Takahashi, H., Komano, H., Kawaguchi, N., Kitamura, N., Nakanishi, S., & Natori, S. (1985) *J. Biol. Chem. 260*, 12228–12233.
Takahashi, H., Komano, H., & Natori, S. (1986) *J. Insect Physiol. 32*, 771–779.
Takio, K., Kominami, E., Wakamatsu, N., Katsunuma, N., & Titani, K. (1983) *Biochem. Biophys. Res. Commun. 115*, 902–908.
Takio, K., Kominami, E., Bando, Y., Katsunuma, N., & Titani, K. (1984) *Biochem. Biophys. Res. Commun. 121*, 149–154.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Truman, J. W. (1987) in *Current Topics in Developmental Biology* (Moscona, A. A., & Monroy, A., Eds.) Vol. 21, pp 99–116, Academic Press, New York.
Weinstock, R., Sweet, R., Weiss, M., Cedar, H., & Axel, R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1299–1303.

Complete Sequence of the Human Tissue Factor Gene, a Highly Regulated Cellular Receptor That Initiates the Coagulation Protease Cascade†,‡

Nigel Mackman, James H. Morrissey,* Bruce Fowler, and Thomas S. Edgington
Department of Immunology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037
Received September 6, 1988; Revised Manuscript Received October 12, 1988

ABSTRACT: Tissue factor (TF) is the high-affinity receptor for plasma factors VII and VIIa. TF plays a role in normal hemostasis by initiating the cell-surface assembly and propagation of the coagulation protease cascade. Outside the vasculature, TF expression is highly dependent upon cell type. TF can also be induced by inflammatory mediators to appear on monocytes and vascular endothelial cells as a component of cellular immune responses. As an initial step toward elucidating the regulatory regions involved in control of TF gene expression, we have established the organization of the 12.4 kbp human TF gene and its complete DNA sequence. There are six exons separated by five introns. Within intron 5, we have mapped the single nucleotide difference which leads to the previously described *Msp*I polymorphism; the same intron also contains an apparently polymorphic *Pst*I site. The TF gene also contains three full-length *Alu* repeats and one partial *Alu* repeat. A single major transcription start site was identified 26 bp downstream from a TATA consensus promoter element. The putative promoter and first exon are located within a 1.2 kbp region of very high G + C content which fits the criteria of an HTF island. A cluster of predicted binding sites for a number of known transcription factors was found to coincide with this putative promoter region. These factors included AP-1 and AP-2 which can mediate the effects of phorbol esters, agonists known to induce TF expression in monocytes and vascular endothelial cells.

Tissue factor (TF), a 47-kDa transmembrane glycoprotein, is the major cellular initiator of the coagulation protease cascade. TF functions in normal hemostasis as the high-affinity, cell-surface receptor and essential cofactor for factors VII and VIIa (Broze, 1982; Bach et al., 1986), thereby triggering the cell-surface assembly of a cascade of highly specific serine proteases. TF belongs to a unique class of highly regulated receptor proteins, including thrombomodulin (Esmon, 1987), that act as effector molecules to modulate the extracellular environment.

As expected, TF is not normally expressed at significant levels by cells within the vasculature. However, in a variety of inflammatory settings, TF can be induced to appear on cells of monocyte lineage, the only circulating cell type observed to express TF (Edwards et al., 1979; Levy et al., 1981; Gregory & Edgington, 1985). TF expression can also be induced in vascular endothelial cells by the inflammatory cytokines interleukin 1 (IL-1) and tumor necrosis factor (TNFα), as well as by other inducers such as endotoxin (Bevilacqua et al., 1984; Stern et al., 1985). Induction of TF in monocytes and vascular endothelial cells is thought to be an important component of cellular immune responses [reviewed by Ryan and Geczy (1987)] and is also implicated in the pathogenesis of disseminated intravascular coagulation and septic shock (Edwards & Rickles, 1984; Taylor et al., 1987). In the monocyte, the primary level of regulation of TF expression appears to be transcriptional; endotoxin stimulation of human monocytes results in transient, de novo transcription of the TF gene (Gregory et al., 1988) in a manner similar to the regulation of TNFα (Sariban et al., 1988).

In nonvascular cells, TF expression appears to be regulated by other mechanisms. In cell culture, for instance, TF activity ranges from undetectable levels in some cell types to relatively high levels in other cell types such as fibroblasts (Dvorak et al., 1983; Rodgers et al., 1984). Immunohistochemical localization of TF in tissues has confirmed the cell type specificity of TF expression in vivo (Drake et al., 1988).

The relatively complex pattern of TF expression in different cell types is likely to involve multiple tissue-specific and agonist-responsive cis-acting regulatory elements linked to the TF gene. Detailed analysis of these elements requires

† This is Publication No. 5394-IMM and was supported in part by NIH Grants HL16411 and CA41085. This work was done during the tenure of an Established Investigatorship (J.H.M.) of the American Heart Association and during the tenure of a research fellowship (N.M.) from the American Heart Association, California Affiliate, and with funds contributed by the San Diego County Chapter.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02846.
* To whom correspondence should be addressed.

