expression is restricted to prelymphocytes at a specific stage of differentiation and during gene rearrangements (Deibel *et al.*, 1983). Study of this gene provides us with an excellent opportunity to understand both the activation and repression of gene transcription. We have thus undertaken the analysis of the regulation of the human TdT gene.

Regulation of transcription initiation is an important mechanism for control of gene expression. Most eukaryotic promoters contain a TATA box which is usually located 25–30 nucleotides upstream of the transcription initiation site (Sawadogo and Sentenac, 1990). A TATA-binding protein binds to this TATA motif, (Roeder, 1991) and along with RNA polymerase II and other additional factors directs the precise location of the transcription initiation site (Nakajima *et al.*, 1988; Pugh and Tjian, 1992). Besides these TATA-containing genes there are numerous other genes transcribed by RNA polymerase II that lack a TATA box upstream of the transcription start site (Weis and Reinberg, 1992). A number of studies have revealed that several of the TATA-less genes contain a DNA sequence (termed the initiator element or Inr) that encompasses the transcription start site, and is responsible for specifying the precise initiation site for transcription (Zenzie-Gregory *et al.*, 1992; Weis and Reinberg, 1992). In TATA-less promoters the Inr element plays an important role in the correct positioning of RNA polymerase II. The murine TdT gene, which lacks a TATA box and binding sites for common transcription factors, like Sp1, contains an Inr element that is responsible for the basal transcription of the gene. (Smale and Baltimore, 1989). The human TdT gene also lacks a consensus TATA box, but visual inspection of the 5′ proximal sequences of the gene failed to locate an obvious Inr element.

The 5′ upstream sequences of human and mouse TdT do share some homology. Alignment of the 5′ sequences of the human and mouse TdT genes revealed 72% homology within a region corresponding to −186 to +51 bp in the mouse gene when three insertions of 9, 20, and 27 bp of the human sequence were excluded (Lo *et al.*, 1991). Interestingly, our study shows that the 27-bp insertion corresponds precisely to the position of the transcription start site of the human gene identified *in vivo*. The limited homology between the 5′ sequences of the murine and the human TdT genes coupled with the apparent absence of a mouse Inr element homolog suggested that the mouse and the human TdT genes employed different promoter elements. Our goal was to study the 5′ sequences in detail and to identify control elements participating in human TdT gene regulation. We have demonstrated that transcription of the human TdT gene initiated at a single nucleotide. Transient transfection analyses of the human TdT gene promoter region identified three requisite elements located upstream of the start site, at the start site and interestingly, downstream of the start site that were necessary for correct initiation and optimal promoter activity.

* This work was supported by National Institutes of Health Grant RO1CA19492 and Tobacco Research Council Grant 2872R1. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Dept. of Pathology, Stanford University School of Medicine, Stanford, CA 94305-5324.

§ Present address: Papanikolaou Research Center, Hellenic Anti-Cancer Institute, 171 Alexandras Ave., Athens 11594, Greece.

¶ To whom correspondence should be addressed. Tel.: 919-966-6292; Fax: 919-966-2852.

[1] The abbreviations used are: TdT, terminal deoxynucleotidyl transferase; CAT, chloramphenicol acetyltransferase; CMV, cytomegalo virus; bp, base pair(s); Inr, initiator; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; bp, base pair(s).

MATERIALS AND METHODS

Cell Lines—The cells used in this study were established cell lines of leukemia-lymphoma origin and were found to represent different, early stages of lymphocyte maturation (Drexler *et al.*, 1985). The human lymphoid cell lines KT-1, MOLT-4, (pre-T), NALM-6 (pre-B) which expressed TdT and KOPN-8 (pre-B), KE-37 (mature-T cells) that did not expressed TdT were used. The cells were maintained in RPMI-1640 media supplemented with 10% fetal bovine serum (Trangas and Coleman, 1989).

Isolation of RNA and 5′-RACE—RNA was isolated from cells by guanidine isothiocyanate cell lysis and centrifugation through CsCl cushions (Sambrook *et al.*, 1989). 5′-RACE reactions were as described by Bethesda Research Laboratories. cDNA was synthesized using 2 pmol of antisense primer (RACE5- or RACE4-, see below) and 1 μg of total RNA. After RNase digestion the cDNA was isolated and incubated with TdT and 200 μM dCTP. Five μl of the cDNA containing the added dC sequences was amplified by PCR using 20 pmol of both the appropriate internal antisense primer and BRL anchor primer in a volume of 50 μl. PCR products were separated on an agarose gel and were sized by comparison with the migration of *Hin*cII-digested ϕX174 DNA markers (BRL). The primers used for cDNA synthesis were 5′-TTGTGAGCTGACTTGTACTTTCTGTG-3′ (RACE5-, exon-2 sequences) and 5′-CCGAACCCGAGTTGTTCTCTGCTAC-3′ (RACE4-, exon-2 sequences upstream of RACE5-). The internal primers for PCR were 5′-CATTTTCAACCCTGAACCCTTTCC-3′ (RACE3-, exon-1 sequences), 5′-GAAATTTGATGTCTTGAGGAGAGGAGGCC-3′ (225-, exon-1 sequences upstream of RACE3-), 5′-GGTCTCTTCTTCCGAGGGCTCAAGTGGGAC-3′ (223-, exon-1 sequences upstream of RACE3-) and RACE4-. The oligonucleotides were synthesized at the University of North Carolina Chapel Hill, Lineberger Comprehensive Cancer Center using an Applied Biosystems 394 Synthesizer. All primers were complementary to human TdT mRNA sequences (Riley *et al.*, 1988) except the BRL anchor primer.

