

Identification of Two Sox17 Messenger RNA Isoforms, with and without the High Mobility Group Box Region, and Their Differential Expression in Mouse Spermatogenesis

Yoshiakira Kanai,* Masami Kanai-Azuma,* Toshiaki Noce,[§] Takaomi C. Saido,[‡] Toshihiko Shiroishi,^{||} Yoshihiro Hayashi,[†] and Kazumori Yazaki*

*Department of Ultrastructural Research and [†]Department of Molecular Biology, the Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113; [§]Department of Molecular Reproduction and Development, Mitsubishi Kagaku Institute of Life Sciences, Machida-shi, Tokyo 194; ^{||}Department of Cell Genetics, National Institute of Genetics, Yata, Mishima 411; and [‡]Department of Veterinary Anatomy, the University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Abstract. Two different mRNA isoforms of the mouse *Sox17* gene were isolated from adult mouse testis cDNAs. One form (referred to as form Sox17) encodes an Sry-related protein of 419 amino acids containing a single high mobility group box near the NH₂ terminus, while the other form (referred to as form t-Sox17) shows a unique mRNA isoform of the *Sox17* gene with a partial deletion of the HMG box region. Analysis of genomic DNA revealed that these two isoforms were produced at least by alternative splicing of the exon corresponding to the 5' untranslated region and NH₂-terminal 102 amino acids. RNA analyses in the testis revealed that form Sox17 is expressed in spermatogonia, and the expression clearly declines from the early pachytene spermatocyte stage onward. In contrast, expression of form t-Sox17 began at the pachytene spermatocyte stage and was highly accumulated in round spermatids. Protein analyses revealed that t-Sox17 isoforms, as well as

Sox17 isoforms, were translated into the protein products in the testis, although the amount of t-Sox17 products is lower in comparison to the high accumulation of t-Sox17 mRNA. By the electrophoretic mobility-shift assay and the random selection assay using recombinant Sox17 and t-Sox17 proteins, Sox17 protein is a DNA-binding protein with a similar sequence specificity to Sry and the other members of Sox family proteins, while t-Sox17 shows no apparent DNA-binding activity. Moreover, by a cotransfection experiment using a luciferase reporter gene, Sox17 could stimulate transcription through its binding site, but t-Sox17 had little effect on reporter gene expression. Thus, these findings suggest that Sox17 may function as a transcriptional activator in the premeiotic germ cells, and that a splicing switch into t-Sox17 may lead to the loss of its function in the postmeiotic germ cells.

THE *Sry* gene on human and mouse Y chromosomes is the testis-determining gene that induces differentiation of the genital ridge into the testicular pathway (Sinclair et al., 1990; Gubbay et al., 1990; Koopman et al., 1991). It potentially encodes a DNA-binding protein containing the high mobility group (HMG)¹ box that is present in several transcription factors such as T cell-specific factor TCF-1 (van de Wetering et al., 1991) and RNA

polymerase I transcription factor UBF (Jantzen et al., 1990). Because the Sry protein synthesized in vitro binds to double-stranded DNA in a sequence-dependent manner (Harley et al., 1992, 1994), it is believed to function in testis determination by regulating the expression of other specific genes. Until now, several genes encoding Sry-related HMG box (termed as *Sox* gene) have been identified in mouse and human cDNA, and a member of this family has also been isolated in a number of other vertebrate and invertebrate species (Denny et al., 1992a,b; Chardard et al., 1993; Wright et al., 1993; Gozé et al., 1993). The amino acid sequences of Sox family genes are known to be highly conserved from humans to *Drosophila* (Denny et al., 1992a), and such strong evolutionary conservation suggests that Sox family genes may play important roles in the differentiation and development of the testis and the other cell types and tissues in many species. For example, *Sox1*, *Sox2*, *Sox3*, and *Sox11* are known to

Please address all correspondence to Yoshiakira Kanai, D.V.M., Ph.D., Department of Ultrastructural Research, The Tokyo Metropolitan Institute of Medical Science, 18-22 Honkomagome 3-chome, Bunkyo-ku, Tokyo 113, Japan. Tel.: 81-03-3823-2101 (ext.5163); Fax: 81-03-3823-2965.

1. *Abbreviations used in this paper:* DIG, digoxigenin; EMSA, electrophoretic mobility-shift assay; GST, glutathione S-transferase; HMG, high mobility group; ORF, open reading frame; RT, reverse transcription; UTR, untranslated region.

be expressed in the developmental nervous system (Stevanović et al., 1993; Uwanogho et al., 1995; Jay et al., 1995; Kamachi et al., 1995). *Sox4* is expressed in T lymphocytes and in some pre-B lymphocytes, and Sox4 protein mediates the enhancer effects of the AACAAAG motif in lymphocytes (van de Wetering et al., 1993). Recently, the human *SOX9* gene was isolated from a translocation chromosome breakpoint of a sex-reversed patient with campomelic dysplasia, and mutation analysis and expression patterns indicated that *Sox9* may play a critical role both in testicular differentiation and skeletal development (Foster et al., 1994; Wagner et al., 1994; Wright et al., 1995).

In the mature adult testis, several Sox family genes are expressed. *Sox5* is expressed stage specifically during spermatogenesis (Denny et al., 1992b). *Sry* is also expressed at a high level in the adult testis (Koopman et al., 1991; Rossi et al., 1993), and the most abundant *Sry* transcripts in adult testes are present as circular RNA molecules (Capel et al., 1993), although their functions in spermatogenesis are uncertain. *Sox4* (van de Wetering et al., 1993), *Sox6* (SOX-LZ) (Denny et al., 1992b; Takamatsu et al., 1995; Connor et al., 1995), and *Sox9* (Foster et al., 1994; Wagner et al., 1994) are also reported to be expressed in the adult testis. Although functions of these Sox family genes in the adult testis remain elusive, such testicular expression leads us to the postulation that several members of this gene family may exert an important effect on mammalian spermatogenesis.

In the present study, we isolated two different mRNA isoforms of the mouse *Sox17* gene (form Sox17 and t-Sox17) from adult mouse testis cDNAs. Form Sox17 encodes an Sry-related protein of 419 amino acids containing a single HMG box near the NH₂ terminus, while form t-Sox17 shows a unique mRNA isoform of the *Sox17* gene with a partial deletion of the HMG box region. Moreover, we examined the expression patterns of these two Sox17 isoforms in the testis and revealed that the isoform with an intact HMG box region was expressed in spermatogonia, and that such expression decreased at the pachytene spermatocyte stage. In contrast, expression of the unique isoform without the HMG box region began at the pachytene spermatocyte stage and was accumulated in round spermatids. Protein analyses revealed that t-Sox17 isoforms, as well as Sox17 isoforms, were translated into the protein products in the testis, although the amount of t-Sox17 products is lower in comparison to the high accumulation of t-Sox17 mRNA. To test the functional difference between Sox17 and t-Sox17, we examined their DNA-binding activity by the electrophoretic mobility-shift assay (EMSA) and the transactivation activity by a cotransfection experiment using a luciferase reporter gene. By EMSA, t-Sox17 showed no apparent DNA-binding activity, but Sox17 can specifically bind the AACAAAT motif. Moreover, by a cotransfection experiment, Sox17 could stimulate transcription through its binding site, but t-Sox17 had little effect on reporter gene expression. Thus, the *Sox17* gene is the first Sox family gene that is found to produce multiisoforms with different expression profiles and DNA-binding activities. The biological significance of the differences between these two isoforms in spermatogenesis will be discussed.

