BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **225,** 275–280 (1996)
ARTICLE NO. 1166

# Promoter Structure of the Mouse TATA-Binding Protein (TBP) Gene

Tetsuya Ohbayashi, Edward E. Schmidt,* Yasutaka Makino, Toshihiko Kishimoto, Yo-ichi Nabeshima,† Masami Muramatsu,‡ and Taka-aki Tamura[1]

*Department of Biology, Faculty of Science, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba-263, Japan; *Department of Molecular Biology, Sciences II, University of Geneva, Geneva 4, CH-1211, Switzerland; †Department of Molecular Genetics, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo-187, Japan; and ‡Department of Biochemistry, Saitama Medical School, Moroyama, Iruma-gun, Saitama-350-04, Japan*

Received July 2, 1996

5′-RACE and genomic cloning were used to determine that the mouse TBP (mTBP) gene consists of one 5′-terminal non-coding exon followed by seven protein-coding exons. The region upstream of the first exon lacks a TATA-box. Hence, as with the case of other genes carrying TATA-less promoter, transcription starts from a cluster of sites which are located at the restricted region of mTBP gene. Interestingly, sequences of this region are well conserved between human and mouse TBP genes, suggesting that both mouse and human TBP genes drive mRNA synthesis in a similar way, and the sequence homology between two genes was used to assign a putative start point for the human TBP gene. Mouse, human, and *Trimersurus gramineus* (green habu snake) TBP genes share two GC-rich regions in their promoter regions. Thus, it is probable that diverse species of vertebrates commonly use TATA-less promoter bearing GC-rich regions to direct ubiquitous TBP expression. 

The TATA-binding protein (TBP), originally identified as a factor that binds to the TATA-box of RNA polymerase II-driven genes, is required by all classes of nuclear RNA polymerases (1, 2). TBP-containing SL1, TFIID, and TFIIIB complexes bind to promoter sequences and direct transcription initiation by RNA polymerase I, II, and III, respectively (1, 2). TBP cDNAs have been obtained from various organisms (2). Their C-terminal 180 amino acids, which have been demonstrated to be involved in association with numerous transcription factors as well as interaction with DNA, are highly conserved all the organisms examined thus far. Furthermore, amino terminal half of the protein, which is rather diversed among different species, appears appreciably to be conserved among vertebrates, except for the length of CAG-repeat (Q-run).

Human (3) and green habu snake (4) TBP genes are composed of eight exons. Seven introns are found at similar positions in the protein coding regions and 5′-terminus non-coding regions. Previously, we reported that the mouse TBP (mTBP) gene has seven exons (5). However, based on the conservation of organization of gene structure between human and snake TBPs, we suspected that the first exon may have been overlooked in the previous work. In this work, we have re-investigated the 5′-terminal region of the mTBP gene and demonstrated that the gene actually has eight exons.

Since TBP has a general role in all eucaryotic gene expression, it is particularly interesting to study how TBP gene expression is regulated. Schmidt *et al.* (6) has demonstrated that mTBP mRNA is overexpressed in rodent male germ cells. However, the promoter structure and function of mammalian TBP genes were not clear. In the present work, we identified the

[1] Corresponding author. Fax: 43-290-2824.