Construction of TdTCAT Plasmids—The -930CAT DNA was constructed by first subcloning a 995-bp *Bam*HI fragment of the 5′ noncoding region of the human TdT (995 bp upstream of the ATG) into the *Bam*HI site of pUCI3. pUCI3 is a pUC-based plasmid in which the multiple cloning site has been modified in our laboratory to facilitate the subcloning of 5′-flanking regions of human TdT. A *Hin*dIII/*Pst*I fragment from pUCI3 containing the 995 bp of TdT was then subcloned into the *Hin*dIII/*Pst*I site of pCAT-Enhancer (Promega) to produce the -930CAT DNA. The numbering for the vectors corresponds to the distance in base pairs upstream or downstream from the TdT transcription start site. The vectors containing -656CAT and -356CAT were constructed similarly, except that a *Bam*HI/*Eco*RI or a *Bam*HI/*Acc*I fragment (721 or 421 bp upstream of the ATG) was cloned into the *Bam*HI/*Eco*RI or *Bam*HI/*Acc*I site of PUCI3. The -286CAT, -202CAT, -54CAT, -34CAT, -5CAT, +25CAT, +41CAT, and -54 to +25CAT vectors were constructed by amplifying the appropriate region by using the polymerase chain reaction and then subcloning the DNA fragments. Oligonucleotide primers (Fig. 2) which contained a *Hin*dIII site in the 5′ primers and *Pst*I site in the 3′ primers were used in the reactions. The amplified products were then subcloned into the *Hin*dIII/*Pst*I sites of pCAT-Enhancer. For the constructs -286CAT, -202CAT, and -54CAT, the reverse primer used was R1 which encompasses the translational start site. The reverse primer R2 was used for making the constructs -34CAT, -5CAT, +25CAT and +41CAT. These constructs included 55 bp of exon-1. R3 was used as the reverse primer to make -54 to +25 CAT. The -286CAT(R) DNA (-286 bp upstream of ATG in the reverse orientation) was made by first isolating the fragment from -286CAT DNA. The isolated fragment was cloned by blunt-end ligation into the *Sma*I site of pUCI3. A clone with the fragment in the reverse orientation (determined by DNA sequencing) was then digested by *Hin*dIII/*Pst*I and subcloned into the *Hin*dIII/*Pst*I site of pCAT-Enhancer. The exon-1CAT DNA was made in a way similar to -286CAT(R) except that a *Bam*HI/*Ava*I fragment containing exon-1 of TdT was subcloned into the *Sma*I site of pUCI3. All plasmid constructs were analyzed by dideoxy sequencing (U. S. Biochemical Corp. sequencing kit) and purified by polyethylene glycol precipitation followed by cesium chloride gradient centrifugation.

DNA Transfection—For DNA transfection experiments, cells in log phase at a density between 5×10^5 to 6×10^5 were used. Harvested cells were pelleted, washed, and suspended in a volume of 10^7 cells/300 μl of phosphate-buffered saline + 7.5% fetal bovine serum (Matsushima *et al.*, 1992). For each transfection, 35 μg of the CAT plasmid to be tested and 5 μg of the plasmid pCMVβ (coding for β-galactosidase activity) in a volume not exceeding 20 μl were added to 300 μl of cell suspension in a gene pulser electroporator cuvette (Bio-Rad). The cells were immediately electroporated using a Bio-Rad Gene Pulser. Electroporation parameters were 260 V and 960 microFarrads by using the Bio-Rad capacitance extender. After electroporation the cells were transferred to tissue culture flasks containing 12 ml of media and incubated at 37 °C for 48 h.

Assay of CAT and β-Galactosidase Activities—Forty-eight h after transfection the cells were harvested from tissue culture flasks and transferred to Eppendorf tubes. The cells were washed with phosphate-buffered saline and resuspended in 250 mM Tris-HCl, pH 8.0, at a density of 10^6 cells/25 μl. Cells were lysed by three freeze-thaw cycles. The samples were then centrifuged, and 30 μl of the supernatant was subsequently used to measure β-galactosidase activity by the method of Sambrook *et al.* (1989). The activity was expressed as $OD_{420} \times 100$ extrapolated to 1 h of reaction.

CAT activity in the supernatants was determined after heating for 10 min at 65 °C (Crab and Dixon, 1987). Thirty μl of supernatant was adjusted to 125 μl with addition of 0.25 M Tris-HCl, pH 8.0, 2.5 μl of 10 mg/ml butyryl-CoA (Sigma) and 0.1 μCi of [^{14}C]chloramphenicol (52 mCi/mmol, Dupont-New England Nuclear), and reactions were incubated at 37 °C. Depending on the cell line used the reaction time was varied from 15 min to 1 h to keep product formation in the linear range. The reactions were terminated by adding 300 μl of xylene, and the reaction products were isolated by the phase extraction procedure of Seed and Sheen (1988). CAT activity was expressed as counts/minute/hour of radioactive butyrylated chloramphenicol. These numbers were then normalized for equal transfection efficiency within a given cell type by calculating the CAT activity/100 units of β-galactosidase activity. The normalized data for a particular construct, for transfection experiments conducted on different days, varied by 2-fold. Data for the CAT assays were expressed as the mean ± SD of four to seven independent transfection experiments using at least two different plasmid preparations of the same construct.

Primer Extension Analysis—Forty-eight h after transfection cells (5×10^7 to 10^8) were harvested, and total RNA was isolated by the guanidium/CsCl method (Sambrook *et al.*, 1989). For primer extension 50 μg of total RNA was annealed to ^{32}P-end-labeled oligonucleotide 5′-GGGATATATCAATGGTATATC-3′. This oligonucleotide is complementary to the CAT coding region beginning 43 bp downstream of the CAT translational start site. To anneal primer and target RNA, samples were heated for 8 min at 85 °C in 30 μl of annealing buffer (0.25 M KCl, 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA) and then incubated for 60 min at 50 °C, followed by slow cooling for 15 min at room temperature. After ethanol precipitation samples were resuspended in reverse transcriptase buffer (50 mM Tris-HCl, pH 8.0, 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol) and 0.5 mM deoxynucleotides. After addition of 20 units of RNASIN (Promega) and 200 units Moloney murine leukemia virus reverse transcriptase (BRL), samples were incubated for 60 min at 37 °C. cDNA products were incubated with 20 μg/ml RNase (Ambion) for 30 min and then treated with proteinase K before phenol-chloroform extraction and ethanol precipitation. Samples were analyzed on a 6% polyacrylamide, 8 M urea denaturing sequencing gel. Primer extension cDNA products were visualized by autoradiography, and the sizes of the products were determined by comparison with ^{32}P-end-labeled ϕX174/*Hae*III markers and DNA sequencing ladders.