Materials and Methods

Isolation of Sox17 cDNAs and Genomic DNA and Sequencing

To isolate Sry-related cDNA expressed in the testis, a mouse testis cDNA library constructed with λ gt11 (Clontech Laboratories, Palo Alto, CA) was screened at high stringency with a Sox9 HMG box probe. The Sox9 HMG box fragments were isolated from testis cDNAs by the degenerate PCR approach, as previously described (Denny et al., 1992a; Wright et al., 1993). As a result, two independent positive clones that coded a novel isoform of Sox17 cDNA (referred to as form t-Sox17) were isolated. Because these two clones encoded a unique isoform of the *Sox17* gene with a partially deleted HMG box region, we further isolated the Sox17 cDNA (referred to as form Sox17) encoding an intact HMG box from testis cDNAs using an LA-PCR kit (Takara Biomedicals, Kyoto, Japan) in combination with two sets of primers (forward primer: 5'-ATGGCCCACTCACACT-GCTGGCGGGT-3' [+375 to +400] \times reverse primer: 5'-CCAAC-CGCTTGCGTTCGTCITTTGGC-3' [+884 to +908]; forward primer: 5'AGGCTAGCTTCCGATCCCCTGCCTC-3' [+564 to +587] \times reverse primer: 5'-GCTTCTGGCCCTCAGGTCCGGTCCGGCAA-3' [+1913 to +1940]). Moreover, the mouse genomic library constructed with λ FIX II (Stratagene, La Jolla, CA) was screened, and three independent clones were isolated. Inserted or amplified DNA fragments were subcloned into pBluescript KS⁺ (Stratagene). Nucleotide sequences were determined by the dideoxy chain termination method (Sanger et al., 1977) using a BcaBEST sequencing kit (Takara Biomedicals).

Generation of Probes

For Northern and Southern blot hybridization, ApaI-digested (+1249 to +1482; probe A in Fig. 2 a) and EcoRI-digested (+1750 to +2375; probe B) fragments of t-Sox17 were used as probes that recognize both isoforms. DNA fragments corresponding to the spliced exon (+564 to +863; probe C), which were isolated by PCR, were used as the probe specific for the Sox17 form. For the reverse transcription (RT) PCR analysis, the fragment digested with SacI and ApaI (+336 to +1248 of t-Sox17) was used for detection of both Sox17 and t-Sox17 cDNAs. The isolated DNA fragments were labeled with [α -³²P]dCTP (3,000 Ci/mmol) by a Megaprime DNA labeling system (Amersham, Buckinghamshire, UK). For in situ hybridization analysis, the DNA fragments corresponding to probe B or C were subcloned into pBluescript. Each clone was linearized with the appropriate restriction enzyme, and the sense and antisense RNA probes were generated by in vitro transcription using digoxigenin (DIG)-UTP with T3 and T7 RNA polymerase (Boehringer Mannheim Biochemica, Mannheim, Germany). The resulting sense and antisense DIG-labeled RNA probes were reduced to ~150-bp fragments by limited alkaline hydrolysis before hybridization to tissue sections.

Southern Blot Hybridization

Genomic DNAs were prepared from the liver of adult male mice and digested with BamHI, EcoRI, HindIII, or PstI. 10 μ g of each DNA preparation were electrophoresed in 0.6% agarose-TBE gels and transferred to a nylon membrane (Hybond N; Amersham). Blots were hybridized with appropriate ³²P-labeled DNA probes in a solution containing 50% formamide, 5 \times SSC, 5 \times Denhardt's, 1% SDS, 0.2 M sodium phosphate (pH 6.8), and 200 μ g/ml denatured salmon sperm DNA (ssDNA) at 42°C for 12–14 h. Filters were finally washed with 0.1 \times SSC–0.1% SDS at 65°C for 1 h and autoradiographed with an x-ray film (X-OMAT-AR; Eastman Kodak, Rochester, NY).

Chromosomal Mapping of Sox17 Gene

Interspecific F1 hybrids of C3H/HeJ and a *Mus spretus*-derived strain, SEG, were back-crossed to C3H/HeJ. 50 interspecific back-cross progeny were scored for the segregation of RFLP of the *Sox17* gene and microsatellite markers, D1Mit1, D1Mit4, D1MitMit11, D1Mit12, D1Mit14, and D1Mit17. Map manager v2.5 was used to analyze the data generated in this study.

Northern Blot Hybridization and RT-PCR Analyses

Total RNAs were extracted from various tissues by the guanidinium thiocyanate method (Chirgwin et al., 1979). 25 μ g of each extract were dena-

tured in formamide-formaldehyde buffer, electrophoresed in 1% formaldehyde-agarose gels, and transferred to a nylon membrane. Blots were hybridized with appropriate ³²P-labeled DNA probes and washed with 0.1× SSC–0.1% SDS at 65°C as described above. For RT-PCR analysis, poly(A)⁺ RNAs were prepared from total RNA by oligo(dT) cellulose chromatography, and 0.2 μg of each poly(A)⁺ RNA preparation were reverse transcribed using random hexamers as primers with a RNA PCR kit following the instructions of the manufacturer (Perkin Elmer Cetus, Norwalk, CT). A reverse transcriptase-free reaction was performed as a control experiment. t-Sox17 and Sox17 cDNAs were separately amplified by the primers as follows: forward primer of form t-Sox17: 5'-ATGGC-CCACTCACACTGCTGGCGGGT-3' [+375 to +400], forward primer of form Sox17: 5'-GCCAAAGACGACGCAAGCGGT-3' [+884 to +905]; reverse primer of both forms: 5'-TCATGCGCTTCACTGCTTG-3' [+1075 to +1094]. 5 μl of each sample were electrophoresed with 3% agarose gel and analyzed by Southern blot hybridization, as described above.

In Situ Hybridization

Adult testes were fixed in Bouin's solution for 4 h, dehydrated in ethanol, cleared in xylene, and then routinely embedded in paraffin. Paraffin sections of 5-μm thickness were cut and subjected to in situ hybridization, as described in our previous study (Noce et al., 1992a,b). In short, deparaffinized sections were pretreated with 0.3% Triton X-100 in 10 mM PBS and 20 μg/ml proteinase K in Tris-HCl buffer (pH 7.5) containing CaCl₂, and then hybridized with DIG-labeled RNA probes in a solution containing 50% formamide, 10% dextran sulfate, 5× SSC, 1× Denhardt's, 1% SDS, 100 μg/ml heparin, 10 mM DTT, and 1 mg/ml denatured tRNA and ssDNA at 45°C for 12–16 h. After treatment with RNase A (20 μg/ml; Sigma Immunochemicals, St. Louis, MO) at 37°C for 1 h, the specimens were finally washed twice with 0.1× SSC at 65°C for 1 h. The signals were detected by an immunological method using alkaliphosphatase-conjugated anti-DIG antibody and nitro blue tetrazolium as the chromogen (Boehringer Mannheim Biochemica).