0006-291X/96 $18.00

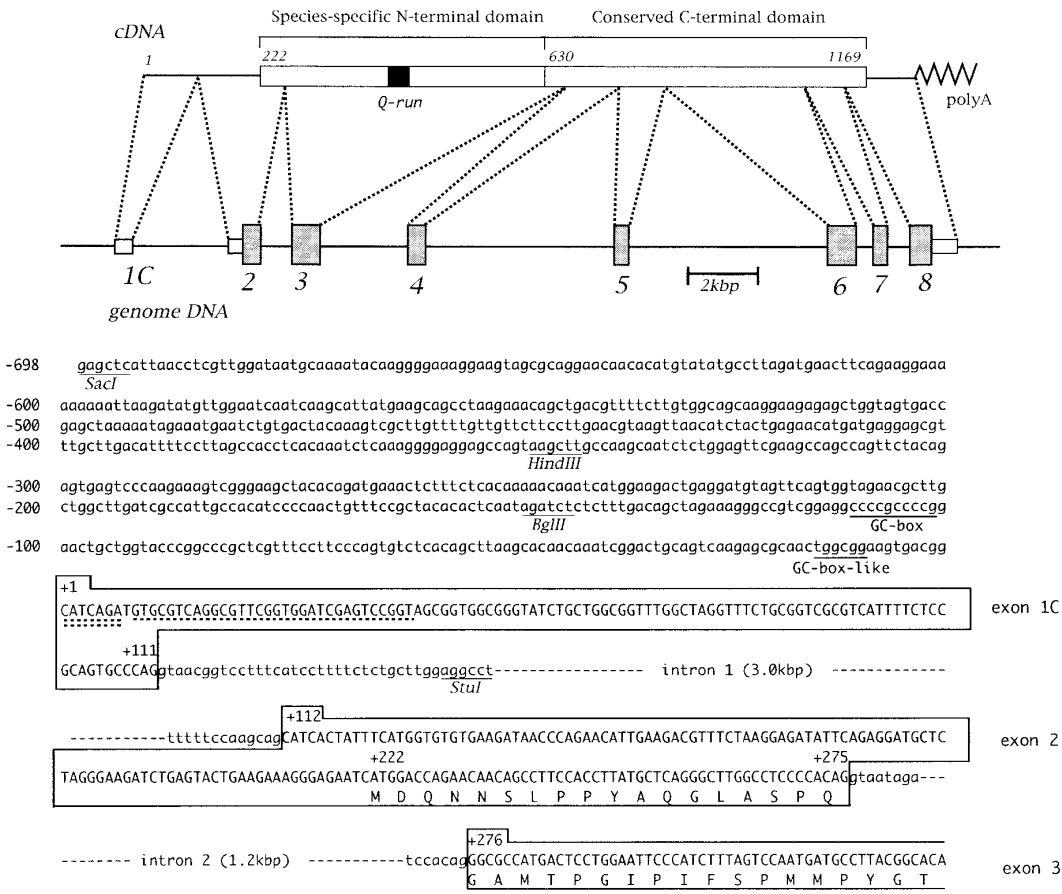


**FIG. 1.** Structure of the mouse TBP gene. (A) Exon−intron relationship of the mouse TBP gene. Numbers indicate nucleotide positions relative to the 1st nucleotide of the major transcription initiation sites (Fig. 2). On the genomic sequence, meshed and open boxes of the genome DNA indicate coding and non-coding exons, respectively. (B) Nucleotide and amino acid sequences of the mTBP genome around the transcription start sites. Nucleotide sequences of exon 1C, 2, and part of exon 3, intron 1, and intron 2 are also shown. Double- and single-dashed lines indicate the major and minor start sites, respectively.

transcription start points on the mTBP gene. We also show that the mTBP gene lacks a TATA-box but contains multiple GC-box or GC-box-like elements instead.

## MATERIALS AND METHODS

*5′-RACE assay.* The 5′-RACE assay was carried out using a brain-derived 5′-RACE kit (Promega) using the manufacture's protocol. The 1st and 2nd PCR primers were 5′-AGTGCAATGGTCTTTAGGTCA AGTTTACAG (positions 485-456 on the cDNA sequence of Tamura *et. al.* , 1991 (7)), and 5′-CTCAGATCTTCCCTAGAGCAT-CCTC (positions −21 - −45) (7), respectively. The PCR products were subcloned into *Bam*HI/*Eco*RI sites of pBluescript SK(+) using the *Bgl*II site (underlined) and *Eco*RI site in the adapter primer.

*Genomic screening and DNA sequencing.* A 129 strain mouse-derived geomic library was screened with a *Pst*I fragment (positions 221-524 (7)) of mTBP cDNA by standard methods (8). Nucleotide sequencing was performed by a dye terminator sequencing kit and autosequencer (PERKIN-ELMER).

*RNase protection.* RNase protection assays were performed as described previously (6, 9). Genomic fragment extending from the *Hin*dIII site at −447 to the *Stu*I site 35 bp downstream of exon 1C (Fig. 1B) was subcloned into pBluescript SK(+) and linearized with *Bgl*II (Fig. 1B). The DNA was transcribed by T7 RNA polymerase and used as a probe. Hybridization mixture contained 100mg of total RNA from the indicated mouse tissues prepared by

sedimentation through CsCl cushions as described (9) and 3.0 femptomoles of radiolabeled probe. Control hybridization mixture contained 100mg of yeast RNA and 3.0 femptomoles of the probe. Protected fragment sizes were determined by comparison to size standards.