RESULTS

Mapping of the Transcription Start Site of the Human TdT Gene—Standard methods for determining the transcription start site of the human TdT gene were unsuccessful in our hands since this gene is transcribed at a very low rate *in vivo*. Therefore, we sought to identify the transcription start site for the human TdT gene by using the powerful 5′-RACE technique. This highly sensitive technique employs isolation of RNA, synthesis of cDNA, addition of C residues to the cDNA, amplification of the product by PCR, and sequencing of this product. To ensure that the template in these reactions was RNA, a series of primers, each of which produced a different length of amplified product was used (Fig. 1*A*). PCR products of the predicted size were observed when RNA from the TdT-positive cell line NALM-6 was used in the reaction (Fig. 1*B*, *lanes 1, 2,* and *5–7*). The sizes of the PCR products detected in each lane were consistent with a transcription start site of approximately 50–75 bases 5′ to the ATG translational start site. In contrast, RNA from the TdT-negative cell line KOPN-8 yielded no specific PCR

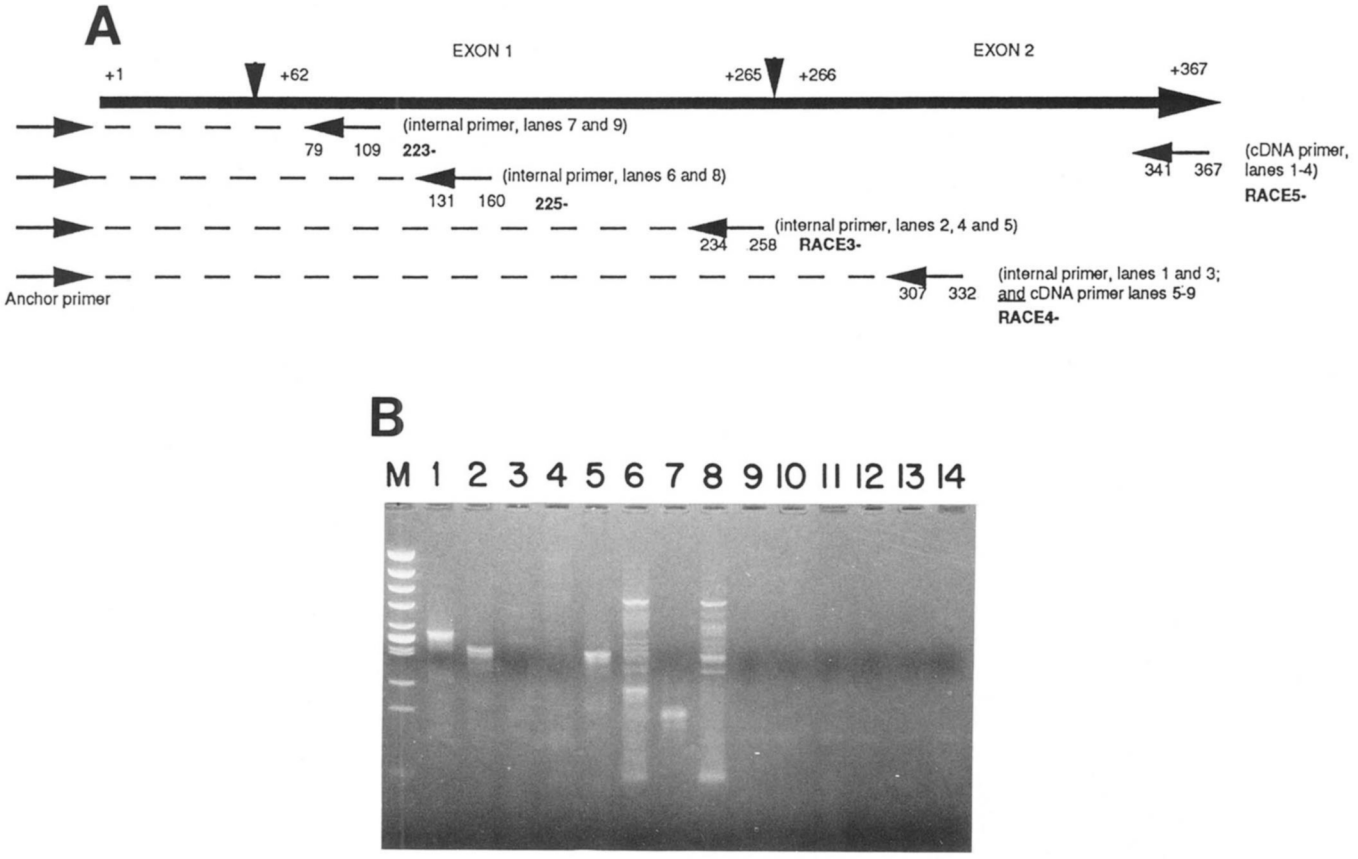

FIG. 1. **Identification of the transcription start for the human TdT gene by 5′-RACE.** *A*, schematic illustration of 5′-RACE experiment shown in *B*. The *bold arrow* pointing to the right represents transcribed sequences from the +1 start site to nucleotide +367. *Vertical arrows* point to the ATG at +62 and the exon1/exon2 boundary at +265/+266. Primers for cDNA synthesis (*RACE5-* and *RACE4-*) and PCR (*RACE4-*, *RACE3-*, *225-*, *223-* and *anchor primer*) are indicated by *horizontal arrows* with their nucleotide location relative to the transcription start site below each arrow. *Dashed lines* represent the PCR products shown in *B*. *A* is an approximate line drawing. *B*, *lanes 1*, *2*, and *5–7*, represent TdT-positive NALM-6 PCR products. *Lanes 3*, *4*, *8*, and *9* represent TdT-negative KOPN-8 PCR products. cDNA was synthesized with the RACE5- antisense primer (*lanes 1–4*) or with RACE4- antisense primer (*lanes 5–9*). Antisense primers upstream from RACE5- (*RACE4-*, *lanes 1* and *3*) or upstream of RACE4- (*RACE3-*, *lanes 2*, *4*, and *5*; *225-*, *lanes 6* and *8*; *223-*, *lanes 7* and *9*) were used as internal antisense primers for PCR. Some background was observed after using the 223- primer (*lanes 6* and *8*). *Lanes 10–14* were negative control 5′-RACE reactions. *Lane 10*, TdT sense primer with NALM-6 RNA. *Lane 11*, C residues were not added to the cDNA. *Lanes 12–14*, no cDNA was used in the PCR.

products in this experiment (Fig. 1*B*, *lanes 3*, *4*, *8*, and *9*). The precise transcription start site (a T) at 62 bp 5′ upstream of ATG was determined by cloning and sequencing one of the PCR products.

