The gene encoding the TATA-binding protein, TBP, is highly overexpressed during the haploid stages of spermatogenesis in rodents. RNase protection analyses for mRNAs containing the previously identified first, second, and eighth exons suggested that most TBP mRNAs in testis did not initiate at the first exon used in somatic cells (here designated exon 1C). Using a sensitive ligase-mediated cDNA amplification method, 5’ end variants of TBP mRNA were identified, and the corresponding cDNAs were cloned from liver and testis. In liver, a single promoter/first exon is used to generate a steady-state level of roughly five molecules of TBP mRNA per diploid cell equivalent. In testis, we detect modest up-regulation of the somatic promoter and recruitment of at least five other promoters. Three of the alternative promoter/first exons, including 1C and two of the testis-specific promoter/first exons, 1D and 1E, contribute roughly equivalent amounts of mRNA which, in sum, account for greater than 90% of all TBP mRNA in testis. As a result, round spermatids contain an estimated 1000 TBP mRNA molecules per haploid cell. Testis TBP mRNA also exhibits several low abundance 5’ end splicing variants; however, all detected TBP mRNA leader sequences splice onto the common exon 2 and are expected to initiate translation at the same site within exon 2. The precise locations of the three major initiation exons are mapped on the gene. The identification of the strong testis-specific promoter/first exons will be important for understanding spermatid-specific tbp gene regulation.

The TATA-binding protein, TBP, is required for transcription initiation by all three nuclear RNA polymerases (1). Somatic tissues contain from 1 to 5 TBP mRNA molecules per cell; these modest differences correspond qualitatively with relative differences in overall transcriptional activity in nuclei from different tissues (2, 3). In contrast, adult rodent testes exhibit roughly 350 molecules of TBP mRNA per cell (2). Testis-specific TBP mRNA overexpression arises primarily as a result of transcriptional up-regulation (2).

During puberty, the first stem cells to undergo spermatogenesis do so almost synchronously, such that different stages of puberty can be correlated to the first appearance of a specific spermatogenic stage in the testis (4). The onset of TBP overexpression during puberty correlates to the appearance of the first haploid cells. Moreover, as shown by immunocytochemistry, TBP protein overexpression occurs in the early haploid cells (2). Because these data suggest that TBP overexpression is restricted to only a subset of the cells in testis, we estimate that TBP mRNA levels in these cells are more than 1000-fold greater than those in somatic cell types. This high level of cell type-specific overexpression makes us suspect that TBP plays a role in spermatogenesis that differs from its ubiquitous functions in somatic cells.

With the goal of understanding the mechanisms regulating spermatid-specific tbp gene expression, we have performed a molecular analysis of the 5’ end of the tbp gene. Sumita et al. (5) previously reported the intron/exon structure of the mouse tbp gene from exons 2 to 8 (formerly designated exons 1–7). In brain, spleen, and liver, use of a single upstream promoter/first exon has been reported (6). Quantitative analyses presented here show that testis-specific tbp expression involves both modest up-regulation of the somatic tbp promoter and recruitment of at least 2 other major and 3 minor promoters. Using a sensitive method for amplifying and cloning cDNA 5’ ends, we characterize 10 TBP mRNA 5’ end variants in testis that differ in promoter usage and/or splicing. The relative contribution of each variant to total testis and liver TBP mRNA levels, the gene structure, and the entire 5’ end genomic sequence are reported.

**EXPERIMENTAL PROCEDURES**

**Animals and RNA Preparation**—Genomic sequencing was performed on mouse 129 ES cell genomic clones. All other samples were prepared from fresh tissues harvested from MORO mice or Sprague-Dawley rats as indicated. Total RNA was prepared by sedimentation through CsCl cushions as described previously (3); nuclear RNA was prepared from citric acid-prepared nuclei sedimented through two sucrose cushions followed by purification of RNA through CsCl cushions as described previously (2).

**RNase Protection Assays**—RNase protection assays were performed as described previously using the indicated amounts of sample RNA supplemented with yeast RNA to 100 μg (2). Control lanes contained probe and 100 μg of yeast RNA; probe control lanes contained roughly a 1:100 dilution of nondigested probe and 10 μg of yeast RNA carrier. Pseudo-pre-mRNA was transcribed from a mouse genomic clone extending from the SacI site 700 bp upstream of exon 1C to the SacI site in...
amplified by using an anti-adapter primer and, first, a primer specific
in preparation.