In Vitro Translation

RNA was synthesized from linearized pBluescript/Sox17 (+564 to +1940) and pBluescript/t-Sox17 (+1 to +2375 of t-Sox17 cDNA) by in vitro transcription with T3 and T7 RNA polymerase. After DNase I treatment, each 1 μg RNA was translated in a 50-μl reaction in the presence of 1 μM [³⁵S]methionine (1,200 Ci/mmol) in a rabbit reticulocyte lysate system according to the instructions of the manufacturer (Amersham). Each protein sample was analyzed by SDS-PAGE and autoradiographed.

Antibody Production, Immunoblotting, and Immunofluorescence

Synthetic peptides corresponding to amino acids 25–39 of the predicted Sox17 NH₂-terminal region (AGLGPCWAESLSPL) or to amino acids 403–419 of the predicted Sox17 COOH-terminal region (VVSDAS-SAVYYCNYPD) were conjugated to KLH (Calbiochem, La Jolla, CA) via a heterobifunctional cross-linker, *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (Pierce Chemical Co., Rockford, IL), as previously described (Saido et al., 1992, 1995). Rabbits were immunized with antigen conjugates (0.5–1 mg/animal) six times at 2–3-wk intervals. Specific antibodies were affinity-purified on the haptenic peptide immobilized on Affi-Gel 10 (Bio-Rad Laboratories, Hercules, CA). We also raised antisera to recombinant GST-Sox17 protein (see below). Rabbits were immunized with 0.2–0.4 mg protein/animal six times at 2–3-wk intervals. Antibodies to glutathione *S*-transferase (GST) were removed by absorption with GST-agarose beads before use.

The COS cells were transfected with pCDM, pCDM/Sox17, or pCDM/t-Sox17 (each 30 μg/6-cm dish, described below), and, after 48 h, the cells were washed with PBS several times. The cells were dissolved in the sample buffer, and each protein sample (30 μg/lane) was used for SDS-PAGE (12.5%) and immunoblotting, as described previously (Noce et al., 1992b). Moreover, 20 testes were collected from mature male mice. After a removal of tunica albuginea, the seminiferous tubules were gently loosened and washed with buffer A (10 mM Tris-HCl, 5 mM EDTA, pH 7.5) containing 250 mM sucrose. The masses of the seminiferous tubules were homogenized in 10 ml buffer A containing 0.5 mM PMSF, 1 ng/ml leupeptin, 5 ng/ml aprotinin, 1 mM benzamide, and 1 ng/ml pepstatin A, and were then sonicated for 30 s. After the extracts were centrifuged at 100,000 g for

30 min, the supernatants containing 9–11 mg of proteins were applied to a DEAE ion exchange column (Tosoh DEAE-Toyopel 650S, 0.5 × 5 cm) equilibrated with buffer A. After the column was washed with buffer A, the bound proteins were eluted by a linear gradient from 0 to 400 mM NaCl (0.5 ml/min, 0.5 ml/fraction). The chromatographic procedures were carried out at 4°C in columns connected to and controlled by an FPLC system (Pharmacia Biotech, Uppsala, Sweden). Each column fraction was also subjected to SDS-PAGE (12.5%) and immunoblotting.

For immunohistochemical staining, we used the methanol-fixed transfected cells and methanol-fixed frozen sections (5-μm thickness) of adult mouse testes. They were first incubated with 1% BSA–5% normal goat serum–PBS for 10 min, and then incubated with anti-recombinant Sox17 antiserum (a dilution to 1:500) for 1 h. After a rinse in PBS, they were incubated with rhodamine-conjugated goat anti-rabbit IgG (diluted to 1:100; Tago, Burlingame, CA) for 45 min, washed again with PBS, and mounted in 90% glycerol–PBS. Specimens were observed with an Axiophoto photomicroscope equipped with an epifluorescence system (Carl Zeiss, Oberkochen, Germany). To check the specificity of the antibody for immunofluorescence, the sections or cells were incubated with the anti-recombinant Sox17 antiserum in the presence of GST-Sox17-agarose beads (10 μg/ml). Nonspecific staining was also checked by the incubation of the sections or cells with rhodamine-conjugated goat anti-rabbit IgG alone.

Preparation of GST Fusion Proteins of Sox17 and t-Sox17, EMSA, and Random Selection Protocol

Sox17 (+719 to +1750) and t-Sox17 (+1034 to +1750) DNA fragments were isolated by PCR from each cDNA clone and were inserted into the EcoRI site of pGEX-4T2 expression vector (Pharmacia Biotech). Each recombinant protein was produced in BL21 (DE3) host bacteria by IPTG induction and purified using glutathione-agarose beads (Smith and Johnson, 1988). The isolation of each recombinant protein was checked by Coomassie blue staining of SDS-polyacrylamide gels.

50 or 250 ng of each purified GST fusion protein was incubated with ³²P-labeled oligonucleotides (100 fmol) in 10 μl binding buffer (10 mM Hepes [pH 7.9], 60 mM KCl, 1 mM DTT, 1 mM EDTA, 1 μg/ml poly (dI-dC), and 12% glycerol) in the presence or absence of an appropriate competitor DNA for 30 min at room temperature. Then, the reactions were electrophoresed through a nondenaturing 6% acrylamide gel run in 1× TBE at 160 V for 1 h. The gels were dried and autoradiographed. The following oligonucleotides containing self-complementary regions were used as probes (Fujita et al., 1992): SCS4, G CAGAACA **AAGCGT** TAGCTCGCCTTGTCTG; SCS6, G CAGAACA **ATGGCGTTAGCTCGCCATTGTTCTG** (underlined sequences show self-complementary regions and bold ones show the motif sequences). These oligonucleotides were labeled by T4 polynucleotide kinase with [γ-³²P] ATP and diluted to 1 pmol/ml. They were boiled (at 100°C for 5 min) and self-annealed at room temperature for 30 min before addition to the reaction mixture. Oligonucleotides containing the consensus sequences of other transcriptional factors (SP1, NF-κB, CREB, and Oct1; Stratagene) were also used as competitor DNAs.

Moreover, sequences binding to GST-Sox17 or GST-t-Sox17 were selected from a pool of random DNA sequences. The random sequence library was prepared by oligonucleotides with the sequences CGA-GGATCCAGTTCAGAC(A/G/T/C)18TCATGGACTGAATTCGCA in combination with primers of the first and last 18 bases. The selection procedure followed the method of Wilson et al. (1993). PCR products from alternative cycles of selection were labeled by T4 polynucleotide kinase with [γ-³²P]ATP and examined by EMSA, as described above. After the last PCR amplification step (at the 7th selection), a fraction of the PCR product was digested with EcoRI and BamHI, cloned into pBluescript, and the individual oligonucleotides were sequenced.

Cell Culture, Transfection and Luciferase Activity Assay

As for effector plasmids of pCDM/Sox17 and pCDM/t-Sox17, the Sox17 (+564 to +1940) and t-Sox17 (+1 to +2375 of t-Sox17) DNA fragments isolated from each cDNA clone were separately inserted into pCDM8. Moreover, DNA fragments of Sox17 deletion constructs (pCDM/Sox17[d343-419], pCDM/Sox17[d295-419], and pCDM/Sox17[d164-419], which are designated by the amino acid numbers of deletion borders) were prepared by PCR using a forward primer (5'-CCAAGCTTAGCTTCCGATCCCTGCTCAGGGT-3' [+569 to +592]), in which the

cDNA in each effector plasmid and the number of H4 sites in the reporter plasmid were confirmed by sequencing. pEF-LacZ is used for the internal control of the transfection efficiency (gift from Drs. H. Hamada and Y. Saijoh, Tokyo Metropolitan Institute of Medical Science).

