The gene for human TATA-binding-protein-associated factor (TAF_{II}) 170: structure, promoter and chromosomal localization

Jan A. VAN DER KNAAP*, Vincent VAN DEN BOOM*¹, Jeroen KUIPERS†, Michiel J. T. VAN EIJK‡², Peter C. VAN DER VLIET* and H. Th. Marc TIMMERS*³

*Laboratory for Physiological Chemistry and Center for Biomedical Genetics, Utrecht University, P.O. Box 80042, 3508 TA Utrecht, The Netherlands,

†Department of Immunology, University Hospital Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands, and ‡Laboratory of Metabolic Diseases,

Wilhelmina Children's Hospital, P.O. Box 85090, 3508 AB Utrecht, The Netherlands

The TATA-binding protein (TBP) plays a central role in eukaryotic transcription and forms protein complexes with TBP-associated factors (TAFs). The genes encoding TAF_{II} proteins frequently map to chromosomal regions altered in human neoplasias. TAF_{II}170 of B-TFIID is a member of the SF2 superfamily of putative helicases. Members of this superfamily have also been implicated in several human genetic disorders. In this study we have isolated human genomic clones encoding TAF_{II}170 and we show that the gene contains 37 introns. Ribonuclease-protection experiments revealed that TAF_{II}170 has multiple transcription start sites, consistent with the ob-

servation that the promoter lacks a canonical TATA box and initiator element. Deletion analysis of the promoter region showed that a fragment of 264 bp is sufficient to direct transcription. In addition, we determined the chromosomal localization by two independent methods which mapped the gene to human chromosome 10q22–q23 between the markers D10S185 and WI-1183. The region surrounding these markers has been implicated in several human disorders.

Key words: chromosome 10, RNA polymerase II, SNF2, TFIID.

INTRODUCTION

Initiation of transcription by RNA polymerase II (pol II) requires the assistance of auxiliary factors. One of these factors is the TATA-binding protein (TBP). Although TBP can act as a monomer *in vitro*, in mammalian cells all TBP is complexed to TBP-associated factors (TAFs) [1]. At least four TBP-containing complexes have been identified. SL1 consists of TBP and three TAF_Is and is involved in RNA polymerase I transcription. TFIIB contains two TAF_{III}s and acts in RNA polymerase III transcription. In case of pol II transcription two distinct complexes have been isolated, TFIID and B-TFIID [1].

Multiple functions can be assigned to TAFs (reviewed in [2–4]). (i) TAFs of each TBP-containing complex determine the RNA polymerase specificity. (ii) Numerous reports show that TAF_{II}s are co-activators capable of directly participating in signal-transducing pathways controlling cell proliferation. (iii) Some TAF_{II}s interact with specific promoter elements and thereby stabilize promoter binding by TFIID. (iv) TAF_{II}s can also act as negative regulators of transcription. A complication in the analysis of TAF function is that some TAF_{II}s in TFIID are also present in other protein complexes that lack TBP, but which contain histone acetyltransferase activity (for review, see [4]). The best-studied pol II complex is TFIID, composed of TBP and at least ten TAF_{II}s (for review, see [2]). Some heterogeneity has been reported in the composition of TFIID. Mapping of the chromosomal location of nine of the TFIID TAF_{II}s showed that six of them are positioned in regions commonly altered in human neoplasias [5].

The other complex involved in pol II transcription is B-TFIID, which consists of TBP and TAF_{II}170 (also known as TAF172) [6,7]. We have found that approximately one third of TBP present in HeLa-cell extracts resides in B-TFIID (J. A. van der Knaap and H.Th.M. Timmers, unpublished work, and [1]). Like TBP, B-TFIID is able to support basal transcription by pol II in a reconstituted system. However, B-TFIID is unable to commit a template in transcription and hence binds less stably to the TATA box than TBP. In addition, TAF_{II}170 is unable to support activated transcription [1]. TAF_{II}170 exhibits ATPase activity and cloning of its cDNA revealed that it is a member of the SNF2 family of proteins [8]. The SNF2 family is part of the SF2 superfamily of putative helicases [9,10], members of which have been implicated in human genetic disorders like xeroderma pigmentosum (XPB and XPD), Cockayne's syndrome and Bloom's syndrome (for review, see [11]). TAF_{II}170 turned out to be the human homologue of yeast Mot1p. *MOT1* has been identified in genetic screens for elevated expression of several unrelated genes [12–15]. Recent experiments *in vitro* using cell extracts show that Mot1p can either enhance or repress transcription depending on its stoichiometry with TBP [16]. Biochemical evidence showed that Mot1p can remove TBP from the TATA box in an ATP-dependent manner [17,18]. The same activity has been reported for the human homologue [7].