the remaining 68 bases of exon 2 but rather gave unrelated sequences
sequence with this primer. False-primed cDNA products did not exhibit
ing exon 2 primer. Plasmids with no inserts (2% of the clones) would not
ods are described elsewhere.2 cDNA was synthesized from 1
ethanol. Pellets were collected by centrifugation, washed with 70%
EDTA); samples were extracted with chloroform and precipitated with
primer listed above or a distinct primer to exon 2 sequences, 5′
GAGCATCCTC-3′ extending through the poly(A) tail.

clones 95 (mouse) or 62 (rat); and the exon 1E probes were transcribed
amplification products; the exon 1D/2 probes were transcribed from
A/aspan spanning from the SoxI site at −700 to the SoxI site 35 bp downstream of
exon 1C linearized with BglII or Sau3AI, respectively. The Sa/X probe
were amplified using a mouse cDNA clone containing sequences beginning near the 5′ end of exon 2 (35 bp upstream of
BglII site) and extending through the poly(A) tail. Pseudo-mRNAs containing exons 1D-2 or exons 1E-2 were transcribed from
clonesequences in exon 2 primer described above or a distinct primer to exon 2 sequences, 5′
- GAAGTGCAATGGTCTTTAGGTCAAGTT-
- CCATGTTCTGGATCTTGAAG-3′.

An mRNA sample (20 μg) was digested with T4 RNAse, per
containing mRNA, and 1 amol of exon 1C-containing
probe (not shown), the protected fragments that were 35 bases
protected fragment (Fig. 1C) using an anti-adapter primer and, first, a primer specific for sequences in exon 4 (5′-GAAGTGCAGATCCTGCTTATAGAAGTT-
TACG-3′), followed by a 25-base primer (5′-CAGACATCTTTCCCTA-
GAGCCATCCTC-3′), which spanned the Bgl II site in exon 2 (underlined).
Amplified cDNAs were cloned into BamHI/NcoI-cut Bluescript KS+
plasmids (Stratagene) using the Bgl II site in exon 2 and a NotI site in the adapter primer. Clones were sequenced using the Bgl II-containing
exon 2 primer. Plasmids with no inserts (2% of the clones) would not
sequences with this primer. False-primed cDNA products did not exhibit
remaining 68 bases of exon 2 but rather gave unrelated sequences (about 20% of the clones).

Primers Extension Analyses—Poly(A)+-mRNA samples (20 μg) were
mixes with 5.0 fmol of 32P-end-labeled primer (either the Bgl II-containing
primers listed above or a distinct primer to exon 2 sequences, 5′
-TGCTGGTGGTGGGCTCAGT-3′) and hybridized under overnight
in 10 μl of hybridization buffer (10 mM Pipes, pH 7.0; 400 mM NaCl; 1
mM EDTA; 0.05% SDS) at 45 °C. After hybridization, samples received
100 μl of dilution buffer (10 mM Tris, pH 7.5; 300 mM NaCl; 0.5 mM
EDTA); samples were extracted with chloroform and precipitated with ethanol. Pellets were collected by centrifugation, washed with 70%
ethanol, and resuspended in 100 μl of reaction mix (50 mM Tris, pH 8.3
at 25 °C; 75 mM KCl; 3 mM MgCl2; 5 mM dithiothreitol; 0.3 μM each of
dATP, dCTP, dGTP, dTTP; 0.02 units/μl RNasin) at 45 °C. Each sample received 200 units of BRL Superscript I RNase H-free reverse transcriptase,
and the incubation bath was adjusted to 50 °C (tamp time 45–50 °C = 10 min). After a 1-h incubation, each sample received 100 μl
RNase mix (10 mM Tris, pH 7.5; 10 mM EDTA; 5 μg/ml RNase A; 1.8
units/μl RNase T1) and was incubated at room temperature for 10 min.
Samples received 6.5 μl of proteinase K mix (100 μl Tris, pH 7.5; 50
mm EDTA; 10% SDS; 1 mg/ml proteinase K; 0.5 mg/ml yeast RNA) and
were incubated 10 min at 37 °C. Samples were diluted with 200 μl of (10
mM Tris, pH 7.5; 300 mM NaCl; 0.5 mM EDTA; 0.2% SDS) and
were extracted with phenol/chloroform and precipitated with ethanol.
Products were resuspended in 75% formamide, 75 °C, and resolved on
denaturing polyacrylamide gels.

Reverse Transcriptase-mediated Polymerase Chain Reaction—cDNAs were synthesized using 1 μg of total or poly(A)+-mRNA. cDNAs (2% of
each reaction) were amplified using the Bgl II exon 2 primer described
above and primers specific to exon 1C (5′-GGCGGCGATATCGTCGGG-
GT-3′) and 3′ end of exon 1D (5′-GGCGGCGATATCGTCGGG-3′)
using Taq polymerase and standard amplification conditions. Amplified
products were separated by agarose gel electrophoresis and were visualized
by ethidium bromide staining.