0006-2960/89/0428-1755$01.50/0

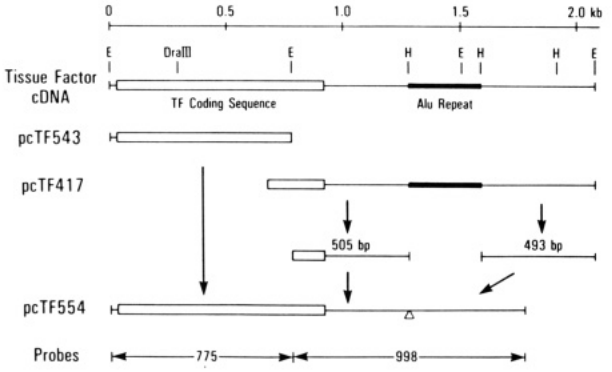


FIGURE 1: Construction of a cDNA for TF containing the complete coding sequence. pcTF543 was made in pUC19 by joining fragments derived from λcTF23 and λcTF2 (Morrissey et al., 1987) at the *Dra*III site. An additional partial TF cDNA clone, pcTF417, was obtained from a GM1380 (normal human fibroblast) cDNA library (Fakhrai, Morrissey, and Edgington, unpublished results). Complete digestion of pcTF417 allowed the isolation of the 505 bp *Eco*RI–*Hin*dIII fragment, and partial digestion allowed the isolation of the 493 bp *Hin*dIII–*Eco*RI fragment. Finally, pcTF554 was constructed in pUC19 in two steps by first ligating the 775 and 505 bp fragments together to form an intermediate plasmid, pCTF553, and then adding the 493 bp fragment. Complete digestion of pCTF554 with *Eco*RI released fragments of 998 and 775 bp in size, which were used as hybridization probes in this study. The numbering of the nucleotides corresponds to that used by Morrissey et al. (1987). The restriction endonuclease sites are as follows: E, *Eco*RI; H, *Hin*dIII. The triangle under the line for pcTF554 represents the point of deletion of the *Alu* repeat.

knowledge of the organization of the gene and promoter region. We now describe the isolation of genomic DNA clones spanning the human TF gene and its complete DNA sequence.

Materials and Methods

cDNA Probes. The 998 bp and 775 bp *Eco*RI fragments from pcTF554 (Figure 1) were used as probes; they encompassed the majority of the TF cDNA (less the *Alu* repeat in the 3′ untranslated region). The probes were labeled with [α-^{32}P]ATP using a random-primer method (Boehringer Manheim).

λ Libraries and Screening. Two human chromosome 1 libraries were used, the first constructed by complete digestion with *Eco*RI (ATCC 57738; American Type Culture Collection, Rockville, MD) and the second by complete digestion with *Hin*dIII (ATCC 57754). In addition, two human genomic libraries were obtained, the first constructed by partial *Eco*RI digestion (ATCC 37385) and the second by partial *Mbo*I digestion (HL1006d; Clontech Laboratories, Palo Alto, CA). From each library, 5×10^5 plaques were screened (3–10 genomic equivalents) by using a modification of the method by Manitatis et al. (1982) in which 1% SDS, 100 μM ATP, and 10 mM sodium pyrophosphate were added to the prehybridization solution. Hybridization and wash conditions were as described previously (Bahnak et al., 1987), with the addition of 0.5% SDS to hybridization and wash solutions.

DNA Sequencing. DNA fragments were subcloned into M13 vectors for subsequent nucleotide sequencing by the dideoxy method using [α-^{35}S]dATP and Sequenase (United States Biochemical, Cleveland, OH). The DNA sequence was determined from both strands for the cDNA clone pcTF417 (Figure 1), all of the 5′ portion of the TF gene, and the entire HTF island, which contained a high G + C content. All areas of discrepancy with previously reported TF cDNA sequences were determined by sequencing from both strands with and without the use of deoxyinosine triphosphate to minimize gel compressions. Fifty-two percent of the intron DNA sequence

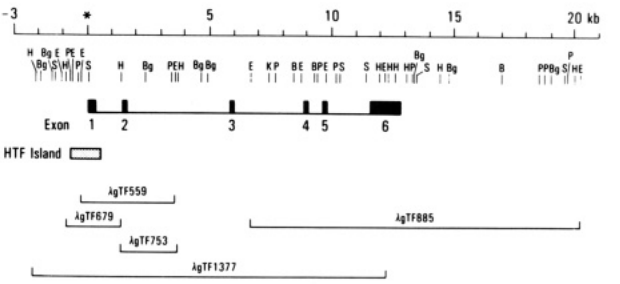


FIGURE 2: Organization of the human TF gene. λgTF559 was isolated from a human chromosome 1 library (ATCC 57738) and contained a 3.9 kbp *Eco*RI insert. λgTF679 and λgTF753 were isolated from a human chromosome 1 library (ATCC 57754) and contained 2.2 and 2.3 kbp *Hin*dIII inserts, respectively. λgTF885 was isolated from a human genomic library (ATCC 37385) and contained a 13.4 kbp insert. λgTF1377 contained a 14.0 kbp insert and was isolated from a human genomic library (Clontech HL1006d). This figure shows the position of exons 1–6. The restriction sites are as follows: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; S, *Sac*I. The asterisk represents the start of transcription. The HTF island is indicated by the stippled box.

was determined from both strands.

DNA Hybridization Analysis. The fetal human fibroblast cell line GM1381 (Human Genetic Mutant Cell Repository; Corielle Institute, Camden, NJ) cultured in DMEM supplemented with 10% fetal calf serum was used for isolation of DNA (Maniatis et al., 1982). Agarose gels were electrophoresed in TAE buffer with 0.5 μg/mL ethidium bromide. DNA samples were blotted onto GeneScreen Plus (New England Nuclear) and cross-linked by UV irradiation before being dried. Hybridization analyses were as above.

Primer Extensions. Poly (A+) RNA was isolated by oligo(dT) chromatography from total cellular RNA purified by the guanidine thiocyanate/CsCl method (Maniatis et al., 1982). The sequence of the 20-mer oligonucleotide primer, 5′-GGCACAGTATTTGTAGTGCC-3′, is complementary to nucleotides 131–152 of the cDNA sequence (Morrissey et al., 1987). This corresponds to the last nucleotide of exon 1 and the first 19 nucleotides of exon 2. The primer was phosphorylated with [γ-^{32}P]ATP (Amersham) using T4 polynucleotide kinase to a specific activity of 2.6×10^8 cpm/μg. Primer extension was performed according to Bonthron et al. (1988) with several modifications. The hybridization reaction contained the following in a final volume of 12 μL: 4 μg of poly (A+) RNA (or 10 μg of yeast tRNA), 1×10^6 cpm of oligonucleotide primer, 100 mM KCl, 25 mM Tris-HCl, pH 8.3. The reaction mixture was heated to 70 °C for 15 min and then incubated at 37 °C for 1 h. Extension was carried out in a final volume of 40 μL containing 30 mM KCl, 50 mM Tris-HCl, pH 8.3, 5 mM DTT, 50 μg/mL actinomycin D, 1.25 mM each dNTP (A, C, T, and G), 100 units/mL RNasin (Promega), 8 mM $MgCl_2$, and 1250 units/mL reverse transcriptase (Life Sciences) and incubated at 42 °C for 30 min. The reaction was stopped by addition of 1.6 μL of 0.5 M EDTA, extracted once with phenol/chloroform, and ethanol precipitated.