To analyse possible involvement of TAF_{II}170 in human disorders, we studied the gene encoding TAF_{II}170. Here we report the chromosomal localization. We find that the TAF_{II}170 gene maps to a chromosomal region frequently altered in specific neoplasias. In addition, we have isolated genomic P1-artificial

Abbreviations used: TBP, TATA-binding protein; TAF, TBP-associated factor; pol II, RNA polymerase II; FISH, fluorescence *in situ* hybridization; PAC, P1-artificial chromosome; CREB, cAMP-response-element-binding protein; ATF, activating transcription factor; AP, activator protein.

¹ Present address: Department of Cell Biology and Genetics, Medical Genetics Center, Erasmus University Rotterdam, P. O. Box 1738, 3000 DR Rotterdam, The Netherlands.

² Present address: Keygene NV, P.O. Box 216, 6700 AE Wageningen, The Netherlands.

³ To whom correspondence should be addressed (e-mail H.T.M.Timmers@med.uu.nl).

The nucleotide sequence data reported will appear in the GenBank Nucleotide Sequence Database under the accession numbers AF166116, AF166117 and AF166118.

chromosome (PAC) clones containing the *TAF_{II}170* gene, determined the intron–exon borders and analysed the *TAF_{II}170* promoter.

EXPERIMENTAL

Isolation of genomic clones for *TAF_{II}170*

The human PAC library RPCI-1 (#FPAC-3386; Genome Systems, St. Louis, MO, U.S.A.) was screened with a cDNA probe ranging from nt +49 to +1869 relative to the translation start site by the YAC Screening Center (Leiden, The Netherlands). Positive clones were isolated and the insert size was determined by digestion with *NotI*, pulsed-field gel electrophoresis and Southern blotting.

Identification of exon–intron borders

A library of the *TAF_{II}170* gene was constructed by digesting the genomic PAC clone 135-i12 with *XhoII*, random insertion of the fragments into pBSII-KS+ digested with *BamHI* and subsequent transformation into the *Escherichia coli* strain DH5 α . The bacterial colonies were replica-plated to Hybond-N filters (Amersham), which were subsequently replicated for screening purposes. Plasmids containing a fragment of the *TAF_{II}170* gene were identified by colony hybridization with radiolabelled *TAF_{II}170* cDNA probes according to standard procedures [19]. Positive clones were sequenced with *TAF_{II}170* oligonucleotides using the ABI Prism® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit in the GeneAmp® PCR System 2400 and analysed with the ABI Prism 310 Genetic Analyser (Applied Biosystems, Foster City, CA, U.S.A.). Resulting DNA sequences were compared with *TAF_{II}170* cDNA sequences.

Promoter plasmids

The plasmid p135Xho5, a subclone of PAC 135-i12, contained 4.5 kb upstream of the translation start site and 0.5 kb downstream. From this plasmid a 1446-bp *PstI*–*NcoI* fragment, containing the region of *TAF_{II}170* directly upstream of the translation start site, was cloned into pBSIIKS-ATG-FLAG digested with *PstI* and *NcoI*. This construct, pBS-TAF170-P/N, was used as starting material for the construction of the pGL3-TAF170 promoter constructs. pGL3-TAF170-P/N was made by inserting the 1474-bp *XbaI* (blunt-ended with Klenow fragment)–*NcoI* fragment of pBS-TAF170-P/N into pGL3-enhancer (#E1771, Promega) digested with *Asp718*, blunt-ended with Klenow fragment, and *NcoI*. pGL3-TAF170-H/N was made by digesting pGL3-TAF170-P/N with *PstI* and *HindIII*, blunt-ending with S1 nuclease, and reclosure by ligation. pGL3-TAF170-A/N was constructed by digesting pGL3-TAF170-P/N with *Asp718* and reclosure by ligation. pGL3-TAF170-P/Nrev was made by blunt-ending both ends of the 1474-bp *XbaI*–*NcoI* fragment of pBS-TAF170-P/N with Klenow fragment and inserting it into pGL3-enhancer digested with *Asp718* and *NcoI* (double blunt-ended with Klenow fragment). pGL3-TAF170-P/A was made by introducing the 1185-bp double blunt-ended *PstI*–*Asp718* fragment of pBS-TAF170-P/N into pGL3-enhancer digested with *Asp718* and *HindIII* (double blunt-ended with Klenow fragment). pGL3-TAF170-H/A contained the double blunt-ended 264-bp *HindIII*–*Asp718* fragment ligated to pGL3-enhancer digested with *Asp718* and *HindIII* and blunt-ended with Klenow fragment. pGL3-promTAF170-P/Arev was made by inserting the 1201-bp *Asp718*–*BamHI* fragment of pBS-TAF170-P/N into pGL3-enhancer digested with *Asp718* and *BglII*. dNcoI-GL3 was made by digesting pGL3-enhancer

with *Asp718* and *NcoI*, blunt-ending the ends of the vector with Klenow fragment and reclosure by ligation. pGL3-dHind3 was constructed by removing the *Asp718*–*HindIII* fragment from pGL3-enhancer, blunt-ending the ends and reclosure by ligation.