* E. E. Schmidt, L. Marzaccii, U. Schibler, and J. Geiselmann,
script in preparation.
pre-mRNA, per sample contained roughly 8 amol of TBP mRNA and 0.6 amol of TBP. We analyzed fragments size for pre-mRNA. due to UTP deprivation during high specific activity transcription and to radiolysis thereafter, it is nearly impossible to produce and maintain full-length high specific activity probes of this size, so we expected to find many smaller protected fragments on the gels. Therefore experiments were designed such that true exonic mRNA-derived signals could be distinguished from probe heterogeneity products based on three criteria. First, products arising from probe heterogeneity should appear upon hybridization to either cellular pre-mRNA or synthetic pseudo-pre-mRNA, whereas signals arising from spliced TBP mRNA should not appear in pseudo-pre-mRNA controls. Second, relative to pre-mRNA signals, signals corresponding to bona fide mRNA should be over-represented in nuclear as compared with total RNA preparations (Fig. 1C). Finally, signals corresponding to mRNA should be over-represented in polyadenylated mRNA preparations as compared with signals arising from pre-mRNA.

Using the SaX probe (Fig. 3A), which contains 83 bases of exon 1C and 342 bases of sequence downstream of the exon 1C splice site, we were able to validate the method. Exon 1C-containing mRNA appears as an 83-base band which is under-represented in nuclear RNA preparations and is not detected using pseudo-pre-mRNA; unspliced pre-mRNA appears as a 425-base band which is abundant in nuclear RNA preparations and is indistinguishable from the signal obtained with pseudo-pre-mRNA. Interestingly, this probe also revealed a protected fragment about 350 bases long in total and nuclear RNA but not in pseudo-pre-mRNA (Fig. 3A) or poly(A)⁺ mRNA.
Fig. 3. RNase protection search for alternate exons between exons 1C and 2. Probes to the regions diagrammed at the top of the figure were used to search for alternate exons in testis. To aid identification of exonic fragments, pseudo-pre-mRNA controls were included, nuclear and total RNA preparations were compared, and total and poly(A)+-enriched preparations were compared. The Bg/St probe showed no evidence for mRNA or pre-mRNA species containing sequences from upstream of exon 1C (<0.1 amol/µg total RNA; Fig. 1). A–C, bands corresponding to putative exonic sequences are demarcated by dark arrowheads on the autoradiograms and are indicated at the right of each panel. A, the Sa/X probe yielded one major (425 bases long) and one minor protected fragment (roughly 310 bases long) with pseudo-pre-mRNA. Both bands also appear at similar levels in total and nuclear testis RNA preparations and thus are attributed to pre-mRNA. The Sa/X probe also yielded a cluster of protected fragments of approximately 152 and 50 bases long which is enriched in total as compared with nuclear RNA preparations. A series of fragments centered around 200 bases in length are also enriched in total as compared with nuclear RNA and thus likely represent mRNA sequences. The 350-base fragment seen with total and nuclear RNA but not with pseudo-pre-mRNA (demarcated with an open arrowhead on the autoradiogram and indicated to the right of the panel) is the expected size for intronic sequences which have been cleaved from exon 1C.

(bands were subsequently found to arise from hybridization to alternate exon with multiple initiation sites. Indeed, these bands are centered around 200 bases in length that fit the criteria for mRNA or pre-mRNA species containing sequences from upstream of exon 1C (<0.1 amol/µg total RNA; Fig. 1). A–C, bands corresponding to putative exonic sequences are demarcated by dark arrowheads on the autoradiograms and are indicated at the right of each panel. A, the Sa/X probe yielded one major (425 bases long) and one minor protected fragment (roughly 310 bases long) with pseudo-pre-mRNA. Both bands also appear at similar levels in total and nuclear testis RNA preparations and thus are attributed to pre-mRNA. Exon 1C appears as an 83-base band which is enriched in total as compared with nuclear RNA preparations. A series of fragments centered around 200 bases in length are also enriched in total as compared with nuclear RNA and thus likely represent mRNA sequences. The 350-base fragment seen with total and nuclear RNA but not with pseudo-pre-mRNA (demarcated with an open arrowhead on the autoradiogram and indicated to the right of the panel) is the expected size for intronic sequences which have been cleaved from exon 1C. Markers from top: 510, 396, 344, 298, 221, 154, and 74/75 bases. B and C, as for A using the indicated probes. For the X/A probe (B), protected fragments of approximately 152 and 50 bases long were under-represented in nuclear RNA and over-represented in poly(A)+ mRNA preparations. For the A/R1 probe, fragments of about 55 bases were under-represented in nuclear RNA as compared with total RNA preparations. All of these fragments represent putative exons. For the R1/Bg probe, no evidence of exonic fragments was found (not shown).