Results and Discussion

Identification of Genomic Clones Spanning the Human TF Gene. We screened two human chromosome 1 libraries with a TF cDNA probe on the basis of localization of the TF gene to this chromosome (Carson et al., 1985; Scarpati et al., 1987b). Several positive clones were identified, sequenced, and found to contain the first two exons of the TF gene together with the 5′ flanking sequence (λgTF559, λgTF679, and λgTF753; see Figure 2). Attempts to identify additional

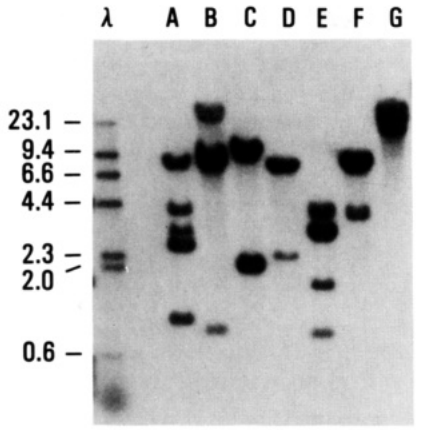


FIGURE 3: Southern blot analysis of the human TF gene. To confirm the presence of a single gene for TF in humans and to ensure that no rearrangment of the DNA had occurred in the formation of the clones, human DNA was probed with the 775 and 998 bp *Eco*RI fragments of the TF cDNA (see Materials and Methods). Six micrograms of GM1381 fibroblast DNA, digested to completion with restriction endonuclease, was used per lane. Lane A, *Eco*RI; lane B, *Bam*HI; lane C, *Sac*I; lane D, *Hin*dIII; lane E, *Pst*I; lane F, *Bgl*II; lane G, *Kpn*I. The sizes of the markers are shown in kilobase pairs. The autoradiogram was exposed for 3 days.

positive clones from these libraries were unsuccessful, leading to a screening of partial *Eco*RI and partial *Mbo*I human libraries which contained larger inserts (circa 16 kbp). Clone λgTF885 extended from the middle to the 3′ end of the TF gene and included exons 4, 5, and 6, whereas clone λgTF1377 contained exons 1–5 and part of exon 6.

The inserts from the positive clones were ligated into pUC19, and the approximate position of the exons was initially established by DNA hybridization using a cDNA probe. A restriction map of the TF gene for seven restriction endonuclease enzymes is shown in Figure 2.

Organization of the Tissue Factor Gene. Previous studies suggested a single gene for TF in humans because a 641 bp partial cDNA probe detected a unique band in a *Sac*I digest of genomic DNA (Morrissey et al., 1987), and the gene had a unique chromosome location (Carson et al., 1985; Scarpati et al., 1987b). Results from DNA hybridization analysis of human fibroblast DNA probed with the nearly full-length TF cDNA (Figure 3) established that no rearrangement of the TF gene had occurred in the formation of these clones; i.e., the pattern of hybridizing bands of genomic DNA was identical with that predicted from the restriction map of the clones (Figure 2). Therefore, the data are accounted for as having arisen from a single gene for TF in humans, and there was no evidence for cross-hybridizing pseudogenes.

The human TF gene spans 12.4 kbp of chromosome 1 DNA and is organized into six exons separated by five introns. The complete sequence of the TF gene, together with 798 bp upstream of the initiation of transcription and 633 bp downstream of the polyadenylation site, is shown in Figure 4. The location and size of each exon and intron are listed in Table I. Exon 1 encodes the N-terminal signal peptide which is removed during processing of the protein. The amino acid sequence of TF contains three repeats of an uncommon tripeptide tryptophan-lysine-serine (WKS), previously hypothesized as a sequence motif that may be associated with structural loci that function to recognize serine proteases (Morrissey et al., 1987). Each of the WKS motifs is encoded by a separate exon (exons 2, 3, and 4). The different locations of the motifs within the respective exons and analysis of the DNA sequence surrounding each WKS motif revealed no evidence that the three

Table I: Location and Size of Exons and Introns in the Human TF Gene

exon	nucleotide positions	length (bp)	amino acids
1	1–223	223	−32 to +1
2	1392–1503	112	2 to 39
3	5594–5793	200	40 to 105
4	8491–8669	179	106 to 165
5	9277–9436	160	166 to 218
6	11157–12434	1278	219 to 263

intron	nucleotide positions	length (bp)	type[a]	no. of *Alu* repeats
1	224–1391	1168	I	0
2	1504–5593	4090	II	1 (partial)
3	5794–8490	2697	I	1
4	8670–9276	607	O	0
5	9437–11156	1720	I	1

[a] Intron type is according to Sharp (1981); O indicates a splice between codons, I indicates a splice after the first nucleotide of a codon, and II indicates that the splice site occurs after the second nucleotide.