To confirm the size of the 5′ non-translated region of the RNA, another experiment was performed. ^{32}P-End-labeled oligonucleotides complementary to sequences upstream or downstream of the putative transcription start site were hybridized to mRNA from TdT-positive cells by using Northern analysis. These experiments showed that oligonucleotides antisense to mRNA sequences downstream but not to DNA sequences upstream of the proposed transcription start site hybridized to the RNA and were detected on the Northern blots (data not shown).

Functional Analysis of the 5′-Flanking Sequence of the Human TdT Gene—In this study we sought to establish a functional assay for potential promoter elements in the 5′ upstream region of the human TdT gene. This region was sequenced when the gene was cloned (Riley *et al.*, 1988). Within 995 bp upstream from the translation start site ATG, a number of known transcription factor-binding sites were present (Fig. 2). AP2-binding sites were identified in the regions −55 to −62 and +33 to +41. There was an Ets-binding site at −101 to −95 which in the murine TdT gene has been shown to be essential for transcription (Ernst *et al.*, 1993). There were several binding sites for INF-1, a binding site for Oct-1, and a TATA box. In order to test the function of these elements *in vivo*, a transient transfection assay was developed.

The cells we elected to use for these experiments were undifferentiated lymphoblastoid cells that were well established in continuous suspension cell culture. The TdT-positive lines used were MOLT-4, NALM-6, and KT-1 cells. The TdT-negative lymphoblastoid line was KE-37. These permanent cell lines are the only source for TdT expression (Drexler *et al.*, 1985) and are highly resistant to transfection (Lo *et al.*, 1991).

Calcium phosphate, DEAE-dextran, and lipofection transfection protocols failed to promote efficient transfer of control plasmids. By contrast, efficient transfection of all cell lines was obtained by electroporation following optimization of the density of the cells at the time of harvest, the cell concentration, buffer composition of the cell suspension to be electroporated, and the temperature at which the cell suspension was kept prior to electroporation.

Using the optimized electroporation protocol, a series of CAT constructs containing various lengths of the 5′-flanking sequences of the human TdT gene was transiently transfected into the KT-1, MOLT-4, or NALM-6 cells (TdT expressing) or the KE-37 (TdT non-expressing) cells. The CAT activity was measured 48-h post-transfection. CAT activities were corrected for transfection efficiencies by cotransfecting the cells with pCMVβ DNA and normalizing the CAT activity/100 units of β-galactosidase activity. The results from the first series of constructs are shown in Fig. 3. Identical data were observed with all three positive cell lines, but for simplicity, only the results obtained with KT-1 cells are illustrated in the figure.

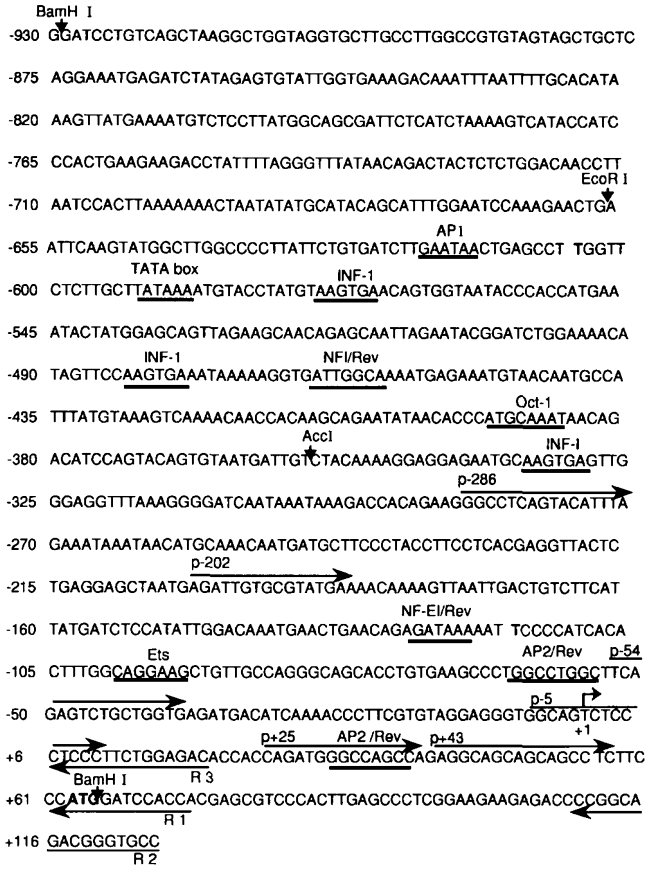

FIG. 2. **Nucleotide sequence of the 5′-flanking region and part of exon 1 of the human TdT gene.** The transcription initiation site is marked as *+1*. Primers used to generate deletion constructs are indicated by *arrows* above and below the sequences. The *Bam*HI, *Eco*RI, and *Acc*I sites are marked. The putative binding regions, based on sequence analysis for well established transcription factors are *underlined*.

The pattern of expression of CAT activity, varied in TdT-positive (*KT-1*) and TdT-negative (*KE-37*) lymphoblastoid cells (Fig. 3). In both cell types the regions 54 and 34 bp upstream from the transcription start site (*-54CAT*) and (*-34CAT*) were capable of directing maximum CAT synthesis in cells. This activity was assigned a value of 100%, which corresponded to 400 pmol of product when KT-1 cell extracts were used in the CAT assays and 500 pmol when KE-37 cell extracts were used. The CAT activity obtained with cells transfected with the other constructs were scaled to these optimal activities.

Elongation of the 5′ upstream sequence beyond -54 diminished CAT gene expression in both cell types (Fig. 3). However, a differential inhibition of transcription was observed. The plasmid constructs -930CAT and -656CAT exhibited no promoter activity over the basic vector (pCAT-Enhancer) alone in KT-1 cells whereas in KE-37 both of these constructs showed promoter activity that was 25–30% of that observed with the -34CAT construct. Likewise, the -356CAT and -202CAT constructs exhibited detectable CAT activity in the KT-1 cells although it was consistently lower than in KE-37 cells.