## RESULTS AND DISCUSSION

*The mouse TBP gene consists of eight exons.* Whereas other vertebrate *tbp* gene had been shown to consist of 8 exons, an earlier study on the mTBP gene revealed 7 exons. To determine whether the mTBP gene, like other vertebrates, had an upstream initiator exon that we previously overlooked, we performed a 5′-RACE using mouse brain RNA and two primers derived from known mTBP cDNA sequences. The RACE products were cloned and nucleotide sequences were determined. One of the RACE products contained 145 bp of sequence upstream of the previously identified cDNA (data not shown). A *Pst*I digested fragment of mTBP cDNA (positions 221-524) (7) as a probe, we screened 129 mouse-derived genomic library, and isolated one clone. Sequencing of the clone revealed the RACE-derived sequences (data not shown). Using this clone, we determined the exon-intron relationship (Fig. 1A), and identified a new initiator exon at the 5′-terminus. Thus, the mTBP gene also has 8 exons like other vertebrates. We designated the newly identified 5′-terminal exon as exon 1C, and the previously reported ''exon 1'' (Sumita *et al.* (5)) was renamed as exon 2.

*The mTBP gene has multiple transcription start sites and no TATA-box.* We determined the transcription start sites of the mTBP by RNase protection (Fig. 2). Total RNA preparations from mouse brain, spleen, and liver were used to map the start sites. The three tissues yielded a similar pattern of protected fragments; yeast RNA control yielded no protected fragments (Fig. 2). Upon mapping sizes of the protected fragments, we noted the pattern of fragments with sizes ranging from 77- to 111-bases in length was reiterated by fragments exactly 35-bases longer. Each of these longer fragments was roughly 10-fold less abundant than its respective shorter counterpart (Fig. 2). Because the probe contained 35 bp of intron 1 (Fig. 1), the sizes and relative abundances of the larger fragments are consistent with their being non-spliced pre-mRNAs. Thus, for example, initiation at position +1 (Fig. 1B) would give a pre-mRNA that hybridized to 146 bases of the probe, and this precursor would be spliced to give a mRNA that hybridized to 111 bases of the probe (Fig. 2). This interpretation is corroborated by the finding that, relative to the shorter protected fragments, the longer species are more abundant in nuclear RNAs as compared to whole cell RNA preparations, and less abundant in poly$(A)^+$ mRNA preparations than in total RNAs. The major initiation sites maped to +1, +2, +5, and, +7; minor start sites were scattered between +9 and +40.

The start sites lied about 220 bp upstream of the translation start site. The mTBP gene harbors no TATA-box near the initiation sites (Fig. 1B). The same region of the snake TBP (sTBP) gene (Fig. 3B) and human TBP (hTBP) gene also lacks a TATA-box (Fig. 3A). Schmidt *et al.* further demonstrate that the mTBP gene uses multiple promoters in testis, all of which lack a typical TATA-box (manuscript in preparation). Thus, it appears that the vertebrate TBP genes are all ''TATA-less''. This is in sharp contract to lower eucaryotes such as *Acanthamoeba castellanii* (10), *Schizosaccharomyces pombe* (11), and *Saccharomyces cerevisiae* (12), all of which have a TATA-box in their upstream regions.

*Structural similarities between mammalian TBP genes around transcription start sites.* We compared nucleotide sequences around the transcription start sites for three vertebrate TBP genes. The gene structure and transcription start sites for sTBP have been reported (4); the transcription start site of the hTBP has not been determined yet, though the surrounding sequences have been reported (5). From sequence alignment with mTBP and hTBP genomes (Fig. 3), we found the first exons of mouse and human TBPs shared 65% sequence identity with the greatest conservation (87%) near the major transcription start site. In contrast, sequences upstream of the start sites displayed only 58% similarity. Based on this sequence

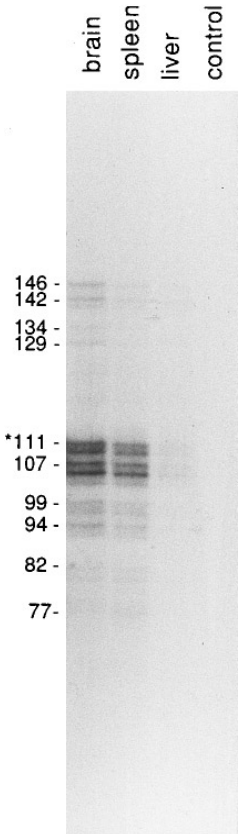


**FIG. 2.** Transcription start sites of mTBP gene. The RNase protection probe was a 325-bp antisence transcript that contains 293-bases of genomic *tbp* sequence as indicated under Materials and Methods. RNAs from mouse brain, spleen, and liver, and yeast were analyzed. Protected fragment sizes are indicated at the left.

similarity, we tentatively assigned a putative transcription start site for hTBP gene. Little similarity existed between the mTBP and sTBP gene. The sequence of the sTBP exhibited only $<54\%$ similarity to mTBP within the domains presented in Fig. 3B, despite the high conservation found protein-coding sequence (4).