L929 cells were seeded 1 d before transfection. Transfection was performed by the DOTAP reagent (Boehringer Mannheim Biochemica) following the manufacturer's instructions. pCDM8 effector plasmid (0–5 μ g), luciferase reporter plasmid (1 μ g), and pEF-LacZ (0.1 μ g) were transfected into 0.5×10^5 cells on 1-cm dishes. After 48 h, the culture medium was removed, and the cells were washed with cold PBS several times. After addition of the lysis buffer (Toyo Ink MFG, Tokyo, Japan), a part of the cell extract was separately reacted with luciferase (Luciferase Assay System; Promega) or β -galactosidase substrate (LumiGAL Detection Kit; Clontech), and then each activity was measured by the TD-4000 Lumiphotometer (Futaba Medical, Tokyo, Japan). Each luciferase level was normalized to β -galactosidase level, and the relative luciferase level for pCDM8 transfected with pH4 \times 4-Luci was set at 1 as basal activity.

Results

Isolation of Sox17 cDNAs and Genomic DNA

To isolate the Sry-related gene expressed in the adult testis, a mouse testis cDNA library was screened at high stringency with a Sox9 HMG box probe, and two independent clones that encode a novel mRNA isoform of the *Sox17* gene were isolated. Sequence analysis by a GenBank DNA database search revealed that these clones were identical to mouse Sox17 cDNA in a partial sequence of the Sox17 HMG box region (Dunn et al., 1995), but they lacked the sequences corresponding to the upper half of the HMG box region. This unique clone contained a single open reading frame (ORF) of 873 bp after an initiation methionine located within the HMG box region, and it predicted a truncated 291-amino acid protein with a predicted molecular mass of 30.7 kD which lacked most of the DNA-binding domain (Fig. 1 and 2 a). To examine whether the clone without the HMG box region was a unique mRNA isoform of the *Sox17* gene, we isolated the genomic clones of the *Sox17* gene and the cDNA clones containing a complete ORF with an intact Sox17 HMG box from testis cDNAs, and compared their sequences with those of the clones without the HMG box region. The Sox17 cDNAs showed a 1,257-bp ORF encoding an Sry-related 419-amino acid protein with a predicted molecular mass of 44.6 kD which contained a single intact HMG box near the NH₂ terminus (Figs. 1 and 2 a), and they showed 100% identity at the nucleotide level to those of the unique clones, except for a 391-bp segment containing the initiation methionine and the upper half of Sox17 HMG box. Moreover, comparison of the genomic clones with those of the two cDNAs revealed that the *Sox17* gene contains at least one intron in the center of the HMG box region and two introns in the 5' untranslated region (UTR), and the inserted 391-bp segment of Sox17 cDNAs forms a single exon in the mouse *Sox17* gene (Fig. 2, b and c). By genomic Southern blot analysis using probes A and B (indicated in Fig. 2 a), which can hybridize both intact Sox17 cDNA and the unique cDNA without the HMG box region, a single band was found in all samples digested with various restriction enzymes (Fig. 3). Moreover, both of these probes detected the same 8.6-kb HindIII and 4.6-kb PstI bands in genomic mouse DNA as expected because the Sox17 cDNA sequences between the positions of probes A and B have no restriction sites of these enzymes.

In the DNA samples digested with BamHI or EcoRI, of which restriction site is included in the sequences, probe A or B hybridized the 5-kb BamHI and 9.3-kb EcoRI fragments or the 6-kb BamHI and 1.5-kb EcoRI fragments, respectively. Such fragments coincide exactly with the restriction map in our genomic clones (Fig. 2 b). Therefore, this result proves that the clone without the HMG box region is an isoform of the Sox17 transcript produced at least by splicing out the 391-bp exon containing the upper half of the Sox17 HMG box. Since the clone without the HMG

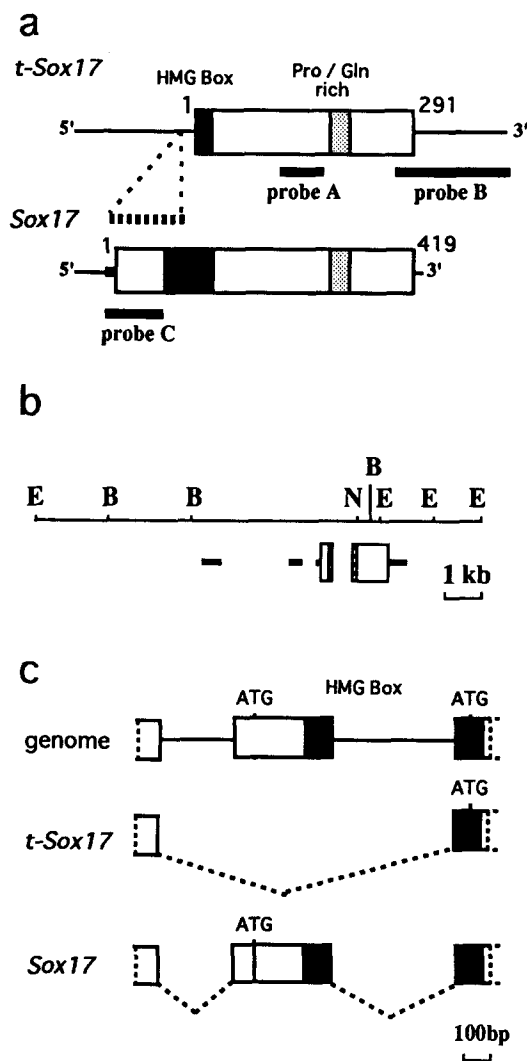


Figure 2. (a) Schematic representation of Sox17 and t-Sox17 cDNAs. The box indicates the ORF (black box, HMG box; dashed box, proline/glutamine-rich region), and the bars indicate the noncoding regions in each isoform. Three bold lines indicate the position of each probe; probes A and B recognize both forms, and probe C is specific for form Sox17. A dashed bold line indicates the 391-bp segment that is deleted in form t-Sox17. (b) Genomic organization of the *Sox17* gene. The box indicates the ORF (hatched box, HMG box), and the solid bars indicate the noncoding regions of Sox17 cDNA. B, BamHI; E, EcoRI; N, NotI. (c) Schematic representation of the genomic structure corresponding to the HMG box region of the *Sox17* gene. The boxes indicate exons, and bars indicate introns. Form t-Sox17 is produced by at least splicing out the exon containing the initiation methionine and the upper half of the HMG box region.