Based on the size and distribution of this fragment, we suspect that it represents intronic sequences that have been excised from exon 1C. This species is present at about 0.2 amol/µg total RNA, which is 1/25 and 1/5 of the concentrations of exon 1C-containing TBP mRNA and pre-mRNA, respectively.

The Sa/X probe also yielded a cluster of protected fragments centered around 200 bases in length that fit the criteria for mRNA. Thus, we predicted that this region likely contains an alternate exon 1C with multiple initiation sites. Indeed, these bands were subsequently found to arise from hybridization to TBP mRNAs containing the various initiation sites for exon 1D (see below).

With the X/A (Fig. 3B) and A/R1 (Fig. 3C) probes we also detected protected fragments that fit the criteria for alternate exons. Thus, within the region between XhoI and ApaI, we found evidence for exonic regions of 152 and 50 bases in length (Fig. 3B). Between ApaI and EcoRI, we found evidence for an exonic region of roughly 55 bases in length (Fig. 3C). The region between EcoRI and exon 2 showed no evidence of containing exonic sequences (not shown).

Cloning TBP mRNA 5’ End Variants from Liver and Testis—

Our data indicated that additional promoters and alternate first exons were used in testis. Since the complexity observed by RNase protection suggested that there were multiple “missing exons,” we wished to perform an exhaustive search that could recover all possible mRNA 5’ end variants. Therefore, a sensitive ligation-mediated method for amplifying and cloning the 5’ ends of cDNAs was used to clone TBP cDNA 5’ ends using rat testis, mouse testis, and mouse liver RNA preparations. This method is analogous to a “RACE”; however, it is more efficient at recovering rare cDNA ends. Of 144 clones, 74% contained TBP cDNAs. Restriction digestion revealed that, although exons 1C and 2 contain no “types” (this does not include different initiation site variants within a cluster; see below) were identified. No species-specific differences in exon or initiation site usage were observed between rat and mouse. Of the six first exons recovered, three
(1C, 1D, and 1E) were recovered numerous times; the other three were cloned only once each. However, as these latter three first exons were precisely spliced onto exon 2, we consider them rare, but bona fide, alternate first exons.

From liver, we analyzed 16 TBP-containing clones; 10 contained exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below). From testis, we sequenced 72 TBP-containing clones, and 5 clones that initiated at an internal position in exon 1E (see Fig. 8).

Clones with no exon designation were not sequenced. Footnotes are as follows: 1 clone 32 was a mixed colony containing one clone each with exon 1C and 1D, and one alternate first exon (designated * in Fig. 6), which was in turn spliced to the 3' end of exon 1E, and finally to exon 2.

Of these, 31 clones (43%) were from mRNAs that did not contain a TBP insert, including both clones with no insert and those with inserts not having non-primer-encoded TBP exon 2 sequence. Clones with no exon designation were not sequenced. Footnotes are as follows: 1 clone 32 was a mixed colony containing one clone each with exon 1C and 1D, and one alternate first exon (designated * in Fig. 6), which was in turn spliced to the 3' end of exon 1E, and finally to exon 2; 4 clones that used the alternate upstream splice donor site in exon 1D; 2 clones that initiated at an internal position in exon 1D (see Fig. 8); 1 clone 74 contained the 5' site in exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below). From testis, we sequenced 72 TBP-containing clones, and 5 clones that initiated at an internal position in exon 1E (see Fig. 8).

(1C, 1D, and 1E) were recovered numerous times; the other three were cloned only once each. However, as these latter three first exons were precisely spliced onto exon 2, we consider them rare, but bona fide, alternate first exons.

From liver, we analyzed 16 TBP-containing clones; 10 contained exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below). From testis, we sequenced 72 TBP-containing clones, and 5 clones that initiated at an internal position in exon 1E (see Fig. 8).

Clones with no exon designation were not sequenced. Footnotes are as follows: 1 clone 32 was a mixed colony containing one clone each with exon 1C and 1D, and one alternate first exon (designated * in Fig. 6), which was in turn spliced to the 3' end of exon 1E, and finally to exon 2; 4 clones that used the alternate upstream splice donor site in exon 1D; 2 clones that initiated at a downstream position in exon 1D (see Fig. 8); 1 clone 74 contained the 5' site in exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below).

From testis, we sequenced 72 TBP-containing clones, and 5 clones that initiated at an internal position in exon 1E (see Fig. 8).