*Promoters of the vertebrate TBPs commonly have GC-box motifs.* The mTBP gene around $-116$ had a striking sequence identity (91%) with those of human around $-146$. The homologous regions encompassed about 23 bp and were GC-rich. This region contains a putative inverted GC-box. The sTBP gene contained a GC-box at $-107$ but exhibited little similarity to mTBP or hTBP sequences in this region. Degenerate GC-boxes were also found at positions $-15$ and $-15$ in mTBP and hTBP genes, respectively within the highly conserved region near transcription start sites mentioned above (Fig. 3B). Moreover, the sTBP promoter contains an inverted GC-box at position $-36$. Thus, all of the vertebrate TBP promoters appeared to harbor one distal and one proximal GC-box or GC-box-like elements.

Ubiquitously expressed genes often lack a TATA-box but have GC-boxes (Sp1-binding sites) within their promoters (13–15). In some cases, GC-boxes are known to play a role in transcriptional activation. Moreover, TATA-less promoters frequently direct transcription from multiple sites. Our present work shows that the vertebrate TBP promoters fall into this class of ubiquitously active, TATA-less, and GC-box-containing promoters. There is a good correlation

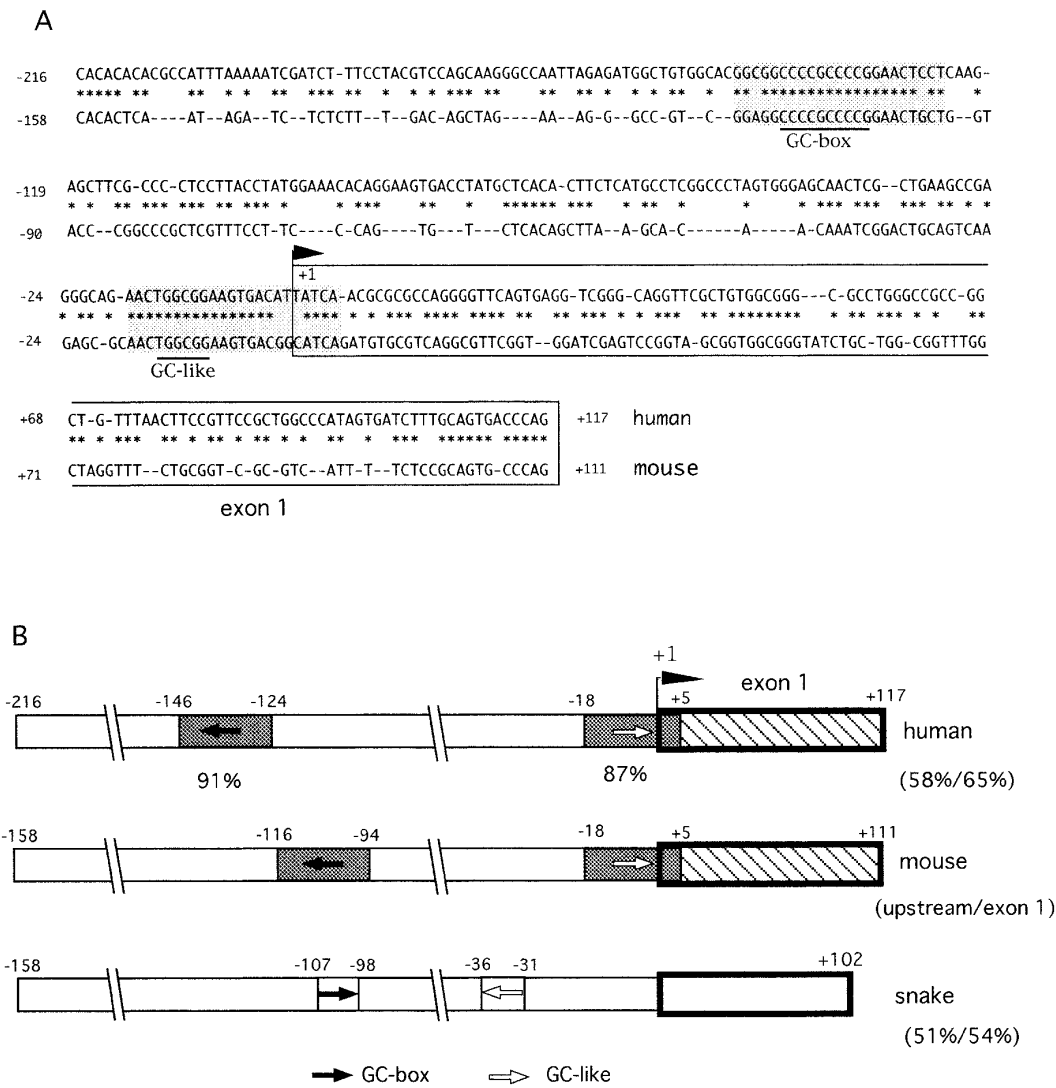


**FIG. 3.** Comparisons of the vertebrate TBP genomes around their promoter regions. (A) Nucleotide sequence alignment of mouse and human (3) TBPs. GC-box and GC-like elements are underlined. Regions of sequence similarity are indicated by shadowing. Asterisks represent matched bases between human and mouse. (B) GC-box and GC-like elements in promoters of vertebrate TBP genes. The human TBP transcription start site was tentatively assigned by homology analysis as described above. Boxes with mesh and oblique lines represent >85% and >65% sequence identities between human and mouse, respectively. Each first exon is indicated by a bold open box. Sequence identities for human (3) and snake (4) TBP genomes against the mTBP in upstream region and exon 1 are also presented.

between net amounts of Sp1 (16) and TBP (6) in the mouse testis. We also know that Sp1-deficient lower eucaryotes (13, 14) have a TATA-box in TBP genes. These findings will be important for understanding how somatic cells regulate *tbp* gene expression and thus maintain homeostasis.