When 29 additional bases were removed from the -34CAT plasmid to form the -5CAT construct, CAT expression was still obtained. However, the level was reduced compared with -34CAT, and the reduction was more severe in the TdT-positive than in the TdT-negative cell line. Deletion of bases past the transcription start site (+41CAT and exon-1CAT) abolished all promoter activity in both cell types. The -286CAT(R) construct in which the 5′-flanking region was cloned in the reverse orientation elicited CAT activity comparable to the same construct in the correct orientation.

Detailed analyses of the region around the transcription start site revealed that 50% of maximum CAT activity was obtained with the sequences spanning +25 to ATG (Fig. 4). Deletion of the region +25 to ATG (*i.e.* the -54 to +25CAT construct) did not abolish CAT activity indicating that the -54 to +25 region had promoter activity but did not retain the optimal activity associated with -54CAT. Thus the sequence from +25 to +40 was necessary but not sufficient for optimal transcription.

Results identical to those in Figs. 3 and 4 were obtained when constructs containing pCAT-Basic vectors which lack the SV40 enhancer region were transfected into lymphoid cells (data not shown). Determination of the effect of the SV40 enhancer on transcription efficiency in this system was thus precluded.

To test TdT promoter activity in non-lymphoid cells, all the pCAT-Enhancer TdT constructs were transfected into Hela cells. However, a high background level of transcription was observed in these cells that masked the effects of the various TdT gene constructs. Thus, Hela cells were not used as a non-lymphoid control for transfection experiments.

Primer Extension Analysis of RNA Isolated from Transfected Cells—To confirm that reaction products measured by CAT activity assays, initiated at the TdT transcription start, primer extension analysis was carried out on RNA isolated from transfected cells. The results of two representative experiments are shown in Fig. 5. KE-37 and KT-1 cells were transfected with the constructs -54CAT (Fig. 5, *lanes 2*, *5*, *7*, and *9*), +25CAT (Fig. 5, *lanes 3* and *6*), and -54 to +25CAT (Fig. 5, *lanes 8* and *10*). The control DNA in these experiments was the basic vector pCAT-Enhancer (Fig. 5, *lanes 1* and *4*). The RNA employed in the primer extension reaction was isolated from cells after 48 h in culture. The predicted cDNA reaction product from -54CAT was 171 bp long. The vector sequence contributed 109 bp, and the transcription start was expected to be at about 62 bp from the translation start site. A major band was visible corresponding to 70 bp from the ATG when primer extension analysis was done on RNA isolated from -54CAT-transfected cells. This site mapped closely (8 bases longer) to the transcription start site determined by 5′-RACE (Fig. 1) and by *in vitro* transcription assay (data not shown). The difference in the transcription start site is probably due to technical details of the assays. Manipulation of the TdT mRNA during the 5′-RACE procedure could easily result in loss of a few bases. Alternatively, the sizes of the cDNA products measured after primer extension could be slightly inaccurate due to lane to lane variability in the electrophoresis procedure. The doublet observed at the major start site might represent transcription initiation at either of two bases. Alternatively, the doublet is an artifact of the primer extension reaction. A faint band (at least 4-fold less intense than the major band) was also observed at 42 bp upstream from ATG when -54CAT was used as the transfecting DNA. Both of these major and minor bands were consistently absent in analyses of RNA isolated from cells transfected with the +25CAT construct. This result suggested that sequences upstream from the start site and surrounding the start site were important for correct initiation of transcription.

The construct containing sequences spanning -54 to +25 was found to initiate transcription accurately since a cDNA product of the expected size (131 bp) was observed after transfection of this DNA construct into KT-1 and KE-37 cells (Fig. 5, *lanes 8* and *10*). These data indicated that accurate transcription could occur when only two parts of the basal promoter were present. The primer extension analysis of CAT mRNA from transfected cells coupled with the results of the CAT assay (Fig. 4) sug-

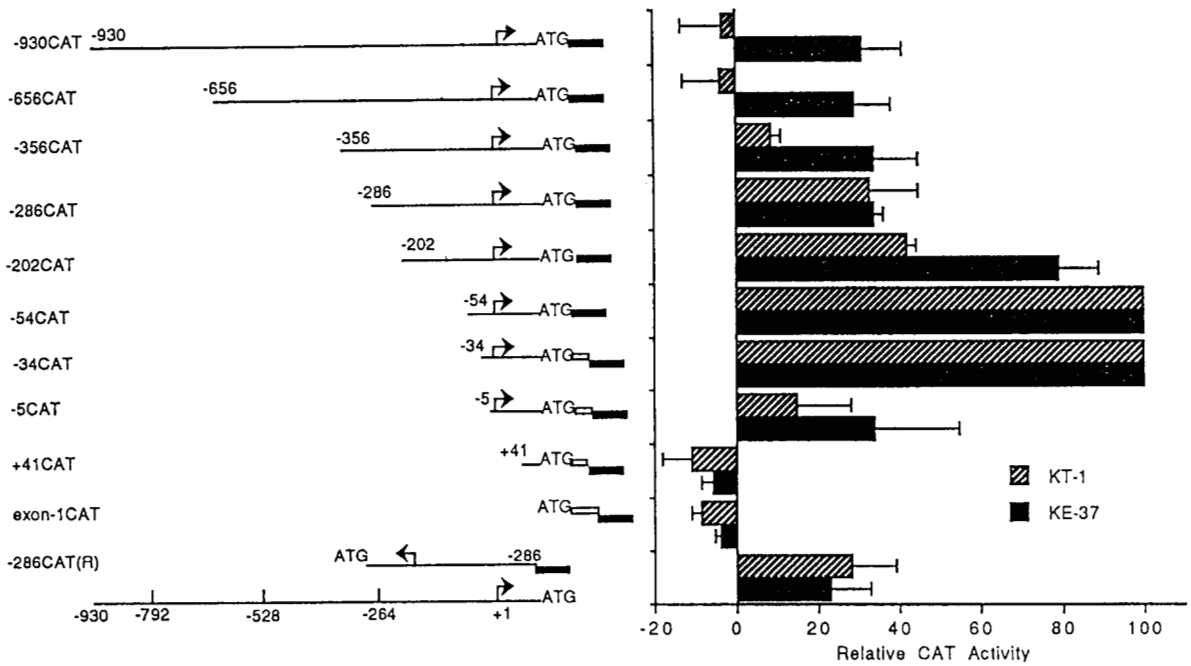

FIG. 3. **Functional analysis of the 5′-flanking sequence of the human TdT gene.** The *line diagram* on the left of the figure represents the deletion TdT constructs used in transient transfection assays. *Open boxes* represent exon-1 sequence, and *solid boxes* represent CAT coding sequences. Nucleotide positions with respect to +1 as the transcription start site are indicated above the line. CAT and β-galactosidase enzyme activities were measured in KT-1 and KE-37 cell lines transfected with 35 μg of the TdT deletion construct DNAs and 5 μg of pCMVβ DNA as described under "Materials and Methods." The CAT activity of the −34 CAT construct was assigned a value of 100%. This represents 400 pmol of butyrylated chloramphenicol in KT-1 cells and 500 pmol in KE-37 cells. Data are mean values from four to seven experiments. At least two different plasmid preparations were used for each transfection assay.