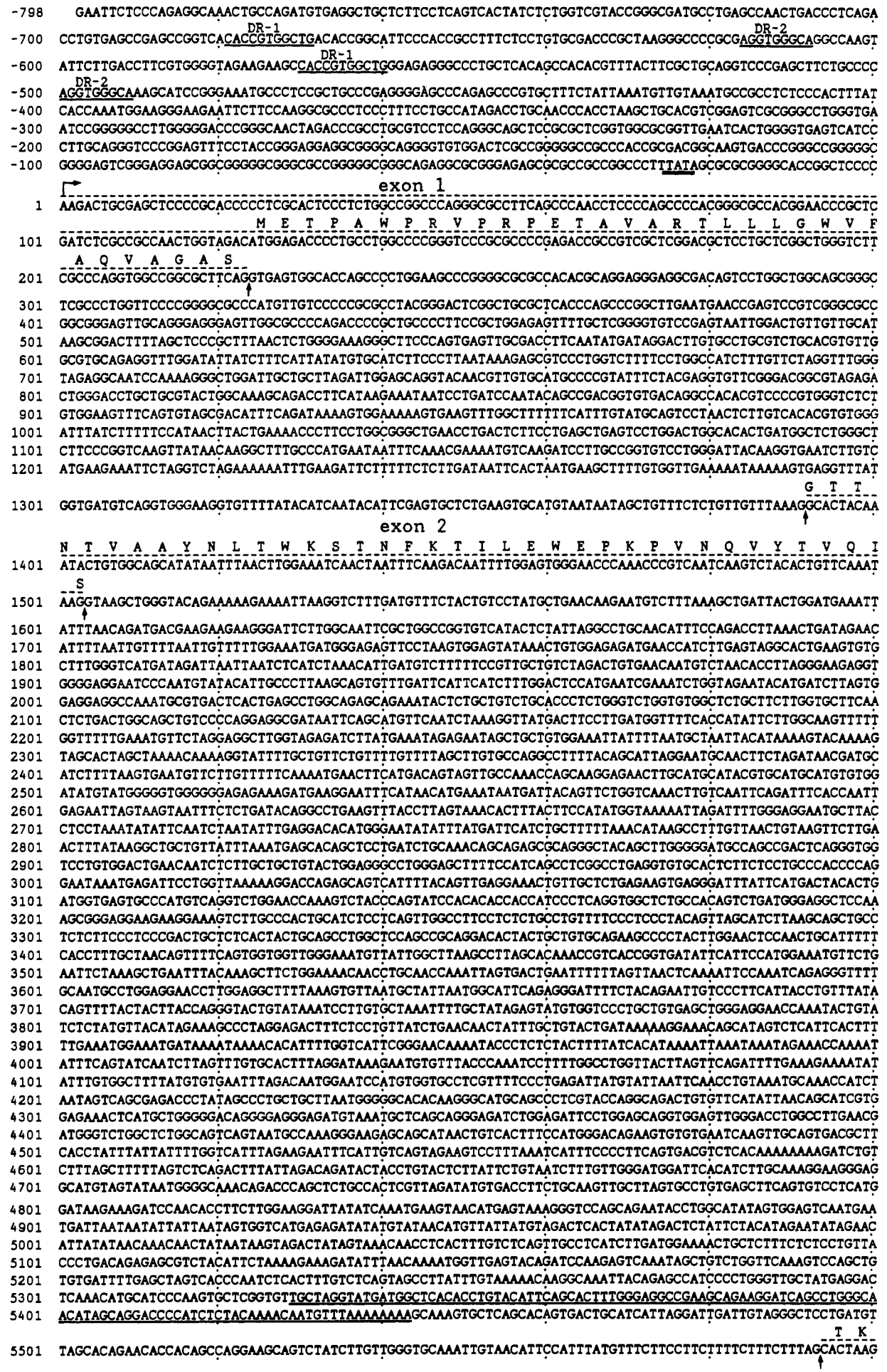
Table II: Splice Junction Sequences

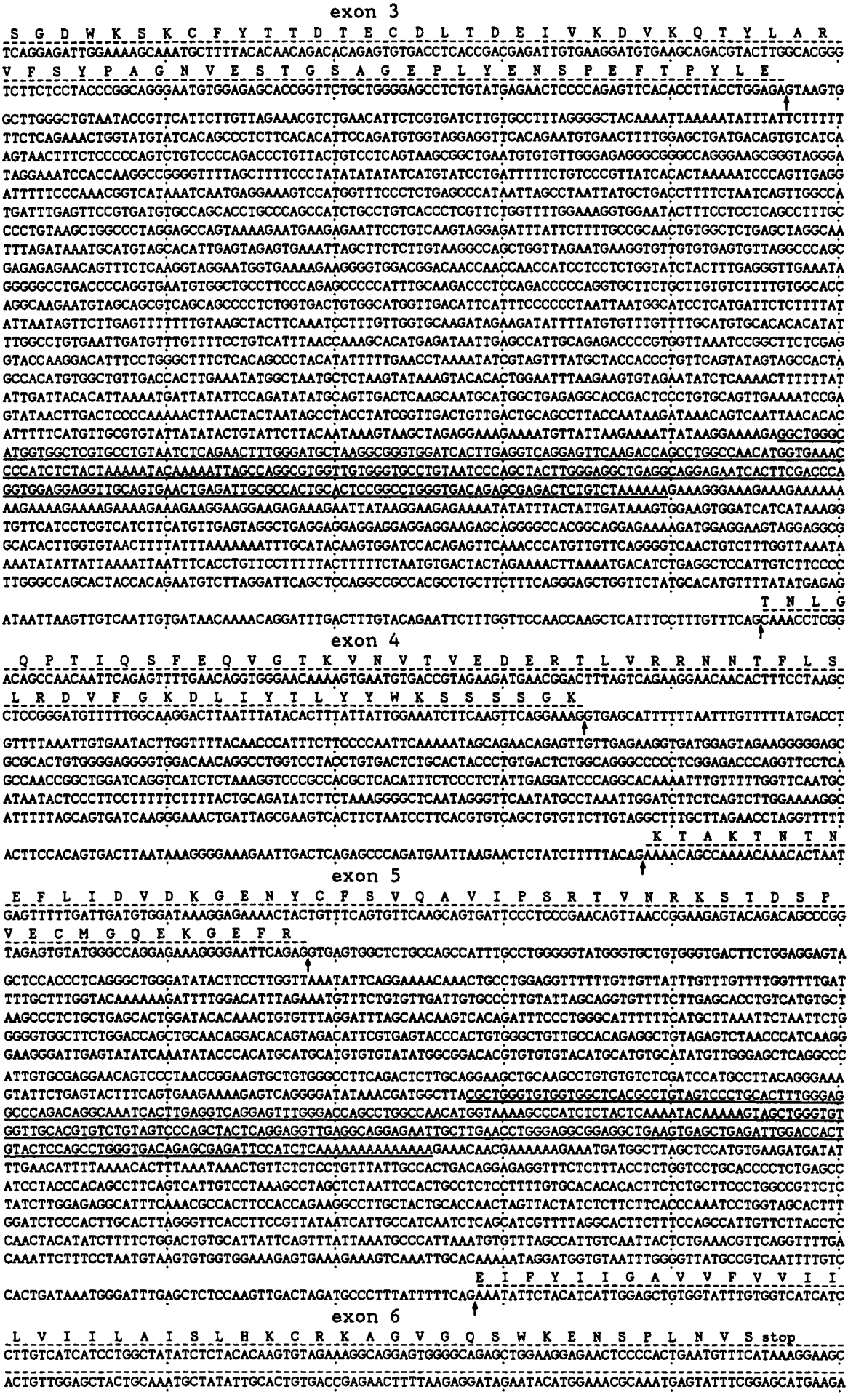
Intron	Exon	Intron	Exon
1	TCAG	GTGAGT.....CTCTGTTGTTTAAAG	GCAC
2	TAAG	GTAAGC.....CTTTTCTTTCTTTAG	CACT
3	GAGA	GTAAGT.....TTTCCTTTGTTTCAG	CAAA
4	AAAG	GTGAGC.....TCTATCTTTTTACAG	AAAA
5	AGAG	GTGAGT.....CCCTTTATTTTTCAG	AAAT
Consensus Sequence (from Mount, 1982)	C AG A	GT AGT..... TTTTTTTTTTT T N AG CCCCCCCCCCC C	G