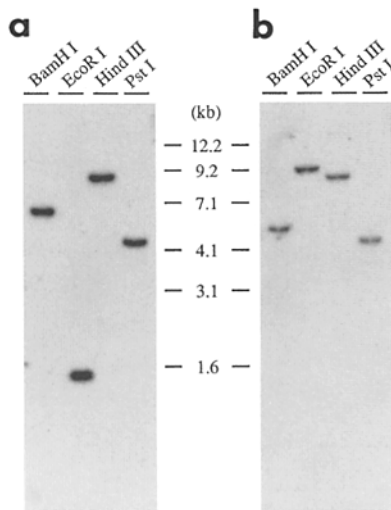


Figure 3. Southern blot analysis of genomic DNA hybridized with probe B (3' UTR probe; *a*) and probe A (coding probe; *b*) of the *Sox17* gene, which can recognize both forms Sox17 and t-Sox17. DNA was digested with (from left to right) BamHI, EcoRI, HindIII, and PstI. Molecular weight markers are shown in the center.

box region is a unique mRNA isoform encoding a truncated Sox17 protein, we termed it as form t-Sox17 (a truncated Sox17). In this paper, we also referred to the Sox17 cDNA with an intact HMG box region as form Sox17. Moreover, the genotype of the back-cross progeny for *Sox17* was determined by Southern analysis using the probe B. BamHI digestion gave RFLP between C3H/HeJ and SEG strains. Linkage analysis with 50 interspecific back-cross progeny mapped the *Sox17* gene to chromosome 1. It is located in the region proximal to D1Mit1 with 2% recombination.

The sequences of the Sox17 HMG box region obtained in this study showed 90.6% and 74.3% similarity with mouse Sox18 HMG box (Dunn et al., 1995) and Sox9 HMG box (Wright et al., 1995) at the amino acid level, respectively. The sequences of the region outside the HMG box showed no significant homology. The COOH-terminal half of the predicted Sox17 protein, however, had a high proportion of proline residues, and a proline- and glutamine-rich region was observed near the COOH terminus, which is commonly found in human *SOX9* (Foster et al., 1994) and mouse *Sox9* (Wright et al., 1995). In addition, a stretch of nine amino acids, SDASSAVYY, at positions 7–15 from the COOH terminus was also found in the same positions of mouse Sox18 cDNA.

Northern Blot and RT-PCR Analyses of the Expression of Sox17 and t-Sox17 Isoforms

Northern blot analysis using probe B (3' UTR probe) revealed that three kinds of transcripts (~1.8, 2.8, and 3.1 kb) were expressed mainly in the lung and testis (Fig. 4 *a*). The 1.8- and 2.8-kb transcripts were observed at low levels in the lung, while the 3.1-kb transcript was expressed abundantly in the testis. Moreover, expression of the 3.1-kb transcript in the testis was regulated during postnatal development. The 3.1-kb transcript was not expressed in the

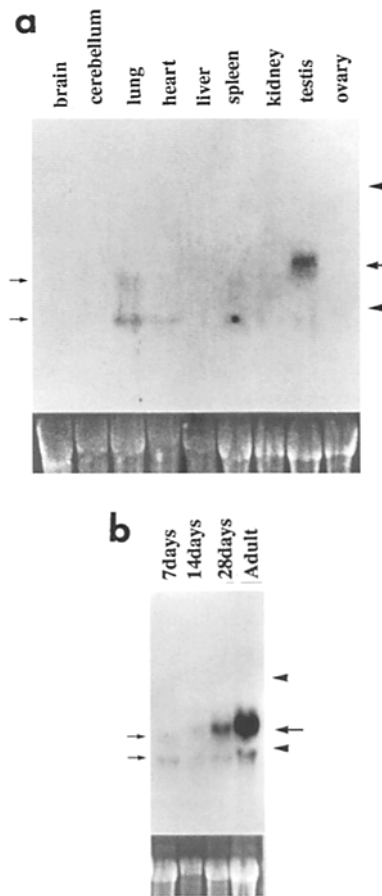


Figure 4. Northern blot analysis showing Sox17 expression in various tissues of adult mice (*a*) and in the testes of prepubertal mice of different ages (*b*). Filters were hybridized with probe B (3' UTR probe), which recognizes both Sox17 and t-Sox17 isoforms. (*a*) Three transcripts (~1.8 kb, 2.8 kb and 3.1 kb) are expressed in the lung and testis. The 1.8- and 2.8-kb transcripts (*small arrows*) are mainly observed at low levels in the lung, while the 3.1-kb transcript (*large arrow*) is abundantly expressed in the testis. (*b*) Expression of the 3.1-kb transcript (*large arrow*) in the testis is regulated during postnatal development. Expression of the 3.1-kb transcript is not detected in the immature testis on day 7 p.p. (7days), but it clearly appears in the testis during days 14 (14days)–28 (28days) p.p. and increases in the mature adult testis (Adult). The positions of rRNAs are marked by arrowheads. The lower panels show the 28S rRNA bands stained with acridine orange.

immature testis on day 7 postpartum (p.p.), but it clearly appeared in the testis during days 14–28 p.p., and increased in the mature adult testis (Fig. 4 *b*). Moreover, to separately detect form Sox17 and form t-Sox17 expressions, we performed RT-PCR analysis. At the adult stage, both Sox17 and t-Sox17 isoforms were expressed in the testis. In the lung, the expression of both forms could be detected, but the expression of t-Sox17 form showed quite a low level (Fig. 5 *a*). As for changes in the expression of each isoform in the testes during postnatal development, form t-Sox17 could not be detected in the immature testis on day 7 p.p. The expression of form t-Sox17 was detected in the testis on day 14 p.p., and increased clearly during days 14–28 p.p.; this coincides with the expression pattern

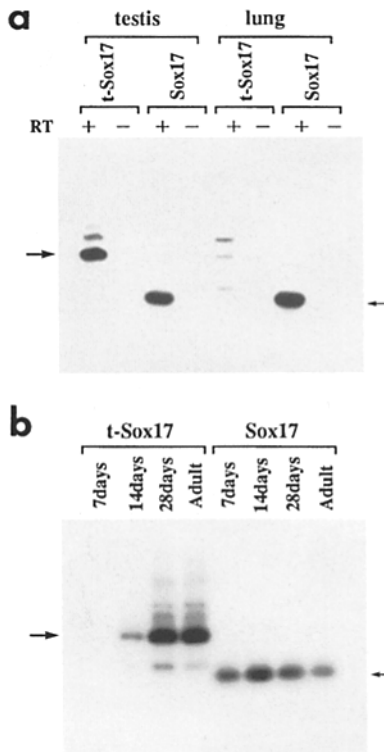


Figure 5. RT-PCR analysis showing the expression of Sox17 and t-Sox17 isoforms in the testis and lung of the adult mice (*a*) and in the testis of prepubertal mice of different ages (*b*). The PCR products were hybridized with the probe that recognizes both forms, as described in Materials and Methods. (*a*) At the adult stage, both Sox17 (small arrow) and t-Sox17 (large arrow) isoforms are detected in the testis. In the lung, the expression of both forms can be detected, but the t-Sox17 form is expressed at a low level. (*b*) In prepubertal mice of different ages, the t-Sox17 form (large arrow) cannot be detected in the immature testis on day 7 p.p. (7days). The expression of form t-Sox17 is detected in the testis on day 14 p.p. (14days), and it clearly increased in the testes on day 28 p.p. (28days) and at the adult stage (Adult). In contrast, the Sox17 form (small arrow) is expressed throughout postnatal development. RT + or - in each lane of *a* indicates that RT was added or omitted, respectively, to the PCR reaction sample.

of the 3.1-kb transcript during postnatal development (Fig. 4 *b*). In contrast, the Sox17 form was expressed throughout postnatal development (Fig. 5 *b*). These findings therefore indicate that the Sox17 form is expressed at a low level in the testis throughout postnatal development, while the t-Sox17 form is expressed abundantly in the testis, predominantly in postmeiotic germ cells.