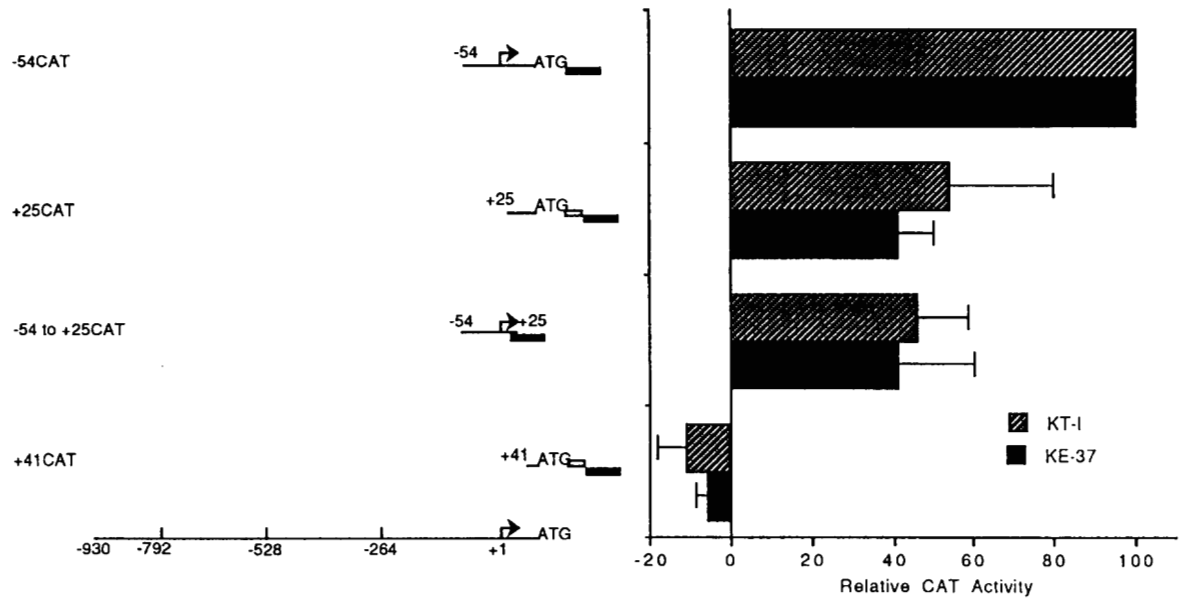

FIG. 4. **Functional analysis of sequences downstream of the transcription start site.** The *line diagram* on the left of the figure represents the deletion TdT constructs used in transient transfection assays. *Open boxes* represent exon-1 sequence, and *solid boxes* represent CAT coding sequences. Nucleotide positions with respect to +1 as the transcription start site are indicated above the line. Shown are the relative CAT activities of −54CAT, +25CAT, −54 to +25CAT, and +41CAT in KT-1 and KE-37 cells. Transfection procedures were carried out as described in Fig. 3 and under "Materials and Methods."

gested that all three segments of the promoter (the primary element within −34 bp, the element spanning the transcription start site and the downstream element) were necessary for optimal promoter activity.

DISCUSSION

An important step in identification of promoter elements is an accurate positioning of transcription start. Previously, Koiwai and Morita (1988) had identified two transcription start sites for the human TdT gene, at 80 and at 106 bp upstream from the ATG. However, our *in vivo* experiments (5′-RACE technique) consistently identified a single start site at 62 bp upstream from translation start. While the reasons for the discrepancy are not clear the transfection experiments support the site we identified.

The human TdT gene lacks TATA and CAAT consensus sequences immediately 5′ to the initiation site, a characteristic that is shared by other eukaryotic genes. The 5′-flanking se-