exons containing the WKS repeats have arisen by tandem duplication. Exons 1–5 have an average length of 175 bp, which is similar to the average length reported for genes of higher eukaryotes (Naora & Deacon, 1982). In contrast, exon 6 contains 1278 bp, which corresponds to over half the entire mRNA. It encodes both the transmembrane and cytoplasmic domains of TF and all of the extensive 3′ untranslated region.

The five introns each begin with a GT dinucleotide and end with an AG, sequences thought to be necessary for correct RNA splicing (Breathnach & Chambon, 1981). The intron–exon splice junction sequences are shown in Table II and agree closely with the consensus sequence from Mount (1982). The organization of the TF gene presented confirms the location of two of the three intron/exon boundaries deduced previously from the sequence of cDNA clones which were thought to be derived from partially processed mRNAs (Scarpati et al., 1987b). However, the splice junction between exon 1 and exon 2 disagrees by one nucleotide from that proposed by Scarpati et al. (1987b).

Properties of the TF Transcript. The start of transcription was mapped by primer extension (Figure 5) using poly(A+) RNA from GM1381 cells, a normal fetal fibroblast cell line constitutively expressing TF (Morrissey et al., 1987). The major start site was an adenine residue identified in Figure 4 as position +1. A faint band was also observed by primer extension which may indicate a minor start site located 3 bp downstream. A TATA consensus promoter element is located 26 bp upstream from the cap site (Breathnach & Chambon, 1981), although no CCAAT box was found. Therefore, the complete 5′ untranslated region of TF mRNA is 123 bp in length. The single AUG triplet in exon 1 confirms the previous


DR-1
DR-2
DR-1
DR-2
exon 1
M E T P A W P R V P R P E T A V A R T L L L G W V F
A Q V A G A S
exon 2
N T V A A Y N L T W K S T N F K T I L E W E P K P V N Q V Y T V Q I
S
G T T
T K


exon 3
S G D W K S K C F Y T T D T E C D L T D E I V K D V K Q T Y L A R
V F S Y P A G N V E S T G S A G E P L Y E N S P E F T P Y L E
exon 4
Q P T I Q S F E Q V G T K V N V T V E D E R T L V R R N N T F L S
L R D V F G K D L I Y T L Y Y W K S S S S G K
T N L G
K T A K T N T N
exon 5
E F L I D V D K G E N Y C F S V Q A V I P S R T V N R K S T D S P
V E C M G Q E K G E F R
E I F Y I I G A V V F V V I I
exon 6
L V I I L A I S L H K C R K A G V G Q S W K E N S P L N V S stop

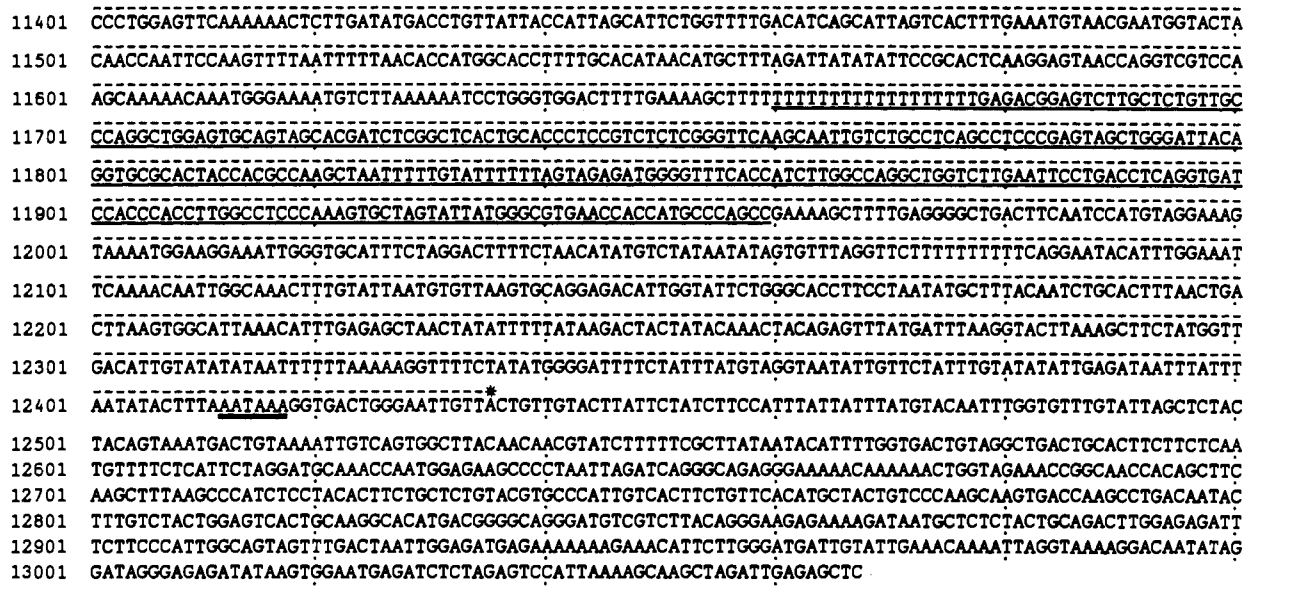


FIGURE 4: Complete nucleotide sequence of the human TF gene. The complete sequence of the TF gene is shown, together with 798 bp upstream of the cap site (↱) and 633 bp downstream of the polyadenylation site. The nucleotides are numbered using the major transcription start site as +1. The TATA box and polyadenylation signal are indicated by bold underlining. The polyadenylation site is represented by an asterisk. The exons are numbered, and their extent is shown by a dashed line above the nucleotide sequence. The coding region has been translated by using the single-letter amino acid code. The symbol (↑) designates the location of the intron/exon boundaries. The three full-length *Alu* repeats and one partial *Alu* repeat are underlined. In the region upstream of the cap site, the 11 and 9 bp direct repeats are indicated as DR-1 and DR-2, respectively.