Clones with no exon designation were not sequenced. Footnotes are as follows: 1 clone 32 was a mixed colony containing one clone each with exon 1C and 1D, and one alternate first exon (designated * in Fig. 6), which was in turn spliced to the 3' end of exon 1E, and finally to exon 2; 4 clones that used the alternate upstream splice donor site in exon 1D; 2 clones that initiated at a downstream position in exon 1D (see Fig. 8); 1 clone 74 contained the 5' site in exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below).

From liver, we analyzed 16 TBP-containing clones; 10 contained exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below). From testis, we sequenced 72 TBP-containing clones, and 5 clones that initiated at an internal position in exon 1E (see Fig. 8).

Clones with no exon designation were not sequenced. Footnotes are as follows: 1 clone 32 was a mixed colony containing one clone each with exon 1C and 1D, and one alternate first exon (designated * in Fig. 6), which was in turn spliced to the 3' end of exon 1E, and finally to exon 2; 4 clones that used the alternate upstream splice donor site in exon 1D; 2 clones that initiated at a downstream position in exon 1D (see Fig. 8); 1 clone 74 contained the 5' site in exon 1C and 6 contained exon 1E (63 and 37%, respectively; see below).
Testis-specific TBP mRNA 5' End Heterogeneity

**Fig. 6. Gene structure, initiation site usage, and splicing patterns for the tbp gene in liver and testis.** At top is shown a map of exon regions of the TBP gene between the SacI site 700 base pairs upstream of exon 1C and the BglII site in exon 2. The entire region depicted has been sequenced and submitted to GenBank. Note, the restriction enzyme sites shown are for reference to probes and sequences in other figures in this paper; several of the enzymes shown will also cut at one or more other positions in the region shown. Below, bent arrows above the exons represent regions of transcription initiation; arrow thickness suggests relative rates of initiation at each region. In liver, only transcripts initiated at exon 1C and spliced onto exon 2 could be detected by RNase protection. However, cDNA clones containing exon 1E/2 fusions were obtained from liver, making it possible that these exist at a low level in liver (see text). Therefore exon 1C/2 fusions are estimated to represent greater than 95% of the TBP mRNA in liver. In testis, the eight mRNA types depicted as well as two others containing unmapped first exons (see text) were identified. The relative abundance of each mRNA type in testis as determined by RNase protection is depicted at the right.

containing exon 1B were too rare to detect in either testis or liver (less than 0.1 amol/µg RNA; not shown). Quantitative comparison of liver RNA to serial dilutions of testis RNA by RNase protection revealed the relative testis specificity of each exon type (Fig. 5A). A probe to exons 2 and 3 confirmed that overall TBP mRNA levels are 25-fold higher per equal DNA equivalent of tissue in testis (top panel). mRNA containing exon 1C was only 4-fold more abundant in testis than in liver. In contrast, although mRNAs containing exons 1D or 1E could not be detected in liver, dilutions of testis RNA indicated that our assay had the ability to detect 1/125 of the testis-specific signal. Thus, up-regulation of mRNAs containing exons 1D or 1E in testis is greater than 125-fold in magnitude. Similar RNase protection comparisons for all three first exons in testis, liver, brain, lung, and thymus confirmed that exons 1D and 1E were testis-specific (not shown). In a final attempt to detect exon 1D-containing TBP mRNA in liver, we used a reverse transcriptase-mediated polymerase chain reaction (RT-PCR) assay (Fig. 5B). The results showed that, whereas exon 1C/2 mRNAs could be detected in liver- and testis-polyadenylated mRNA samples and in liver, testis, and brain total RNA samples, exon 1D/2 mRNAs could only be detected in the testis samples. In conclusion, exons 1D and 1E appear to be truly testis-specific, whereas exon 1C is used in all tissues and is up-regulated modestly (4-fold above liver or 20-fold above thymus levels) in testis.

**tbp Initiation Sites and Promoter Sequences**—The ligation-mediated cDNA amplification method allows precise identification of the 3' nucleotide of the first-strand cDNA synthesis product. Thus, assuming that the template RNA was intact and the reverse transcriptase was processive, the sequence of the cDNA clone allows precise identification of the transcription initiation nucleotide. The initiation sites thus identified on 5 clones containing exon 1C confirmed that this exon initiates at two clusters of sites as suggested by RNase protection (Fig. 1). Exons 1D and 1E were also found to initiate at multiple sites. For these two exons, two zones of initiation separated by 130 or 55 bases, respectively, were identified (Fig. 6 and see below).