The expression patterns in the testes during the postnatal development clearly indicate that the 3.1-kb transcript or the 1.8- and 2.8-kb transcripts may be identical to form t-Sox17 or Sox17, respectively. The t-Sox17 transcript is longer than the Sox17 transcript probably because of the difference in the length of their noncoding sequences. This leads us to the suggestion that form t-Sox17 may be produced, not only by splicing out of the 391-bp exon corresponding to the NH₂-terminal coding region, but also by the use of the different noncoding regions(s), e.g., the use of the alternative transcriptional start or polyadenylation site and/or the alternative splicing of the other exon(s).

In addition, all RT-PCR products of forms Sox17 and t-Sox17 actually represent true mRNA products because the genomic DNA regions corresponding to the amplified Sox17 or t-Sox17 cDNAs contained one or more intron(s), and no amplification product was detected in samples where reverse transcriptase had been omitted from the reactions (Fig. 5 *a*, RT “-”). Moreover, each PCR product was confirmed by sequence analysis as the Sox17 or t-Sox17 cDNAs, and the t-Sox17 PCR products corresponding to the two different minor bands (one small band and another larger band that were weakly detected in the t-Sox17 sample [Fig. 5, *a* and *b*]) were produced by deletion of the segment (+474 to +565) and an insertion at the +474 position. These minor bands most likely resulted from the splice variants in the 5' UTR of the t-Sox17 form.

In Situ Hybridization Analysis

As described above, the testicular expression of each isoform was regulated differently during postnatal development. Moreover, we examined the localization of each isoform in the adult testis by in situ hybridization. When using probe B (the 3' UTR probe, which recognizes both forms Sox17 and t-Sox17), positive signals were mainly observed in the inner cell layer of seminiferous tubules of adult testis (Fig. 6, *a* and *c*). Late pachytene spermatocytes and the round, elongated spermatids were positive for hybridization with this probe, and more intense positive signals were notably restricted to early round spermatids (stages I–VII) (Fig. 6 *d*). On the other hand, with probe C (the probe specific for the Sox17 form), the distribution of the signals showed a different pattern from probe B. Signals were detected mainly in the basal layer of seminiferous tubules (Fig. 6, *e* and *g*), showing that the Sox17 form was expressed in spermatogonia, and such expression was clearly reduced from the early pachytene spermatocyte stage and onward (Fig. 6 *h*). These results suggest that positive signals of probe B in meiotic and postmeiotic germ cells reflect localization of form t-Sox17, and that t-Sox17 form is expressed in spermatogenic cells, from the late pachytene spermatocyte to the elongated spermatid stages, especially in round spermatids at a high level. Expression patterns of forms Sox17 and t-Sox17 by in situ hybridization were clearly consistent with the results by Northern hybridization and RT-PCR analyses using testis RNA from prepubertal mice of different ages (Figs. 4 *b* and 5 *b*). In addition, no appreciable positive signals were detected in Sertoli or Leydig cells in the cases using these antisense probes, although some positive somatic cells are located in the testicular interstitium (Fig. 6 *h*). Moreover, sense probes showed no positive signals throughout any sections of the testis (Fig. 6, *b* and *f*).

Analyses of Protein Products of Sox17 and t-Sox17 Isoforms

To test whether each RNA isoform is translated into the protein product, we first carried out in vitro transcription and translation analysis using Sox17 (+564 to +1940) and t-Sox17 (+1 to +2375 of t-Sox17) cDNAs. By SDS-PAGE analysis, a doublet of the translated proteins was detected in both samples from Sox17 and t-Sox17 cDNAs. One band was ~0.4 kD smaller than another protein (Fig. 7 *a*),

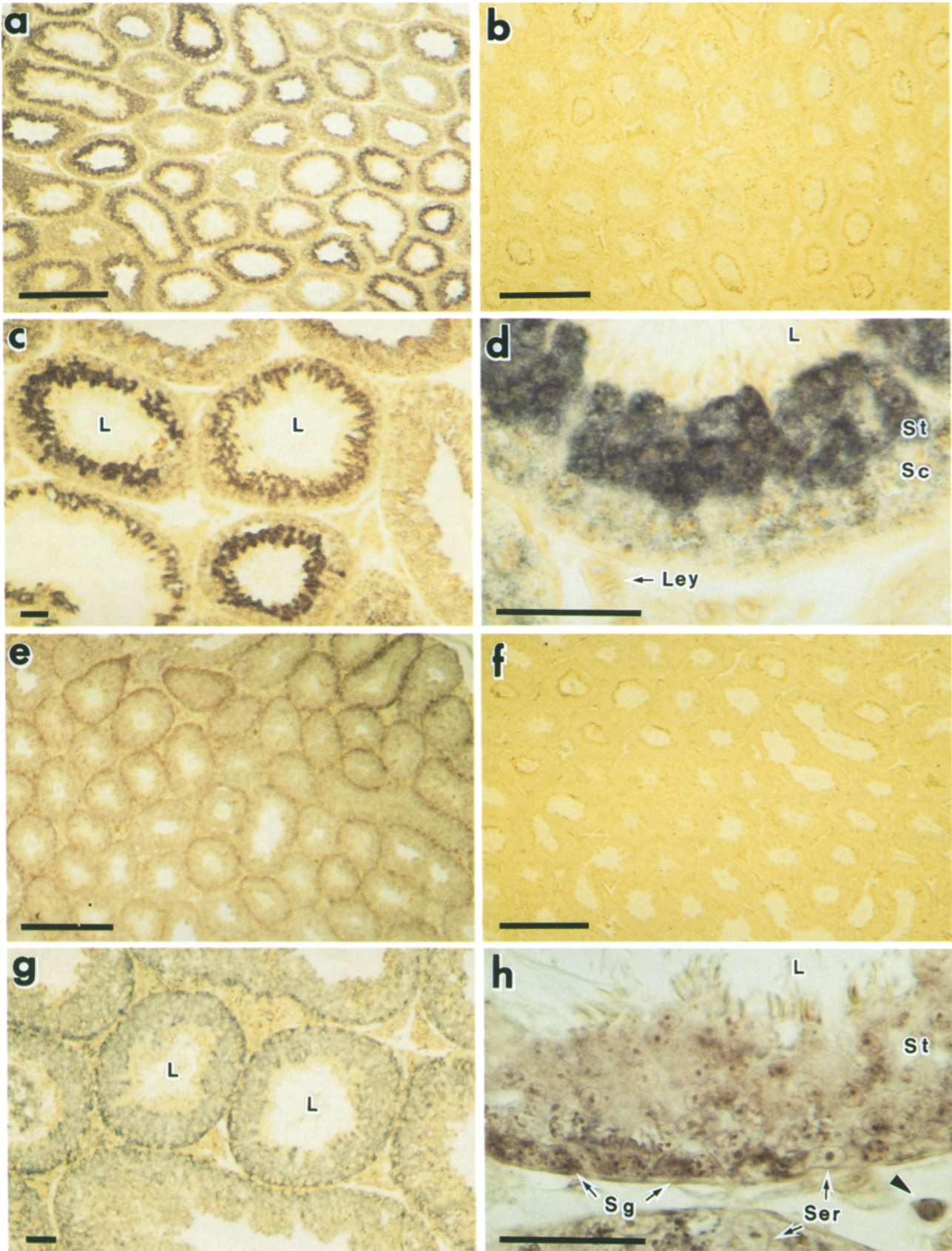


Figure 6. In situ hybridization analysis showing the localization of the Sox17 and t-Sox17 isoforms in the adult testis by using probe B (3' UTR probe, *a-d*) and probe C (the probe specific for form Sox17, *e-h*). In the case of using probe B, which recognizes both forms Sox17 and t-Sox17, positive signals are observed mainly in the inner cell layer of seminiferous tubules of the adult testis (*a* and *c*). Late pachytene spermatocytes, and round, elongated spermatids are positive for hybridization with this probe, and more intense positive sig-