assignment of the initiation of translation. The predicted size of the mature TF transcript is in good agreement with the major 2.3 kb mRNA identified in RNA blots from several cell types (Morrissey et al., 1987; Scarpati et al., 1987b). Minor higher molecular weight transcripts have also been observed in some cells and tissues (Scarpati et al., 1987b; Fisher et al., 1987; Gregory et al., 1988). At present, there is no evidence for either an upstream promoter or a second polyadenylation signal to explain these larger transcripts, and therefore, it seems most likely that they have arisen by incomplete or alternative splicing as proposed by Scarpati et al. (1987b). The GM1381 fibroblast line used here expresses only the 2.3 kb mRNA species (Morrissey et al., 1987).

An AU-rich domain in the last 250 nucleotides of the TF transcript was previously identified (Scarpati et al., 1987b) as being similar to regions which specify rapid turnover of mRNAs encoding transiently expressed cytokines (Caput et al., 1986; Shaw & Kamen, 1986). The 3′ untranslated region of exon 6 contains only one polyadenylation signal which fits the consensus AATAAA (Proudfoot & Brownlee, 1976). In addition, five nucleotides spanning the polyadenylation site (TACTG) agree closely to the consensus CAYTG described by Berget et al. (1984) which is thought to function in cleavage site selection.

Nucleotide Differences between the TF Gene and cDNAs. The nucleotide differences between the TF gene sequence and the cDNA sequences reported previously (Fisher et al., 1987; Morrissey et al., 1987; Scarpati et al., 1987b; Spicer et al., 1987) are listed in Table III. Discrepancies at the extreme 5′ ends of some of the TF cDNA sequences are probably atrributable to artifacts introduced by resolution of secondary structure during in vitro synthesis of the cDNA (Land et al., 1981). There is a single nucleotide difference in the coding region which corresponds to a single amino acid change within the transmembrane domain of the mature protein. This was previously identified by Scarpati et al. (1987b), who proposed it as a probable sequence polymorphism. A surprising number of differences occur in the 3′ untranslated region in published cDNA sequences, although the gene sequence presented here agrees completely with that of the cDNA derived from pcTF417. It is not known whether the differences in published cDNA sequences are due to nucleotide sequence polymorphisms, cloning artifacts, or sequencing errors.

By comparing the sequences of λgTF885 and λgTF1377 (Figure 2) within intron 5, we have confirmed that the *Msp*I restriction fragment length polymorphism (RFLP) described

Table III: Sequence Differences between the TF Gene and Previously Published TF cDNA Sequences[a]

nucleotide position from Figure 4	gene	TF sequence from cDNA 1[b]	cDNA 2[c]	cDNA 3[d]	cDNA 4[e]
13	TCCCCGC			CGGGCGA	
47	A	A	G	A	
51	C	C	A	C	
91	GAAC	GAAC	GAAC	GAAC	CGTT
11184	T	T	C	T	T
11579	C	C	C	T	C
11585	G	G	G	T	G
11656	TT	--	-T	--	TT
11722	C	C	T	C	C
11735	CTG	CTG	CTG	CTTG	CTG
11816	G	A	A	G	G
11896	G	T	G	-	G
11923	GTG	GTG	GTG	GATG	GTG
12072	TT	TT	TTT	TT	TT
12111	TG	TG	TG	TGG	TG
12164	C	C	C	G	C
12330	GT	GT	GT	GTT	GT
12356	T	T	C	T	T
12420	TGA	TGA	TGGA	TGA	TGA
12428	A	A	-	A	A

[a] Blanks indicate sequence not available. [b] Fisher et al. (1987). [c] Scarpati et al. (1987b). [d] Spicer et al. (1987). [e] Morrissey et al. (1987) and the present study (pcTF417).

Table IV: Nucleotide Differences within Intron 5

polymorphism	nucleotide position from Figure 4	λgTF885	λTF1377
*Pst*I-CTGCAG	9820	CTGCAA	CTGCAG
*Msp*I-CCGG	10024	CCGG	CTGG

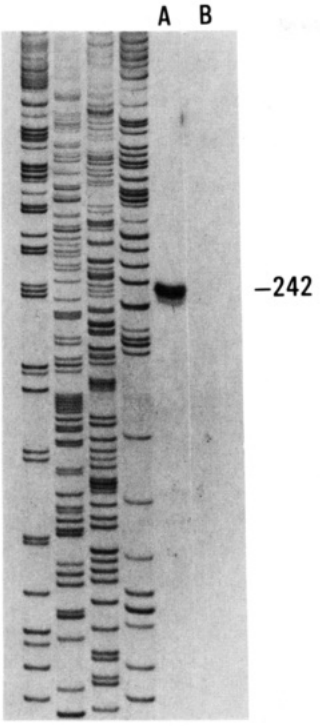


FIGURE 5: Primer extensions. Samples were analyzed on a 7% sequencing gel along with a sequencing reaction using a known DNA template as a standard. Lane A, primer extension using 4 μg of poly(A+) RNA from GM1381 cells as template; lane B, negative control using 10 μg of yeast tRNA as template. The length (in bases) of the major extension product is indicated. The autoradiogram was exposed for 18 h. A separate experiment using 4 μg of poly(A+) RNA from the MDA-MB-435S breast carcinoma cell line as a negative control template in place of the yeast tRNA gave identical results (data not shown). MDA-MB-435S cells have no detectable TF mRNA (Morrissey et al., 1987) and gave no detectable extension product.