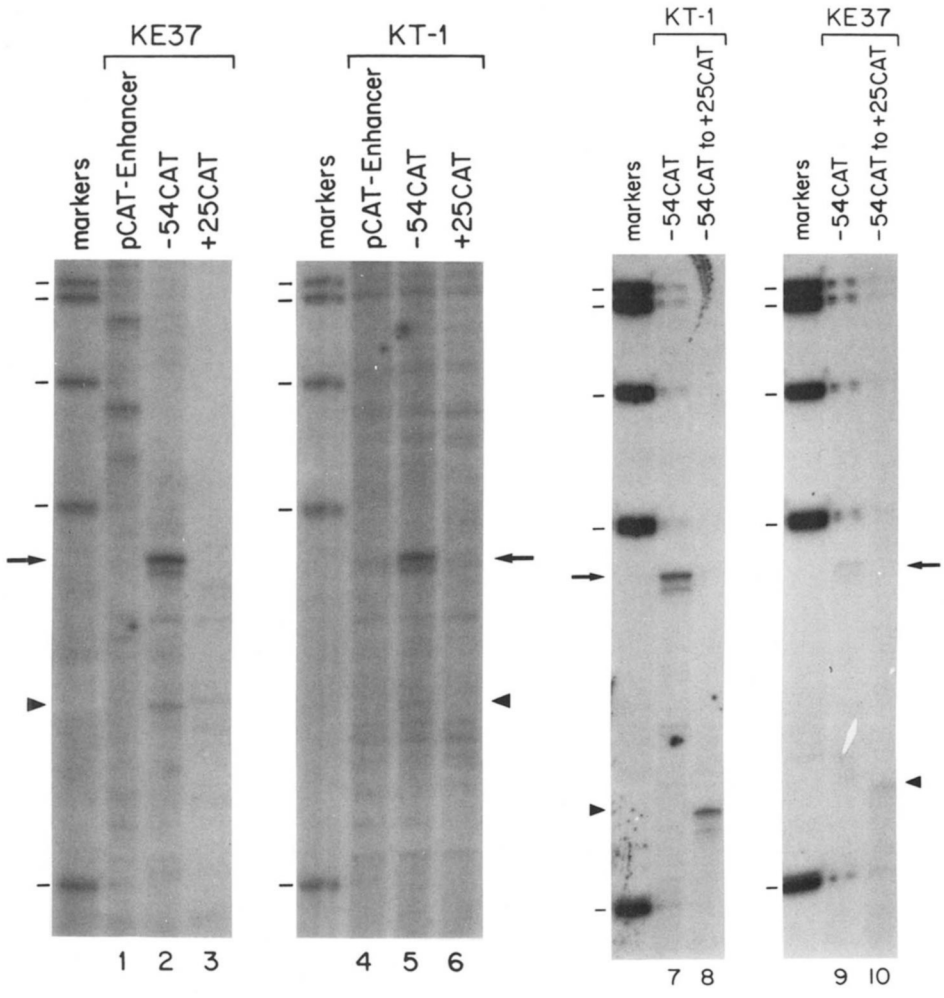

FIG. 5. **Primer extension analysis of RNA derived from TdT-CAT DNA-transfected cells.** Primer extension reactions were performed on 50 μg of total RNA isolated from KE-37 and KT-1 cells transfected with 35 μg of the indicated plasmid DNA. 5′-end-labeled oligonucleotide complementary to the CAT coding region ("Materials and Methods") was used for primer extension. The cDNA reaction products were electrophoresed on a denaturing gel, and the products were visualized by autoradiography. In *lanes 1–6*, the *upper arrow* indicates the predicted +1 transcription start site, and the *lower arrow* indicates a minor transcription start site. The *control lane* contains the primer extension data with 50 μg of RNA isolated from cells transfected with pCAT-Enhancer. *Lanes 7–10* were from another experiment. The *upper arrow* and *lower arrow* in *lanes 7–10* both indicate the predicated +1 transcription start site. The cDNA products in *lanes 8* and *10* were 40 bp shorter than those in *lanes 7* and *9* because of the deletion of sequences from +25 to ATG during cloning of the −54 to +25CAT construct. *Hae*III-digested ØX174 size markers (^{32}P-labeled) are shown on the left side of each panel. The *dashes*, –, on the left of the figure correspond to the bands in the marker lane and represent 281, 271, 234, 194, and 118 bases, respectively.

quences of the human and mouse TdT genes share limited homology. The mouse TdT gene was difficult to study because transcription rates of this gene were very low (Smale and Baltimore, 1989; Lo *et al.*, 1991). No CAT activity was detected when TdTCAT fusion plasmids were used for transfections of mouse lymphoma cells (Lo *et al.*, 1991). Therefore, to identify regulatory elements, Smale and co-workers resorted to the use of replicating vectors in transient transfection assays (Lo *et al.*, 1991; Ernst *et al.*, 1993). However, the authors of that study suggested replicating vectors probably masked all but the strongest control elements. We initially confronted a similar problem in our study of the human TdT gene. Transient transfection assays using CAT reporter vectors fused to different regions of the human TdT gene were difficult to establish. However, optimizing conditions for electroporation helped to increase the transfection efficiency and made these experiments feasible.

With transient transfection assays of immature lymphoid cells firmly established, we tested a series of TdTCAT fusion genes in transfection assays. The optimal levels of CAT activity were consistently observed when the sequence spanning −54/−34 to ATG was used in transfection assays. Further analyses identified the basal promoter of the human TdT gene within the region −34 to +40 bp. Moreover, three distinct control elements in this region were identified. The first control element was located 34 bp upstream from the transcription start site, but we have not yet determined its precise boundaries. Negative regulation was also evident upstream from −54 bp. This transcription inhibition effect was far more prominent in TdT-positive than in TdT-negative cell lines. In fact, when the −930CAT and −656CAT DNA were used to transfect KT-1 cells, virtually no CAT activity was detected. The same constructs yielded 25 times more CAT activity in TdT-negative than in TdT-positive cells. With the construct −356CAT, three times more CAT activity was detected in TdT-negative than in TdT-positive cells. These results suggest that a negative regulatory element exists in the region −356 to −656 and that it is differentially active in the TdT-positive cells. Inspection of the 5′ sequence for putative transcription factor-binding sites identified potential AP1, INF-1, NF-1, and Oct-1 elements in this region (−356 to −656). Any one of these or a different protein binding within this region could possibly exert a negative effect on TdT transcription in TdT-positive cells. The sequences between −54 to −356 also exerted negative regulation, and the effect was roughly equivalent in both KT-1 and KE-37 cells. A variety of negative elements, in different gene promoters has been characterized (Baniahmad *et al.*, 1987; Wang and Brand, 1990; Rincon-Limas *et al.*, 1991). The exact position and nature of the negative elements in human TdT 5′ sequences will be the subject of further investigation.

The second of the three elements within the human basal promoter was found to span the transcription start site. When a construct was used that contained only 5 bp upstream from the transcription start site along with downstream sequences to the ATG, CAT activity was readily detected although at a reduced level. Furthermore, no specific transcripts were observed in primer extension experiments when sequences both upstream and around the start site were removed. The fact that this construct (−5CAT) was able to drive CAT activity, and that transcription *in vivo* initiated at a single nucleotide in the

absence of a TATA sequence indicated that an Inr type element was present in the human promoter sequence (Weis and Reinberg, 1992; Kollmar and Farnham, 1993). There was no visible homology between the human TdT initiation site and the mouse TdT Inr element, but we did observe a similarity to an 8-bp Inr consensus sequence (GCAGTCTC) described by Kollmar and Farnham (1993). In order to match the Inr homology to our human TdT sequence, we had to place the start site at an A (as described in Kollmar and Farnham, 1993) two nucleotides upstream from the T initiation site detected by 5′-RACE.