Transient transfections

The human osteosarcoma cell line, U2OS, was cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 300 μ g/ml L-glutamine, 100 units/ml penicillin and 100 μ g/ml streptomycin. Transfections were performed using a modification of the calcium phosphate co-precipitation method [20]. Briefly, U2OS cells were seeded at a density of 8.5×10^3 cells/cm² in six-well plates and after 24 h the DNA/calcium phosphate mixture was added. Each dish received 2 μ g of luciferase construct, 0.25 μ g of the internal control CMV-lacZ and 7.75 μ g of pEP40 carrier DNA. After 20 h medium was refreshed and after an additional 24 h of culture the cells were harvested in 300 μ l of lysis buffer (25 mM Tris/phosphate, pH 7.8/15% glycerol/1% Triton X-100/8 mM MgCl₂/1 mM dithiothreitol). Luciferase assays were performed as described elsewhere [21]. β -Galactosidase activity was determined using the Galacto-light-plus kit (Tropix, Bedford, MA, U.S.A.). Each transfection was performed in duplicate and values are expressed as means $\pm$ S.D. of three independent experiments.

Ribonuclease-protection assay

For the construction of the radiolabelled antisense RNA probe, pBSII-TAF170-P/N was linearized with *SspI*. Transcription with T3 RNA polymerase and purification of the probe by incubation with DNase I and ethanol precipitation were performed according to standard protocols [22]. The probe was hybridized to total RNA from either HeLa cells or the human multipotent embryonal carcinoma NTERA2 clone D1 (NT2D1) cell line or to yeast tRNA, followed by digestion with ribonuclease A and ribonuclease T1. The products were analysed on a denaturing polyacrylamide/urea gel according to standard protocols [22].

Radiation-hybrid mapping

The primer pair 5'-GGGAAACAAGTTATTCTCA-3' and 5'-TCAGCCATGCATTACTAC-3', corresponding to the 3' untranslated region of *TAF_{II}170* (GenBank accession number AF166117), was used for mapping of *TAF_{II}170* on the GeneBridge 4 Radiation Hybrid Panel [23]. The PCR reactions, yielding a 719-bp product, were performed with 25 ng of template DNA, 1.5 mM MgCl₂, 100 μ M dNTPs, 0.5 μ M of each primer and 0.8 units of *Taq* DNA polymerase in a total volume of 10 μ l. The reaction profile was: a 5-min precycle denaturation at 94 °C, 39 cycles of denaturation for 15 s at 94 °C, 30 s of primer annealing at 57 °C and 45 s of extension at 72 °C, followed by a 7-min end extension at 72 °C. PCR products were resolved by 1% agarose gel electrophoresis. The data file was submitted to the Whitehead Institute radiation-hybrid mapping database to place *TAF_{II}170* in the context of the radiation-hybrid mapping framework markers.

Fluorescence *in situ* hybridization (FISH)

FISH was carried out on phytohaemagglutinin-stimulated normal human lymphocytes using standard procedures. In short, after competing for repetitive sequences with a 50 $\times$ excess C₀t-1 DNA (Life Technologies), 250 ng of biotin-labelled pN10 cDNA [8] was hybridized on denatured metaphase chromosomes

for 48 h at 37 °C. After washing at 42 °C in 50 % formamide/ $2 \times$ SSC (where $1 \times$ SSC is 0.15 M NaCl/0.015 M sodium citrate) the hybridized probe was detected by applying an avidin–CY3 conjugate (Jackson Immuno Research, West Grove, PA, U.S.A.). The signal was amplified using a biotinylated goat anti-avidin antibody (Vector Laboratories, Burlingame, CA, U.S.A.) as a second layer and avidin–CY3 again as a third layer. Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI) and the resulting banding pattern was used to assign the chromosome with the probe signal using a fluorescence microscope with CCD camera and Cytovision software (Applied Imaging, Sunderland, Tyne and Wear, U.K.).