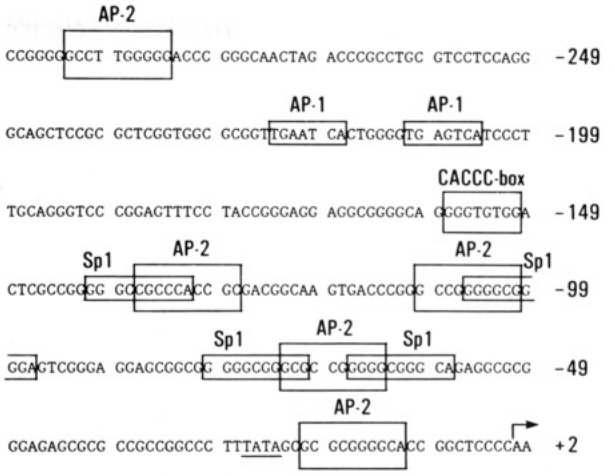


FIGURE 6: Identification of putative transcription factor binding sites. A region of 298 bp upstream of the TF transcription start site (┌→) was compared to the consensus sequences associated with the binding of a number of transcription factors. Sequences showing homology in either direction are boxed using the following consensus: Spl, 5′-GGGGCGG$^{GGC}_{AAT}$-3′; AP-1, 5′-TGANTCA-3′; AP-2, 5′-$^{TCC}_{CGG}$CC$^{A}_{C}$N$^{GCG}_{CGC}$-3′; CACCC box, 5′-CCNCACCC-3′.

by Scarpati et al. (1987a) is due to a nucleotide difference at position 10025 (see Table IV). Moreover, a *Pst*I site within intron 5 (Figure 2) was found to be present in λgTF1377 and absent in λgTF885; this is due to a second nucleotide difference at position 9825 (Table IV). The presence of the 3.0 kbp and 860 bp *Pst*I fragments in Figure 3 indicated that the fibroblast DNA contained this *Pst*I site. Although Scarpati et al. (1987a) found no evidence for *Pst*I polymorphisms within the TF gene in 6 individuals, our data from 12 individuals revealed both predicted homozygous and heterozygous patterns, suggesting that this *Pst*I site may indeed by polymorphic (data not shown).

Alu Repeats. A search of the TF gene sequence for the presence of *Alu* repeats identified three full-length elements and one partial repeat (Figure 4). In general, *Alu* repeat sequences are about 300 bp long and consist of two directly repeating monomer units (Kariya et al., 1987). They are the most abundant family of middle repetitive DNA sequences, being present at about 500 000 copies per haploid human genome. A subfamily of *Alu* repeats is very similar to 7SL DNA and may indicate the origin of this repetitive DNA (Jurka & Smith, 1988). An unusual feature of the TF gene is that one of these *Alu* repeats is within the 3′ untranslated region of exon 6, which was previously noted to be flanked by 11-nucleotide palindromic direct repeats that contain *Hin*dIII sites (Scarpati et al., 1987b). Downstream of the *Alu* repeat within intron 3 is a stretch of 68 purine residues, which may represent part of a purine-rich element recently identified to be closely linked to a subset of *Alu* repeats (Bosma et al., 1988). Other common repeats, such as *Sau*3A (Kiyama et al., 1986) and *Kpn*I (Sun et al., 1984), were not found within the TF gene. Our search for DNA sequence identity in the GenBank database (Release 56.0; July, 1988) revealed no significant homology to the TF gene sequence for any sequences other than the human TF cDNA and *Alu* repeats.

HTF Island. DNA sequence analysis of the TF gene revealed a highly (G + C)-rich 1.2 kbp region spanning the first exon and putative promoter (nucleotide positions −749 to +451; see also Figure 2). The association with the 5′ end of the TF gene suggested that it may represent an HTF island. In general, HTF islands are characterized by high G + C content, hypomethylation, and lack of CpG suppression (Bird, 1986). The G + C content of this region of the TF gene was calculated to be 70% with a 1:1.2 ratio of CpG to GpC dinucleotides. By contrast, a different 1.2 kbp region within the TF gene chosen at random for comparison (nucleotide positions 1854–3054) was 40% G + C with a CpG:GpC ratio of 1:7.8, typical of the bulk of human DNA. Although the function of HTF islands is unknown, they have often been found associated with the 5′ region of "housekeeping" genes and may play a role in constitutive expression of such genes in all cell types (Bird, 1986). HTF islands also appear to encompass the promoter regions of certain tissue-specific genes (Bird, 1986). Therefore, it will be important to elucidate the role, if any, of the HTF island in the regulation of TF transcription.

Potential Cis-Acting Regulatory Elements. Upstream of the TATA box, we have identified two direct repeats of 11 bp (DR-1) and 9 bp (DR-2) in length (Figure 4). These repeats and nearby sequences show some similarity to sequences flanking the genes for TNFα, IL-1, and plasminogen activator inhibitor 1 (data not shown), which are also regulated products of monocytes and/or endothelial cells in inflammatory responses. Functional studies will be required to determine whether these repeats represent binding sites for novel transcription factors.

Comparison of 298 bp of DNA sequence upstream of the transcriptional start site (Figure 6) with consensus sequences for transcription factor binding sites provided a number of matches for Sp1 (Kadonaga et al., 1986), AP-1 (Lee et al., 1987), and AP-2 (Imagawa et al., 1987) as well as a CACCC sequence (Dierks et al., 1983). The concentration of overlapping sequences predicted to bind Sp1 and AP-2 suggests a possible functional role for these transcription factors in the

control of TF gene expression. Moreover, AP-1 and AP-2 binding sites have been demonstrated to act as PMA (phorbol 12-myristate 13-acetate)-responsive elements in the 5′ flanking regions of several genes (Imagawa et al., 1987; Angel et al., 1987). Therefore, the predicted AP-1 or AP-2 sites in the putative TF promoter region may mediate the de novo induction of TF expression, based on the observation that PMA can induce TF expression in monocytes and endothelial cells in a concentration-dependent manner (Lyberg et al., 1983; Nawroth et al., 1985).