Although the sequence of the cDNA 5' end clones presented above allowed precise identification of the transcription initiation sites of individual mRNAs, it was possible that, after amplification and cloning, individual clones might have become either over- or under-represented in the population. Thus, we wished to confirm the relative frequencies of initiation site usage by direct primer-extension analysis. Primer extension on poly(A)^-selected rat testis mRNA using either of two primers specific for sequences in exon 2 confirmed the initiation sites (Fig. 7). Samples containing poly(A)^-selected liver RNA confirmed that the signals were testis-enriched, as expected for TBP mRNA-specific signals. The sizes of the individual products corresponded to those predicted from the cDNA 5' end clones in Fig. 4 and to the sizes of RNase protection products mapped in Figs. 1 and 3. The start sites are diagrammed in Fig. 8.

The regions upstream of all of the initiator exons are TATA-less, which is consistent with their having clustered initiation sites (8). Roughly 35 bases upstream from the major site of internal initiation for exon 1E is the sequence TATAT, which bears some resemblance to a TATA box (Fig. 8) (9). However, the imprecise initiation noted in this region is reminiscent of TATA-less promoters, and thus it is likely that this sequence is not sufficient to direct transcription initiation to a single nu-
cleotide. Comparison of the entire \textit{tbp} 5' region with published data bases revealed no striking similarities (excluding repetitive sequences) with other published sequences. Indeed, with the exception of an SP1 binding site in the exon 1C promoter (6), the sequences upstream of all of the \textit{tbp} initiation sites bear little obvious resemblance to previously identified promoters (see below).

**DISCUSSION**

TBP protein is required for all nuclear transcription initiation and, thus, is a fundamental component of all cells. Unexpectedly for a gene with such ubiquitous and apparently well-defined functions, the \textit{tbp} gene is highly overexpressed in the early haploid stages of spermatogenesis. Because spermatogenesis is a very complex process that involves interactions between many cell types (10), we suspect that a full understanding of the roles of TBP in spermatogenesis can only come through manipulation of TBP expression in animals. Identification of the spermatid-specific regulatory mechanisms for the \textit{tbp} gene is a requisite step on the path to manipulating TBP expression in transgenic animals and, ultimately, to understanding the roles of TBP overexpression in spermatogenesis.

A molecular analysis of the 5' end of the \textit{tbp} gene is presented. Our results show that liver uses almost exclusively a single promoter/first exon and produces predominantly a single species of TBP mRNA. In contrast, testis initiates transcription at no fewer than 3 major and 3 minor first exons and produces at least 10 different TBP mRNA types (6 abundant and 4 rare). This work precisely localizes the testis-specific transcription initiation sites and testis-specific exons on the \textit{tbp} gene. This indicates where to focus a search for the spermatid-specific regulatory mechanisms, and more importantly, it indicates which sequences to target in a rational mutagenesis of putative spermatid-specific regions of the \textit{tbp} gene in transgenic animals.

**Testis-specific TBP mRNAs—**What is the purpose for the testis-specific TBP mRNA heterogeneity? One possibility was that the different mRNAs might generate distinct protein products. However, only mRNAs initiated at the more upstream region of exon 1E have a translational start codon that could produce an alternate protein (72 amino acids long, not in the TBP reading frame), and the ATG for this polypeptide is in a poor context for initiation (11). Thus, we predict that all 10 types of TBP mRNA characterized here only give rise to normal TBP protein. A similar situation has been reported for the cytochrome \(c\) gene, which produces a testis-specific mRNA that is predicted to use the same open reading frame as somatic cytochrome \(c\) mRNA (12). It is possible that the upstream ATGs noted on some of the TBP mRNA 5' end variants might play a role in translational regulation (see below).

A second possibility is that each individual promoter is favored by a specific subset of nuclei types. Indeed, somatic tissue nuclei appear to use exclusively exon 1C for initiation. It was possible, for example, that exons 1C, 1D, and 1E were each preferred by germ cells at a specific developmental stage. However, we have analyzed the developmental onset of accumulation of mRNAs containing exons 1C, 1D, and 1E in testis, and we cannot detect differential temporal accumulation of any of these mRNAs (not shown).

A third possible reason for generating different mRNA species might be to allow different post-transcriptional regulation of TBP protein accumulation. Recently, a study on the \textit{copper-zinc superoxide dismutase} gene showed that, like for \textit{tbp}, testis-specific expression involves recruitment of two testis-specific promoters in addition to the somatic promoter (13). Of the three different mRNA types that accumulate in testis, one (arising from one of the testis-specific promoters) is sequenced as ribonucleoprotein particles; the other two are predominantly polysomal (13). A fraction of the TBP mRNA in whole testis is also sequenced as ribonucleoprotein particles (not shown).
Presumably, these stored mRNAs are translated at a later time. Different sequences in the 5′-nontranslated leader of TBP mRNAs resulting from different first exon usage, possibly including the presence of upstream ATGs, might target mRNAs to be either translated immediately or stored for later use.3