RESULTS AND DISCUSSION

TAF_{II}170 is located on chromosome 10

To determine the chromosomal location of *TAF_{II}170*, several independent methods were adopted. First, FISH analysis was performed with the previously described 3N10 cDNA clone as a probe [8]. Signals were observed on chromosome 10 between band q22 and q23 and no signal was detectable on any other chromosome (Figure 1). This was in agreement with preliminary data obtained by PCR analysis of somatic human–hamster cell hybrids which indicated that the *TAF_{II}170* gene is located on chromosome 10 (results not shown). To verify the FISH data, radiation-hybrid analysis was carried out on the GeneBridge 4 panel consisting of 93 radiation-hybrid clones spanning the whole human genome. By this method the *TAF_{II}170* gene was assigned to chromosome 10, placing it between the markers D10S185 and WI-1183 with a distance of 1.5 cR to each marker. D10S185 has been assigned to the 10q23–q24 boundary [24]. It is interesting to note that allelic loss of D10S185 occurs in 6.25 % of stage-B prostate carcinomas. The loss of heterozygosity of D10S185 in advanced prostate cancers (stages C and D) is even higher (21 %) [25]. Also, in 18 % of squamous cell carcinomas of the lung, loss of heterozygosity of D10S185 has

been observed [26]. Loss of all informative loci at 10q has been reported in 23 out of 36 (64 %) glioblastomas studied. Also, in 41 % of the investigated meningiomas loss of 10q was detected [27]. The occurrence of these disorders cannot be accounted for solely by mutations of the PTEN tumour-suppressor gene, located at 10q23.3 [27]. Recent data also indicate that an as-yet unidentified tumour-suppressor gene on 10q23 is probably involved in breast carcinomas [28,29]. Finally, a complicated form of spastic paraparesis has been mapped between the markers D10S564 and D10S603, which flank D10S185 [30].

In this respect it is interesting to compare *TAF_{II}170* with the other *TAF_{II}s*. Correlations have been found between cytogenetically detectable mutations in human neoplasias and the chromosomal locations of *TAF_{II}100*, *TAF_{II}80*, *TAF_{II}55*, *TAF_{II}31*, *TAF_{II}28* and *TAF_{II}43* [5]. In addition, human *TAF_{II}68* shows sequence similarity with the pro-oncoproteins TLS/FUS and EWS [31] and *TAF_{II}250* is identical to cell cycle gene 1 (*CCG1*) [32,33]. This indicates that *TAF_{II}s* can be considered potential proto-oncogenes. Analogous to *Mot1p*, it can be expected that disruption of *TAF_{II}170* will affect the expression of numerous genes. *TAF_{II}170* is also likely to be an essential protein, because deletion of *MOT1* is lethal [12]. Since *TAF_{II}170* is also a member of the SF2 superfamily of putative helicases, it is an intriguing possibility that *TAF_{II}170* is involved in the human disorders described above.

Genomic structure of the *TAF_{II}170* gene

Screening of the human PAC library RPCI-1 with a cDNA fragment of 1.8 kb from the N-terminus of *TAF_{II}170* as a probe resulted in the identification of seven positive clones: 14-a6, 31-15, 54-b17, 135-i12, 219-h14, 305-h15 and 306-h15. Southern-blot analysis and PCR indicated that PAC 135-i12 possesses an insert of approx. 130 kb and was most likely to contain the complete *TAF_{II}170* gene (results not shown). To determine the exon–intron borders a sub-library was constructed of PAC 135-i12 and PAC 306-h15 by digesting the clones with *Xho*II. This sub-library was screened with *TAF_{II}170* cDNA probes and positive clones were sequenced with oligonucleotides specific for *TAF_{II}170* cDNA. Comparison with the *TAF_{II}170* cDNA sequence resulted in the identification of 37 introns (Table 1). All introns appeared to be of the U2 type except for intron 24, which contained a perfect consensus sequence for the U12-dependent spliceosome [34]. The ATPase domain of *TAF_{II}170* (amino acids 1263–1779) spans exons 27–37. Comparison of this domain with the ATPase domain of its closest human homologue, *ERCC6*, showed that both have the same number of introns in this domain. Although they are not located at identical positions, the borders of the *TAF_{II}170* introns 27, 31, 33 and 35 are almost at the same location as the corresponding ones in *ERCC6* (results not shown). Since the formation of the *MOT1/TAF_{II}170* and *ERCC6* sub-families occurred before the separation of fungal and animal ancestors [10], there might be an evolutionary pressure to form similar exon–intron structures in this domain. We recently described the isolation of two cDNA clones (N5D and 3N10) from a human B-lymphocyte cell library, which seemed to contain a non-spliced intron [8]. Careful inspection of the sequence of intron 9 (GenBank accession number AF166118) revealed that the genomic sequence in these two cDNA clones is due to a cryptic 5' splice site. This site is located 128 bp downstream of the correct site. Use of this cryptic splice site would result in a frame shift and yield a truncated protein of 350 amino acids (amino acids 1–330 + VSVLMQQLSPSINSWHL-FPM).

Figure 1 *TAF_{II}170* is located on human chromosome 10, between bands q22 and q23

The cDNA clone 3N10, containing the complete open reading frame of *TAF_{II}170* except for the first 54 bp, was hybridized to human metaphase chromosomes as described in the Experimental section. The *TAF_{II}170* fluorescence signals are indicated by arrows. The insert shows a magnification of chromosome 10 and a schematic representation.