ACKNOWLEDGMENTS

We thank Dr. Habib Fakhrai for preparing the oligonucleotide, David Revak for excellent technical assistance, and Joy Lozano for typing the manuscript.

Registry No. TF, 9035-58-9; TF gene, 118334-24-0; TF gene mRNA, 118334-27-3; amino acids −32 to 263, 118334-29-5; amino acids 2–263, 118334-28-4.

REFERENCES

Angel, P., Baumann, I., Stein, B., Deliuw, H., Rahmsdorf, H. J., & Herrlick, P. (1987) *Mol. Cell. Biol. 7*, 2256–2266.

Bach, R., Gentry, R., & Nemerson, Y. (1986) *Biochemistry 25*, 4007–4020.

Bahnak, B. R., Howk, R., Morrissey, J. H., Ricca, G. A., Edgington, T. S., Jaye, M. C., Drohan, W. W., & Fair, D. S. (1987) *Blood 69*, 224–230.

Berget, S. M. (1984) *Nature 309*, 179–182.

Bevilacqua, M., Pober, J., Majeau, G., Cotran, R., & Gimbrone, M. (1984) *J. Exp. Med. 160*, 618–623.

Bird, A. P. (1986) *Nature 321*, 209–213.

Bonthron, D. T., Morton, C. C., Orkin, S. H., & Collins, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1492–1496.

Bosma, P. J., Van der Berg, E. A., & Koolstra, T. (1988) *J. Biol. Chem. 263*, 9129–9149.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.

Broze, G. J. (1982) *J. Clin. Invest. 70*, 526–535.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., & Cerami, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1670–1674.

Carson, S. D., Henry, W. M., & Shows, T. B. (1985) *Science 229*, 991–993.

Dierks, P., van Ooyen, A., Cochran, M., Dobkin, C., Reiser, J., & Weissmann, C. (1983) *Cell 32*, 695–706.

Drake, T. A., Morrissey, J. H., & Edgington, T. S. (1988) *FASEB J. 2*, 1410.

Dvorak, H., DeWater, L., Bitzer, A., Dvorak, A., Anderson, D., Harvey, V., Bach, R., Davis, G., DeWolf, W., & Carvalho, A. (1983) *Cancer Res. 43*, 4334–4342.

Edwards, R., & Rickles, F. (1984) *Prog. Hemostasis Thromb. 7*, 183–209.

Edwards, R., Rickles, F., & Bobrove, A. (1979) *Blood 54*, 359–370.

Esmon, C. T. (1987) *Science 235*, 1348–1352.

Fisher, K. L., Gorman, C. M., Vehar, G. A., O'Brien, D. P., & Lawn, R. M. (1987) *Thromb. Res. 48*, 89–99.

Gregory, S. A., & Edgington, T. S. (1985) *J. Clin. Invest. 76*, 2440–2445.

Gregory, S. A., Morrissey, J., Altieri, D., & Edgington, T. (1988) *FASEB J. 2*, 1078.

Imagawa, M., Chin, R., & Karin, M. (1987) *Cell 51*, 251–260.

Jurka, J., & Smith, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4775–4778.

Kadonaga, J. T., Jones, K. A., & Tjian, R. (1986) *Trends Biochem. Sci. (Pers. Ed.) 11*, 20–23.

Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S., & Matsubarg, K. (1987) *Gene 53*, 1–10.

Kiyama, R., Matsui, H., & Oishi, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4665–4669.

Land, H., Grez, M., Hauser, H., Lindenmaiser, W., & Schutz, G. (1981) *Nucleic Acids Res. 9*, 2251–2266.

Lee, W., Haslinger, A., Karin, M., & Tjian, R. (1987) *Nature 325*, 368–372.

Levy, G., Schwartz, B., & Edgington, T. S. (1981) *J. Immunol. 127*, 357–363.

Lyberg, T., Galdal, K. S., Evensen, S. A., & Pyrdz, H. (1983) *Br. J. Haematol. 53*, 85–95.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Morrissey, J. H., Fakhrai, H., & Edgington, T. S. (1987) *Cell 50*, 129–135.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Naora, H., & Deacon, N. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6196–6200.

Nawroth, P. P., Stern, D. M., Kisiel, W., & Bach, R. (1985) *Thromb. Res. 40*, 677–691.

Proudfoot, N., & Browlee, G. (1976) *Nature 263*, 211–214.

Rodgers, G., Broze, G., & Shuman, M. (1984) *Blood 63*, 434–438.

Ryan, J., & Geczy, C. (1987) *Immunol. Cell Biol. 65*, 127–139.

Sariban, E., Imamura, K., Leubbers, R., & Kufe, D. (1988) *J. Clin. Invest. 81*, 1506–1510.

Scarpati, E. M., Sadler, J. E., O'Connell, P. O., Nakamura, Y., Leppert, M., Ballard, L., Lathrop, G. M., Lalouel, J.-M., & White, R. (1987a) *Nucleic Acids Res. 15*, 9098.

Scarpati, E. M., Wen, D., Broze, G. J., Miletich, J. P., Flandermeyer, R. R., Siegel, N. R., & Sadler, J. E. (1987b) *Biochemistry 26*, 5234–5238.

Sharp, P. A. (1981) *Cell 23*, 643–646.

Shaw, G., & Kamen, R. (1986) *Cell 46*, 659–667.

Spicer, E. K., Horton, R., Bloem, L., Bach, R., Williams, K. R., Guha, A., Kraus, J., Lin, T.-C., Nemerson, Y., & Konigsberg, W. H. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5148–5152.

Stern, D., Bank, I., Nawroth, P., Cassimeris, J., Kisiel, W., Fenton, J., Dinarello, C., Chess, L., & Jaffe, E. (1985) *J. Exp. Med. 162*, 1223–1235.

Sun, L., Paulson, K., Schmid, C., Kadyk, L., & Leinward, L. (1984) *Nucleic Acids Res. 12*, 2669–2690.

Taylor, F. B., Chang, A., Esmon, C. T., D'Angelo, A., Vigano-D'Angelo, S., & Blick, K. E. (1987) *J. Clin. Invest. 79*, 918–925.