The numerous splicing variations that arise in testis are another curiosity. We can currently find no rational explanation for why transcripts initiating at the upstream region of exon 1D must use two splice donor sites (positions 168 and 1179 in Fig. 8). Spermatids might simply provide a "sloppy" or "promiscuous" splicing environment, such that otherwise cryptic splice donor and acceptor sites can be used. Indeed, numerous genes have been shown to exhibit splicing patterns in testis that are not found in somatic tissues (e.g., Refs. 14–16). In the case of TBP mRNA, where all of the alternatively spliced mRNAs are expected to produce the same protein, such alternate splicing would not be deleterious.

3 E. E. Schmidt and U. Schibler, submitted for publication.

---

**Fig. 8.** Sequences and initiation sites for *tbp* exons. For convenience, all sequences shown are from TBP clones derived from a mouse 129 ES cell genomic library. Initiation site data are from MORO mouse and Sprague-Dawley rat tissues. With the exception of the *Pvu*II site polymorphism in exon 1D (see below), strain- and species-specific differences in exonic sequences were negligible. Exonic sequences are shown by capital letters, arbitrarily assigning position +1 to the most upstream cDNA that was cloned and sequenced for each initiation exon. For each initiation exon, 200 bases of upstream promoter sequence are shown. Initiation sites are indicated by vertical lines above the sequence; line thickness reflects the number of independent cDNA clones initiating at each site that were sequenced to the initiation site. For exon 1C, the initiation sites determined for sequenced clones map to the two clusters of initiation sites predicted in Figs. 1 and 2. The underlined region upstream of exon 1C represents the putative SP1 binding site (6). The 3′ end of each first exon is labeled with an inverted black triangle and with the position of the splice site relative to +1 for that exon. In each case, the first 10 bases of exon 2 are shown; these are underlined by a stippled box. For exon 1D, the alternative splice donor site for clones 31, 32, 34, and 74 is shown. The *Pvu*II site noted by restriction mapping and sequencing of exon 1D cDNA clones from both MORO mice and Sprague-Dawley rats (Fig. 4) did not exist in the mouse 129 cell genomic clone; this sequence polymorphism is indicated. For exon 1E, the TATA-like sequence upstream of the internal initiation zone is boxed. The alternate splice acceptor site for clone 74 is indicated. At bottom, the sequence of the alternate intervening exon for clone 74 is listed; this sequence was spliced between the exon 1D alternate donor and the exon 1E alternate acceptor sites.
Testis-specific Transcription of the tbp Gene—Comparison to public data bases indicates that the tbp gene 5′ end exhibits no notable sequence similarities to previously described spermatid-specific promoters. A search for putative transcription factor recognition sequences revealed little evidence of what factors might be regulating this gene. Thus, although the entire 5′ region of tbp is peppered with putative recognition sites for the “testis-determining factor,” SRY (frequency, 1 site per 168 bases over 5416 base pairs; 17), a similar frequency (1 site in 281 bases) was found in the 111,400-base pair 0–2.4-min region of the Escherichia coli genome. Thus, the concentration of putative SRY sites in the tbp gene does not appear to differ significantly from that found in an arbitrary and physiologically irrelevant sequence.

Putative binding sites were noted for two other spermatid-enriched transcription factors, a site for the SRY-related Sox-5 protein (18) and several putative cAMP-response elements (CREs; 19) which are the binding sites for CREM and CREB. The one putative Sox-5 site is located roughly 850 base pairs downstream of the exon 1E promoter. Five putative CREs are found as follows: one roughly 165 base pairs upstream of exon 1B, one at the major point of initiation for exon 1C, one roughly 220 base pairs, and one roughly 35 base pairs upstream of exon 1E, and one roughly 200 base pairs downstream of the promoter of exon 1E. CREB is expressed in many testis cell types, whereas CREM is predominantly restricted to the germ cells (20). Two recent reports show that CREM-deficient mice cannot complete spermatogenesis (21, 22). However, in one of these reports, TBP mRNA was used as a control for an RNase protection experiment, and its expression did not appear diminished in the CREM-deficient mice (22). Thus, it appears that the major CRE-binding protein in spermatids is dispensable for TBP overexpression. Binding sites for other putative spermatid-specific transcription factors such as Tet-1 (23) or Zfy-1 (24) have not yet been selected from random sequences, and thus remain largely undefined. It remains possible that, despite a lack of obvious sequence identity, the tbp gene might share testis-specific regulatory signals with other spermatid-specific genes. Accurate delineation of what cis-regulatory sequences are important for testis-specific TBP overexpression will require a functional analysis of these sequences in transgenic animals.