Table 1 Exon-intron organization of *TAF_{II}170*

Exon sequences are in upper case and intron sequences in lower case. The intron sequences of the 3' splice acceptor of introns 4, 14 and 20 have not been determined (indicated by n). 9* indicates the cryptic splice site as found in the cDNA clones 3N10 and N5D. The size of exon 1 is the distance between the most upstream transcription start site and the first intron. The size of exon 38 is the distance between the last intron in the coding region and the predicted polyadenylation signal found in clone 3N10.

Exon	Exon size (bp)	5' Splice donor	3' Splice acceptor
1	289	GTCTCCAGgtgggtcttg	gttatcttcagGCTAGATC
2	124	TGTCTAAAgtaagaactt	aatcttttaagGTGTGAT
3	115	CAGACAAggtgcttttaa	cctcatgtagAACCTACT
4	147	AAAATCAGgtctgtagtg	nnnnnnnnnnGTGAAGTG
5	164	AACAACCTgtaggtaaaa	attcttttagACTCTTCA
6	137	GAGAAGAGgtagtaattct	tatttttttagCAATGATA
7	130	TTGAAGAGgtactcttga	tcttattcagACAAATGA
8	69	CCTGGGAGgttaagatttc	ttcttttcagGTTTCGACA
9	90	TAGAAGAGgtaagtgtat	attttaacagATGATTCA
9*	218	ATGAAAAGgtaaatatca	
10	96	CTGATGAAGtaagtatca	tcgtctgtagGTTGTGGC
11	177	TCCGTCAGgttaaatattt	tgttttacagGATGTAAT
12	141	CACAAAAGgttaattaaa	gatcttatagGTGCCCTT
13	125	CAATGCAGgttaattattt	tctttattagTATTCAGC
14	121	AGGACCAGgttaagaactg	nnnnnnnnnnAACTCTTC
15	105	TTCAACAAGgtacacagca	taccctgcagGTTTGGAT
16	150	GAGCCAAGgttaagtgtag	ccctctgtagGAAAAAAC
17	143	GCAGCCAAGtgagtgtac	tcctctcttagGTTGTTAG
18	175	TACAAAAGgttaagatttt	tgtatcttagGAGTGTA
19	204	GTGACTTGgtgagtaaaag	tttttaaatagGTCACTAC
20	427	TTCTAAAGgtatataacct	nnnnnnnnnnGATCCACC
21	191	TTGATGAGgttaagttagt	taatttttaagGCCCAGAA
22	154	TAATTTTGgtatacacat	attttaacagATGGAAAA
23	110	ATCCCTTGgttaactaact	ttgttttttagTTGGTCCA
24	202	ACTTGCCGTgtatcctttt	acttactcagGTGTAATG
25	149	CACTTGAGgttaagtcaaa	atcatcttagGCAGGCAT
26	153	ATCAGCAGgttaagtttta	tctctatcagGATGGTGT
27	122	TGTCATAGgttaatttaag	tattgtaaagGGCCAGG
28	168	AGAATAAGgttaagagttg	ctaaataaagGTTACAGC
29	78	TTCTTTAGgttaagaatta	tctcccttagAAATATTA
30	130	CAATCCAGgttaattattt	ctgtttacagAACACGT
31	142	AGAAGCAGgttatgaaaga	ctattaacagGTGTCTT
32	131	CTCTCCAGgttaggaatc	catctttacagGTTTCAGCT
33	126	TATTCAGgttatagatta	cctcttgtagGCATTACA
34	153	TGAACAAGgtatgtatgt	tcttcctaagTTGCTGTT
35	212	GTTTCCCGgttaagtggct	ctacttttagGTTTAATA
36	160	TTGGGCAGgttaaagtc	tcttatccagAAACGTGT
37	171	TTGATAAGgttaagaact	ttttcttaagGATGGCAA
38	1509		

The *TAF_{II}170* gene has multiple transcription start sites

In order to map the transcription start sites of the *TAF_{II}170* gene we developed a ribonuclease-protection experiment with an RNA probe containing the translation start codon and 869 nucleotides upstream. Analysis of the protected fragments on a denaturing sequencing gel revealed the presence of five major transcription start sites and several weaker ones (Figure 2, lanes 3 and 4). The start sites are located in a window spanning from +1 to +33. No protection of the probe was observed using only tRNA (Figure 2, lane 2). The fact that the protection pattern is the same using RNA samples from different cells lines shows specificity of the signals (see Figure 2). It has been reported previously that the lack of both an initiator and a TATA box can lead to multiple transcription start sites [35]. The results of the ribonuclease-protection experiment are in agreement with the transient transfections (see below).