The third promoter element spanned nucleotides +25 to +40. This element alone was not capable of initiating transcription correctly. Primer extension analyses of the transcription products indicated that initiation was no longer at a single nucleotide when only the third promoter element was contained in the constructs. However, removal of this sequence in the CAT constructs resulted in a 2-fold reduction in CAT activity in the transfection assay when compared with the complete basal promoter. This reduction was significant since the decrease represented a drop from about 500 to 250 pmol of butyrylated product. There are several examples of downstream regulatory elements in other genes. The SV40 major late promoter requires for correct initiation of transcription, sequences centered 28 bp downstream of the CAP site (Ayer and Dynan, 1988). In the HIV promoter sequences spanning +21 to +30 enhances transcription (Zenzie-Gregory *et al.*, 1993). The exact mechanism by which the downstream element in the human TdT promoter affects transcription is still not understood, but is under investigation.

Another feature of the human TdT basal promoter was its bidirectional activity. While not common, bidirectional activity of promoters is not without precedence (Abrams and Schimke, 1989; Rincon-Limas *et al.*, 1991).

The proteins responsible for basal promoter activity of the human TdT gene have not yet been clearly identified. Gel mobility shift assays and DNase I footprinting analyses with various regions of the basal promoter have detected specific protein binding to the sequences −27 to −17, −17 to −6 and +25 to +40 (data not shown). Further experiments are underway to identify these proteins and to assess exactly how they function in transcription of the human TdT gene.

In summary, we have demonstrated that the human TdT gene contains a basal promoter that functions comparably in both TdT expressing and non-expressing cells. The basal promoter of the human TdT gene is not identical to that found in the murine TdT gene. The latter seems to consist only of an Inr element. In addition, an LyF-1 and an Ets sequence upstream from the Inr seems to provide lymphoid specificity to the mouse TdT gene. In contrast, the human TdT basal promoter consists of three elements, including a consensus Inr element. While LyF-1- and Ets-binding sites are present upstream from the human basal promoter (Lo *et al.*, 1991; Ernst *et al.*, 1993), participation of these two proteins in lymphoid-specific transcription of the human TdT gene is not apparent from the experiments we have performed since these sequences had no detectable effect on CAT activity.

Studies of the boundaries of the basal promoter, negative regulatory elements, and tissue-specific enhancer elements in other regions of the human gene are currently underway in our laboratory. These results will help us understand transcriptional regulation of the human TdT gene and will help define the various mechanisms by which tissue- and differentiation-specific gene expression is controlled.

Acknowledgments—We are grateful to Dr. Jenny Ting and members of her laboratory for helping with the transfection experiments. Acknowledgment is due to Dr. Bruce Aronow (University of Cincinnati) for suggesting the 5′-RACE experiment. Eloise Parrish provided excellent technical assistance and Drs. Beverly Mitchell and Judith Lesnaw were of tremendous help in their critiques of the manuscript.

REFERENCES

Abrams, J. M., and Schimke, R. T. (1989) *Mol. Cell. Biol.* **9,** 620–628
Ayer, D. E., and Dynan, W. S. (1988) *Mol. Cell. Biol.* **8,** 2021–2033
Baniahmad, A., Muller, M., Stener, C., and Renkawitz, R. (1987) *EMBO J.* **6,** 2297–2303
Bollum, F. J. (1974) *The Enzymes* (Boyer, P. D., ed) pp. 145–171, Academic Press, New York
Crab, D. W., and Dixon, J. E. (1987) *Anal. Biochem.* **163,** 88–92
Deibel, M. R., Riley, L. K., Coleman, M. S., Cibull, M. L., Fuller, S. A., and Todd, E. (1983) *J. Immunol.* **131,** 195–200
Desiderio, S. V., Yancopolous, G. D., Paskind, M., Thomas, E., Boss, M. A., Landau, N., Alt, F., and Baltimore, D. (1984) *Nature* **311,** 752–755
Drexler, H. G., Gaedicke, G., and Minowada, J. (1985) *Leukemia Res.* **9,** 209–229
Ernst, P., Hahm, K., and Smale, S. T. (1993) *Mol. Cell. Biol.* **13,** 2982–2992
Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., and Mathis, D. (1993) *Science* **261,** 1175–1178
Ikuta, K., Uchida, N., Friedman, J., and Weissman, I. L. (1992) *Annu. Rev. Immunol.* **10,** 759–783
Koiwai, O., and Morita, A. (1988) *Biochem. Biophy. Res. Commun.* **154,** 91–100
Kollmar, R., and Farnham, P. J. (1993) *Proc. Soc. Exp. Biol. Med.* **203,** 127–139
Komori, T., Okada, A., Stewart, V., and Alt, F. W. (1993) *Science* **261,** 1171–1175
Lo, K., Landau, N. R., and Smale, S. T. (1991) *Mol. Cell. Biol.* **13,** 5229–5243
Matsushima, G. K., Lindstrom, Y.-I., and Ting, J.-P. Y. (1992) *Mol. Cell. Biol.* **12,** 5610–5619
Nakajima, J. W., Horikoshi, M., and Roeder, R. G. (1988) *Mol. Cell. Biol.* **8,** 4028–4040
Pugh, B. F., and Tjian, R. (1992) *J. Biol. Chem.* **267,** 679–682
Riley, L. K., Morrow, J. K., Danton, M. J., and Coleman, M. S. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2489–2493
Rincon-Limas, D. E., Krueger, D. A., and Patel, P. I. (1991) *Mol. Cell. Biol.* **11,** 4157–4164
Roeder, R. G. (1991) *Trends. Biol. Chem. Sci.* **16,** 402–407
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Sawadogo, M., and Sentenac, A. (1990) *Annu. Rev. Biochem.* **59,** 711–754
Schatz, D. G., and Baltimore, D. (1988) *Cell* **53,** 107–115
Seed, B., and Sheen, J. Y. (1988) *Gene (Amst.)* **67,** 271–277
Smale, S. T., and Baltimore, D. (1989) *Cell* **57,** 103–113
Trangas, T., and Coleman, M. S. (1989) *Biochem. Biophy. Res. Commun.* **164,** 750–757
Wang, T. C., and Brand, S. J. (1990) *J. Biol. Chem.* **265,** 8908–8914
Weis, L., and Reinberg, D. (1992) *FASEB J.* **6,** 3300–3309
Zenzie-Gregory, B., O'Shea-Greenfield, A., and Smale, S. T. (1992) *J. Biol. Chem.* **267,** 2823–2830
Zenzie-Gregory, B., Sheridan, P., Jones, K. A., and Smale, S. T. (1993) *J. Biol. Chem.* **268,** 15823–15832