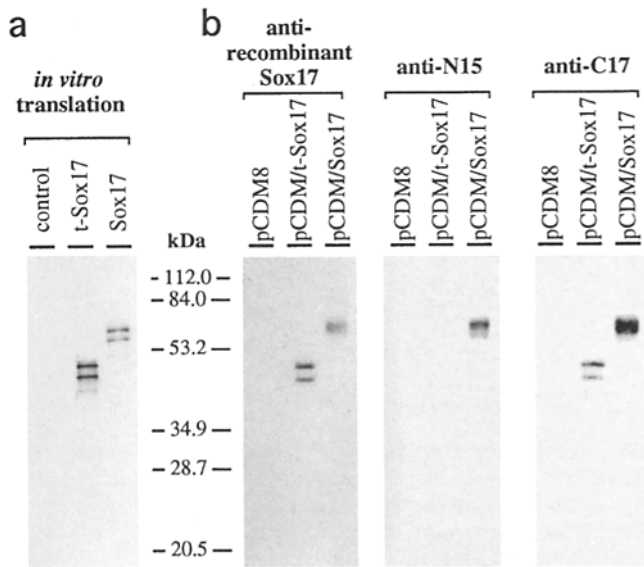


Figure 7. In vitro transcription/translation (*a*) and immunoblot (*b*) analyses of protein products of Sox17 and t-Sox17 mRNA isoforms. (*a*) Sox17 or t-Sox17 cDNA was transcribed and translated in vitro in the presence of [³⁵S]methionine. Samples were analyzed by 12.5% SDS-PAGE and autoradiographed. (*b*) The protein sample of the COS cells transfected with pCDM8 alone, pCDM/t-Sox17, or pCDM/Sox17 was electrophoresed on a 12.5% SDS-polyacrylamide gel, and was immunoblotted by using anti-N15 (15-residue peptide [amino acid residues 25–39] of the Sox17 NH₂-terminal region) or anti-C17 (17-residue peptide [amino acid residues 403–419] of the Sox17 COOH-terminal region) antibody or anti-recombinant Sox17 antiserum. Anti-N15 antibody detects only the Sox17 protein product, while anti-recombinant Sox17 antiserum or anti-C16 antibody recognized both Sox17 and t-Sox17 proteins. Both Sox17 and t-Sox17 proteins migrate slower than expected, possibly because of their high proline content. Molecular weight size standards are shown in the center.

which may be caused by the translation from the second methionine in each ORF (position 24 in Sox17 ORF or position 20 in t-Sox17 ORF). Moreover, the larger t-Sox17 protein was ~14 kD smaller than the larger Sox17 protein, as predicted by their respective ORFs. To confirm the data using in vitro translation products, we raised three kinds of antisera (anti-N15 [15-residue peptide of amino acid residue 25–39 of the Sox17 NH₂-terminal region], anti-C17 [17-residue peptide of amino acid residue 403–419 of the Sox17 COOH-terminal region], and anti-recombinant GST-Sox17 antiserum) in immunized rabbits, and we examined the expression of each protein product in the COS cells transfected with pCDM/t-Sox17 or pCDM/Sox17. On immunoblots of the protein extracts of the COS cells (Fig. 7 *b*), both anti-recombinant Sox17 antiserum and anti-C17 antibody detected a doublet of bands in the cells trans-

ected with pCDM/t-Sox17, which resembles a doublet of in vitro translation products in size (Fig. 7*a*). In the extract of cells transfected with pCDM/Sox17, these antisera also detected a specific band with a similar size to the larger product of in vitro-translated Sox17 proteins. Moreover, immunoreactivity for anti-N15 antibody was present only in the protein extract of the cells transfected with pCDM/Sox17, and was not detected in the extracts with pCDM/t-Sox17. The detected band showed the same size as the one in the immunoblot using anti-recombinant Sox17 antiserum or anti-C17 antibody.

On the immunoblots of the testicular crude extracts, we could not detect a specific band of both Sox17 and t-Sox17 products. To confirm the existence of both isoform products in the testis, the testicular extracts were analyzed by ion exchange HPLC. They were applied to a DEAE exchange column, and each eluted fraction was subjected to SDS-PAGE (12.5%) and immunoblotting (Fig. 8, *a* and *b*). As a result, the Sox17- and t-Sox17-specific bands were detected in the fractions eluted from 250 to 400 mM NaCl (Fig. 8 *b*). Sox17 products were eluted at the lower salt concentration (270–315 mM) in comparison with t-Sox17 products, which may reflect a deficiency of the HMG box region enriched with basic amino acids in the t-Sox17 isoform. These findings suggest that both Sox17 and t-Sox17 mRNAs isoforms are translated into the protein products, although the amount of t-Sox17 products in the testis appeared lower in comparison to the high accumulation of t-Sox17 mRNAs.

By immunofluorescence staining using anti-recombinant Sox17 antiserum, positive reactions are restricted to the nucleus of L929 cells transfected with pCDM/Sox17, suggesting that Sox17 is a nuclear protein like Sry and other Sox proteins (Fig. 9 *a*). In the L929 cells transfected with pCDM/t-Sox17, positive reactions were observed both in the cytoplasm and nucleus (Fig. 9 *b*). Similar patterns were also obtained in the COS cells transfected with pCDM/Sox17 or pCDM/t-Sox17 (Fig. 9, *c* and *d*). Moreover, all of the L929 cells transfected with the COOH-terminal deletion constructs of pCDM/Sox17 (see Fig. 12 *a*) exhibited a nuclear localization of each product (figure not shown). Such intracellular localization patterns in the transfection experiment clearly agree with a previous report showing that the NH₂-terminal part of the HMG box domain functions as a nuclear localization signal in diverse HMG box proteins (Poulat et al., 1995). Thus, t-Sox17 is a truncated Sox17 protein lacking both the DNA-binding domain and nuclear localization signal, although a part of t-Sox17 protein can be transported into the nucleus. In the testicular sections stained with anti-recombinant Sox17 antiserum, positive reactions were detected in the nucleus of the spermatogonia and the round spermatids (Fig. 9 *e*). Moreover, some positive somatic cells are located in the testicular interstitium. The spermatocytes, however, showed

nals are especially restricted to early round spermatids (*d*). On the other hand, with probe C, the distribution of the signals shows a different pattern from probe B, and signals are mainly detected in the basal layer of the seminiferous tubules (*e* and *g*). The Sox17 form is expressed in spermatogonia, and this expression is clearly reduced from the early pachytene spermatocyte stage and onward (*h*). No appreciable positive signals were detected in Sertoli or Leydig cells with these antisense probes, although some positive somatic cells were located in the testicular interstitium (*h*, arrowhead). Moreover, sense probes show no appreciable positive signals in any sections of the testis (*b* and *f*). *L*, lumen of seminiferous tubules; *Sg*, spermatogonium; *Sc*, spermatocyte; *St*, spermatid; *Ser*, Sertoli cell; *Ley*, Leydig cell. *a*, *b*, *e*, and *f*, bar = 1 mm; *c*, *d*, *g*, and *h*, bar = 100 μm.