Figure 2 The *TAF_{II}170* gene has multiple transcription start sites

Ribonuclease-protection analysis was performed on RNA isolated from HeLa or NT2D1 cells. The numbers on the left refer to the sizes of the DNA markers (M). The region containing the major transcription start sites is indicated by the black bar on the right. The five major transcription start sites are indicated by arrows.

The *TAF_{II}170* promoter contains no TATA box and initiator

To investigate the transcriptional regulation of the *TAF_{II}170* gene the nucleotide sequence of the 1446-bp *PstI*-*NcoI* fragment directly upstream of the translation start site was determined (Figure 3; GenBank accession number AF166116). Visual inspection of this region did not reveal the presence of a TATA box or initiator element at the expected positions relative to any of the transcription start sites. Also, no splice acceptor sequences are present between the transcription start sites and the translation start codon. Screening of the Transfac database with MatInspector V2.2 [36] revealed the presence of numerous transcription-factor-binding sites in the 1446-bp *PstI*-*NcoI* fragment. Among these were three binding sites for cAMP-response-

```

      -1172
CTGCAGTACT TCTCATTAGC CTCATTTTAC AGATGAAAAA CAGACTCAAG GATTGAGACA
Pst I

TAGCCTCTTC AGGTAATGAG AAGCCAATGG ACATGCTTTA TAAGCATAGG AGCGACATGA
TAAAAAGCAG TGAGGTTTTT TGAAGATTAA TCTGTTGGCC TATAAAATGG GCTGAACTGA
AAATTGGAGA AATATTATTT TGACTACTAG CATAAAAAATC AAGTTTAAGG GCCGGGCGCA
GTGGCTCACA CCTGTAATCC CAGCACTTTG GGAGGCCGAG GCGGGCAGAT CACTTGAGGC
CAGGAGCAGG AGACAAGCCT GGCCTACATG GCGAAACCCC GTCTCTACTA AAAAATACAA
AAAATTGGCC CGGCGTGGGG CTCGTGTGCC TCCCAGCTAC TCGGAGGTT GAGGCACATT
TAGAATCGCT TGAACCCGGT AAGCGCAGGT TGGCAGTGAG CCGAGATCGC GCCACTGCAC
TCCAGCCTGG GCGACAGAGC GAAACTCGGT CAAAAAATAA AAATCAAGTG TAAGATGATA

      -594
AGGGCCTGGC AGAAGTGTTC AAAAGCCTAA TGAAGTTAAA AATATTGAAA AGCCAGATAA
Ssp I

GTCCCCCAAA GGAACAAAAA TTAAAAATG GTAGTTTTC AATCTGGCTG GCTGAGAAAA
TGATGGTACA AACACATAA ACAATTGAGA AATCAAACCTT GGCAAAATCG TTGGTCAGAA
GACCTTCTCT CCAATACCTA GGCAGAACTT CTGGCAAATA ATGACAAACA TCTGAGCCAT
TTCTTCAGTG TCATAGCGAC TTTTGTCTT CACTCCTCCC ACTGGCTTCA CCACCTGGAC
AAGAGCATGT TAAGGTCCTT CTGATCTTGC TTCTAGATG CAGTACGGTG GGAAGCAGAA

      -251
TCTGGGTAGT GTGAATTAG AAACCAAGCT TTCGAAGTTC CAGGACTTGG ACTAAATCGC
Hind III

CTTTTTCCTT CTCTGTCAA TAAGGGGACC CTTACTTCAC CTCCTTCCCT AGTTTTAAGA
CTAGACACCC TCTCCACATC AGCACAAAGC GCTTTTATTT AGGCGCGTGC CGCCAAACCA
AAGGCGCTCG CGCCGAGCTA AGCTCCTGTC GGTGAGCTTA CGTCACTTCC GGCGCGCTGA

      +1
CGCTCAGTTG TGCGTGCCGA CGCCCGCGTC AGCAAAGAGC GGAGCTGAGG GTACCGGTT
Asp 718

TGAAGTCGTG CGGGTCGGAG GACTGCCGCC TCCGCTACCG TCTTGGACCC CTGCTTACCG
GCCCGCGCGG GGACGAGCTC GGTAGGCGG CAGGGCAGAT GCCCGAGGGC CTGCGCTGGA
ACGCCCCACG CCGCACCGGC CGCTCCCTTG TGCTCCGCGG CCTGGGCCTG CGCCGCTCAG
CTCTCTGGAA ACTAGCGCCT CAGCTGCGCG GCGCGTAGGT CGCGGGGAGC TCCGAACCGC

      +275
CGGCGCCCGG CCATGGCGGT CTCCAGgtgg gt
M A V S R
Nco I