## ACKNOWLEDGMENT

This work was supported by a *Grant-in-Aid for Scientific Research* from the *Japanese Ministry of Education, Science, Sports and Culture.*

## REFERENCES


1. McKnight, S. L. (1996) *Genes Dev.* **10,** 367–381.
2. Hernandez, N. (1993) *Genes Dev.* **7,** 1291–1308.
3. Chalut, C., Gallois, Y., Poterszman, A., Moncollin, V., and Egly, J.-M. (1995) *Gene* **161,** 277–282.
4. Nakashima, K., Nobuhisa, I., Deshimaru, M., Ogawa, T., Shimohigashi, Y., Fukumaki, Y., Hattori, M., Sakaki, Y., Hattori, S., and Ohno, M. (1995) *Gene* **152,** 209–213.
5. Sumita, K., Makino, Y., Katoh, K., Kishimoto, T., Muramatsu, M., Mikoshiba, K., and Tamura, T. (1993) *Nucleic Acids Res.* **21,** 2769.
6. Schmidt, E. E., and Schibler, U. (1995) *Development* **121,** 2373–2383.
7. Tamura, T., Sumita, K., Fujino, I., Aoyama, A., Horikoshi, M., Hoffmann, A., Roeder, R. G., Muramatsu, M., and Mikoshiba, K. (1991) *Nucleic Acids Res.* **19,** 3861–3865.
8. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
9. Schmidt, E. E., and Schibler, U. (1995) *J. Cell Biol.* **128,** 467–483.
10. Wong, J. M., Liu, F., and Bateman, E. (1992) *Nucleic Acids Res.* **20,** 4817–4824.
11. Hoffmann, A., Horikoshi, M., Wang, C. K., Schroeder, S., Weil, P. A., and Roeder, R. G. (1990) *Genes Dev.* **4,** 1141–1148.
12. Schroeder, S. C., Wang, C. K., and Weil, P. A. (1994) *J. Biol. Chem.* **269,** 28335–28346.
13. Kadonaga, J., Jones, K., and Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23.
14. Melton, D. W., McEwan, C., McKie, A. B., and Reid, A. M. (1986) *Cell* **44,** 319–328.
15. Tamura, T., and Mikoshiba, K. (1991) *FEBS Lett.* **282,** 87–90.
16. Saffer, J. D., Jacson, S. P., and Annarella, M. (1991) *Mol. Cell. Biol.* **11,** 2189–2199.