A final point to consider is the reason for the rare promoter/first exons in testis. The mRNAs arising from these exons in testis are at least 10-fold less abundant than TBP mRNA in somatic cells and are predicted to yield the exact same protein product. Thus, we suspect that these transcripts do not have a unique function. Rather, they might result from promiscuous transcription initiation in spermatids. Spermatids contain greatly elevated levels of all measured components of the basal RNA polymerase II transcription machinery (2). A model has been proposed for how increased levels of the transcription machinery should decrease promoter stringency and thus promote transcription initiation at sequences that would otherwise not be recognized as promoters (25). This model might explain many cases of spermatid-specific gene expression, including the rare TBP mRNAs.

In summary, testis-specific up-regulation of the tbp gene involves recruitment of two very strong testis-specific promoter/first exons. This work will be important for further resolving the signals regulating spermatid-specific TBP expression and ultimately for understanding the reason why spermatids contain 1000-fold more TBP mRNA molecules per cell than do somatic cell types.

Acknowledgments—We thank P. Fonjallaz for assistance with PCR, D. Lavery for critically reading the manuscript, and N. Roggli for photography.

REFERENCES
1. Hernandez, N. (1993) Genes Dev. 7, 1291–1308
2. Schmidt, E. E., and Schibler, U. (1995a) Development 121, 2373–2383
3. Schmidt, E. E., and Schibler, U. (1995b) J. Cell Biol. 126, 467–483
4. Nebel, B. R., Amarose, A. P., and Hackett, E. M. (1981) Science 134, 832–833
5. Sumita, K., Makino, Y., Katoh, K., Kishimoto, T., Muramatsu, M., Mikoshiba, K., and Tamura, T.-A. (1993) Nucleic Acids Res. 21, 2769
6. Ohbayashi, T., Schmidt, E. E., Makino, Y., Kishimoto, T., Nabeshima, Y.-I., Muramatsu, M., and Tamura, T.-A. (1996) Biochem. Biophys. Res. Commun. 235, 275–280
7. Gross, M. K., Kainz, M. S., and Merrill, G. F. (1987) Dev. Biol. 122, 439–451
8. Lu, J., Lee, W., Jiang, C., and Keller, E. B. (1994) J. Biol. Chem. 269, 5391–5402
9. Buccheri, P., and Trifinov, E. N. (1986) Nucleic Acids Res. 14, 10009–10026
10. Hecht, N. P. (1995) Dev. Genet. 16, 95–103
11. Kozak, M. (1986) Cell 44, 283–292
12. Hake, L. E., and Hecht, N. B. (1993) J. Biol. Chem. 268, 4788–4797
13. Gu, W., Morales, C., and Hecht, N. B. (1995) J. Biol. Chem. 270, 236–243
14. Fox, N.-C., Funkhouser, J. M., Carter, D. A., and Murphy, D. (1994) J. Biol. Chem. 269, 659–667
15. Foulkes, N. S., Mellstrom, B. Benusiglio, E., and Sasson-Corsi, P. (1992) Nature 355, 80–84
16. Mettus, R. V., Litvin, J., Wali, A., Toscani, K., Latham, K., Hatton, K., and Reddy, E. P. (1994) Oncogene 9, 3077–3086
17. Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1994) EMBO J. 13, 6115–6124
18. Denny, P., Swift, S., Connor, F., and Ashworth, A. (1992) EMBO J. 11, 3705–3712
19. Benbrook, D. M., and Jones, N. C. (1994) Nucleic Acids Res. 22, 1463–1469
20. West, A. P., Sharpe, R. M., and Saunders, P. T. (1994) Biol. Reprod. 50, 869–881
21. Nantel, F., Monaco, L., Foulkes, N. S., Lasnequier, D., LeMeur, M., Herrnkind, K., Dierich, A., Parvinen, M., and Sashon-Corsi, P. (1996) Nature 380, 159–162
22. Blendy, J. A., Kaestner, K. H., Weinhauer, G. F., Nieschlag, E., and Schultz, G. (1996) Nature 380, 162–165
23. Tamura, T., Makino, Y., Mikoshiba, K., and Muramatsu, M. (1992) J. Biol. Chem. 267, 4327–4332
24. Zambrowicz, B. P., Zimmermann, J. W., Harendza, C. J., Simpson, E. M., Page, D. C., Brinster, R. L., and Palmer, R. D. (1994) Development 120, 1549–1559
25. Schmidt, E. E. (1996) Curr. Biol. 6, 768–769

4 TFSEARCH version 1.3, http://www.tokyo-center.genome.ad.jp/SIT/TFSEARCH.html.