```

Figure 3 Sequence of the *TAF_{II}170* promoter and exon 1

The region containing the transcription start sites is indicated in italics. The five most predominant start sites are indicated in bold italics. The transcription start site furthest upstream is designated +1 and the rest of the sequence is numbered relative to this site. Restriction sites are indicated below the sequence and are underlined. Predicted GATA1-binding sites carry bold underlining and predicted CREB, ATF-1 and AP-1 sites are double underlined. A predicted AP-1 site, which is not a CREB or ATF-1 site, is underlined by dashes. Amino acids are depicted by single-letter code. Intron sequence is in lower case.

element-binding protein (CREB), activating transcription factor (ATF)-1 and activator protein (AP)-1 in the -251/+13 region and three GATA-binding sites and one for AP-1 in the -1172/-251 region.

264 bp are sufficient for *TAF_{II}170* promoter activity

Transcription-start-site mapping predicted that the 1446-bp *Pst*I-*Nco*I fragment contained the *TAF_{II}170* promoter. In order

to test for promoter activity we cloned this fragment directly in front of the luciferase gene, which was followed by simian virus 40 enhancer sequences (pGL3-*TAF_{II}170*-P/N; Figure 4). Transient transfection of human U2OS cells and measurement of the luciferase activity in extracts of these cells revealed that this region was able to direct transcription at levels comparable with those of the Herpes simplex virus thymidine kinase promoter (Figure 4B). As expected, *TAF_{II}170* promoter activity is uni-

Figure 4 TAF_{II}170 promoter activity in U2OS cells

(A) Schematic representation of the TAF_{II}170 promoter. P, *Pst*I; H, *Hind*III; A, *Asp*718; N, *Nco*I. The arrow indicates the transcription start site furthest upstream. Numbering refers to the position of the most upstream transcription start site, which was set as +1. (B) Luciferase analysis of U2OS cells transfected with the indicated constructs. The values obtained for pGL3-TAF_{II}170-P/N were arbitrarily set to 100%. pGL3-dNcoI and pGL3-dHind3 contained no promoter construct. 'mock' indicates cells transfected only with CMV-lacZ. This value is $0.07 \pm 0.03\%$.

directional since inversion of this region did not induce luciferase activity above background levels (compare pGL3-TAF_{II}170-P/Nrev with pGL3-dNcoI). We also tested for autoregulation of the TAF_{II}170 gene by co-transfection of a plasmid expressing the TAF_{II}170 protein, but we did not observe a regulatory effect (results not shown).

Deletion of the region upstream of the *Hind*III site at -251 resulted in a 2.7-fold decrease in luciferase activity, whereas further reduction in size of the region until the *Asp*718 site at +14 completely abolished promoter activity. To test whether the +14/+274 (*Asp*718-*Nco*I) fragment was dispensable for promoter activity, we made two constructs lacking this region. pGL3-TAF_{II}170-P/A spanned from -1172 to +13, and pGL3-TAF_{II}170-H/A encompassed nucleotides -251 to +13. Both constructs maintained promoter activity. In addition, inversion of the -1172/+13 fragment abolished all promoter activity. In conclusion, the minimal promoter region of the TAF_{II}170 gene lies within the *Hind*III-*Asp*718 fragment (-251/+13). This suggests that transcription factors belonging to the b-ZIP class (ATF, Fos/Jun and CREB families) are involved in regulation of the TAF_{II}170 gene.

We thank Glenn Barnes and Marcy MacDonald (Molecular Neurogenetics Unit, Massachusetts General Hospital, Boston, MA, U.S.A.) for initial studies on the chromosomal localization. We thank Marjolein de Ruwe and Thomas Albert for critical reading of this manuscript and we are grateful to Wim van Driel, Fiona van den Heuvel and Roel Schiphof for excellent technical assistance. This work was supported by the Netherlands Foundation for Chemical Research (SON), with financial aid from the Netherlands Organization for Scientific Research (NWO) and a PIONIER grant from NWO-MW.

REFERENCES

- 1 Timmers, H. T. M. and Sharp, P. A. (1991) *Genes Dev.* **5**, 1946–1956
- 2 Burley, S. K. and Roeder, R. G. (1996) *Annu. Rev. Biochem.* **65**, 769–799
- 3 Lee, T. I. and Young, R. A. (1998) *Genes Dev.* **12**, 1398–1408
- 4 Bell, B. and Tora, L. (1999) *Exp. Cell Res.* **246**, 11–19
- 5 Purrello, M., Di Pietro, C., Viola, A., Rapisarda, A., Stevens, S., Guermah, M., Tao, Y., Bonaiuto, C., Arcidiacono, A., Messina, A. et al. (1998) *Oncogene* **16**, 1633–1638
- 6 Timmers, H. T. M., Meyers, R. E. and Sharp, P. A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8140–8144
- 7 Chicca, J. J., Auble, D. T. and Pugh, B. F. (1998) *Mol. Cell. Biol.* **18**, 1701–1710
- 8 van der Knaap, J. A., Borst, J. W., van der Vliet, P. C., Gentz, R. and Timmers, H. T. M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11827–11832
- 9 Gorbatenya, A. E. and Koonin, E. V. (1993) *Curr. Opin. Struct. Biol.* **3**, 419–429
- 10 Eisen, J. A., Sweder, K. S. and Hanawalt, P. C. (1995) *Nucleic Acids Res.* **23**, 2715–2723
- 11 Ellis, N. A. (1997) *Curr. Opin. Genet. Dev.* **7**, 354–363
- 12 Davis, J. L., Kunisawa, R. and Thorner, J. (1992) *Mol. Cell. Biol.* **12**, 1879–1892
- 13 Piatti, S., Tazzi, R., Pizzagalli, A., Plevani, P. and Lucchini, G. (1992) *Chromosoma* **102**, S107–S113
- 14 Jiang, Y. W. and Stillman, D. J. (1996) *Genes Dev.* **10**, 604–619
- 15 Prelich, G. (1997) *Mol. Cell. Biol.* **17**, 2057–2065
- 16 Muldrow, T. A., Campbell, A. M., Weil, P. A. and Auble, D. T. (1999) *Mol. Cell. Biol.* **19**, 2835–2845
- 17 Auble, D. T. and Hahn, S. (1993) *Genes Dev.* **7**, 844–856
- 18 Auble, D. T., Hansen, K. E., Mueller, C. G., Lane, W. S., Thorner, J. and Hahn, S. (1994) *Genes Dev.* **8**, 1920–1934
- 19 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn., Cold Spring Harbor Press, Cold Spring Harbor
- 20 van der Eb, A. J. and Graham, F. L. (1980) *Methods Enzymol.* **65**, 826–839
- 21 de Wet, J., Wood, K. V., DeLuca, M., Helinski, D. R. and Subramani, S. (1987) *Mol. Cell. Biol.* **7**, 725–737

- 22 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidmann, J. G., Smith, J. A. and Struhl, K. (1992) *Short Protocols in Molecular Biology*, Greene Publishing Associates and John Wiley & Sons, New York
- 23 Walter, M. A., Spillelt, D. J., Thomas, P., Weissenbach, J. and Goodfellow, P. N. (1994) *Nat. Genet.* **7**, 22–28
- 24 Gray, I. C., Fallowfield, J., Ford, S., Nobile, C., Volpi, E. V. and Spurr, N. K. (1997) *Genomics* **43**, 85–88
- 25 Ittmann, M. (1996) *Cancer Res.* **56**, 2143–2147
- 26 Petersen, S., Rudolf, J., Bockmuhl, U., Gellert, K., Wolf, G., Dietel, M. and Petersen, I. (1998) *Oncogene* **17**, 449–454
- 27 Bostrom, J., Cobbers, J. M., Wolter, M., Tabatabai, G., Weber, R. G., Lichter, P., Collins, V. P. and Reifengerger, G. (1998) *Cancer Res.* **58**, 29–33
- 28 Feilolter, H. E., Coulon, V., McVeigh, J. L., Boag, A. H., Dorion-Bonnet, F., Duboue, B., Latham, W. C., Eng, C., Mulligan, L. M. and Longy, M. (1999) *Br. J. Cancer* **79**, 718–723
- 29 Freihoff, D., Kempe, A., Beste, B., Wappenschmidt, B., Kreyer, E., Hayashi, Y., Meindl, A., Krebs, D., Wiestler, O. D., von Deimling, A. and Schmutzler, R. K. (1999) *Br. J. Cancer* **79**, 754–758
- 30 Seri, M., Cusano, R., Forabosco, P., Cinti, R., Caroli, F., Picco, P., Bini, R., Morra, V. B., De Michele, G., Lerone, M., Silengo, M., Pela, I., Borrone, C., Romeo, G. and Devoto, M. (1999) *Am. J. Hum. Genet.* **64**, 586–593
- 31 Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. and Tora, L. (1996) *EMBO J.* **15**, 5022–5031
- 32 Ruppert, S., Wang, E. H. and Tjian, R. (1993) *Nature (London)* **362**, 175–179
- 33 Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M. and Roeder, R. G. (1993) *Nature (London)* **362**, 179–181
- 34 Burge, C. B., Padgett, R. A. and Sharp, P. A. (1998) *Mol. Cell* **2**, 773–785
- 35 Geng, Y. and Johnson, L. F. (1993) *Mol. Cell. Biol.* **13**, 4894–4903
- 36 Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884

Received 2 September 1999; accepted 10 November 1999