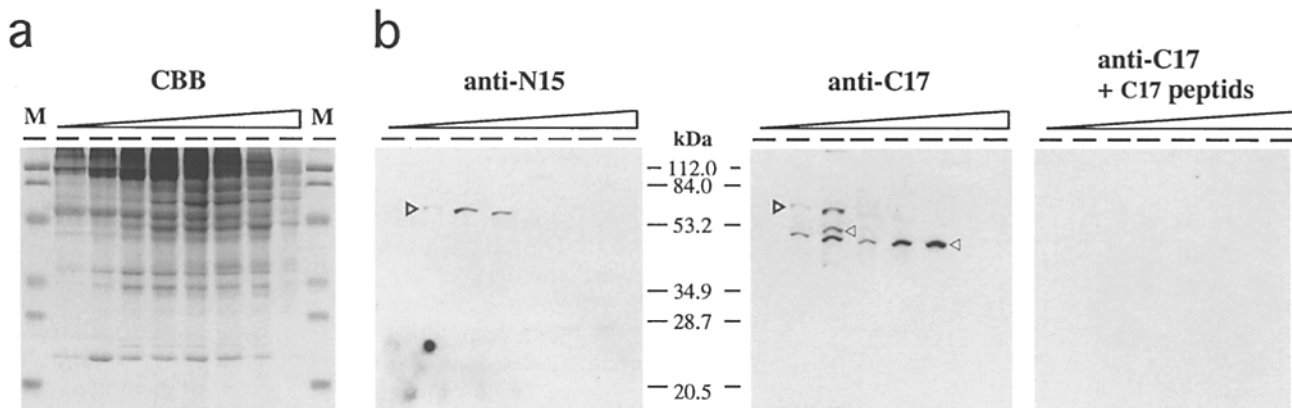


Figure 8. Immunoblot analyses of Sox17 and t-Sox17 products in the testicular extracts fractionated by DEAE chromatography. The testicular extracts were applied to a DEAE ion exchange column and eluted with a linear salt gradient from 0 to 400 mM NaCl. Individual fractions eluted from 250 to 400 mM NaCl (fractions 12–19 from left to right) were electrophoresed on a 12.5% SDS-polyacrylamide gel, and were immunoblotted by using anti-N15 or anti-C17 antibody. (a) CBB staining pattern. *M*, molecular mass markers (from top to bottom, 112.0, 84.0, 53.2, 34.9, 28.7, and 20.5 kD). (b) Immunostaining patterns with anti-N15 or anti-C17 antibody. The boldfaced triangles indicate the Sox17 product, and the regular triangles indicate the t-Sox17 product. A control blot was stained by anti-C17 antibody preincubated with the antigenic hapten peptide (C17; 100 μ M). Molecular mass size standards are shown on the center.

a weak or no apparent positive for anti-Sox17 staining. We could not distinguish Sox17 from t-Sox17 proteins on the sections because the anti-N15 antibody was disqualified due to histochemistry. Such staining patterns in the testis by anti-recombinant Sox17 antiserum, however, are in agreement with the mRNA expression patterns by *in situ* hybridization analysis. Thus, it suggests that the immunohistochemical reactions in the spermatogonia or those in the spermatids may reflect the localization of Sox17 or t-Sox17 proteins, respectively. It further suggests that a switch from Sox17 into t-Sox17 may occur both at protein and RNA levels.

In addition, no positive reaction of anti-recombinant Sox17 was detected in L929 and COS cells transfected with pCDM8. The cytochemical control using the primary antiserum preincubated with GST-Sox17 beads showed no specific positive reaction in the testis (Fig. 9 *d*) or in the L929 and COS cells transfected with pCDM/Sox17 or pCDM/t-Sox17 (not shown).

DNA-binding Activity of Sox17 or t-Sox17

It was shown previously that Sry and several Sox proteins have a sequence-specific DNA-binding activity to the AACAAAT or AACAAAG motif (Denny et al., 1992*b*; van de Wetering et al., 1993; Harley et al., 1994). To test functional differences between Sox17 and t-Sox17 protein products, we first examined the DNA-binding activity of recombinant GST-Sox17 or t-Sox17 protein by EMSA. The probes used were 32 P-labeled oligonucleotides with partially self-complementary sequences that form a stem-loop structure and provide a 13-bp double-stranded region containing each motif (the oligonucleotides containing the motifs AACAAAT and AACAAAG are named SCS6 and SCS4, respectively). Both GST alone and GST-t-Sox17 protein showed no detectable DNA binding to these oligonucleotides (Fig. 10 *a*). In contrast, GST-Sox17 protein clearly showed DNA-binding activity to both SCS6 and SCS4, and its binding activity to SCS6 seemed to show a

higher level compared with that of SCS4 (Fig. 10 *a*). The binding of Sox17 to SCS6 was inhibited by the addition of unlabeled SCS6 oligonucleotides, while a set of different target sequences of other known transcription factors were unable to successfully compete this binding (Fig. 10 *b*). No appreciable competition with these nonspecific target oligonucleotides was observed in the case of Sox17 protein and SCS4 binding (figure not shown). Thus, these data indicate that the Sox17 protein is a DNA-binding protein with a sequence specificity similar to other members of the Sox family proteins.

Moreover, we selected the sequences binding to GST-Sox17 or GST-t-Sox17 from a pool of random DNA sequences using each recombinant protein. The 32 P-labeled PCR products from alternate cycles of selection were analyzed by EMSA (Fig. 10 *a*). In the selection using GST-Sox17, an appreciable enrichment for Sox17-binding oligonucleotides was found by the third cycle of selection, and a large part of the PCR products from the fifth cycle of selection formed protein-DNA complexes. On the other hand, no DNA-protein complex was observed when using t-Sox17 protein, even in the fifth cycle of selection, suggesting that t-Sox17 protein shows no sequence-specific DNA-binding activity. The PCR products binding to GST-Sox17 were sequenced after seven cycles of selection. As a result, all individual clones recovered were AT rich, and AACAAAT occurred in 19 of 31 clones (Fig. 11 *b*), which confirms the results of the DNA binding of Sox17 protein to SCS6 oligonucleotides (Fig. 10 *a*). Interestingly, in 29 of 31 clones, two copies of AACAAAT or its minor variants were found within the random 18-bp region of the original random oligonucleotide, which forms palindrome-like sequences composed of two inverted AACAAAT or its variant sites separated by three to five nucleotides. This result leads us to the possibility that the Sox17 protein binds DNA cooperatively or as a dimer. However, the EMSA and immunoprecipitation analysis using *in vitro*-translated proteins could demonstrate neither a cooperative binding to DNA nor a direct interaction at present (data not shown).

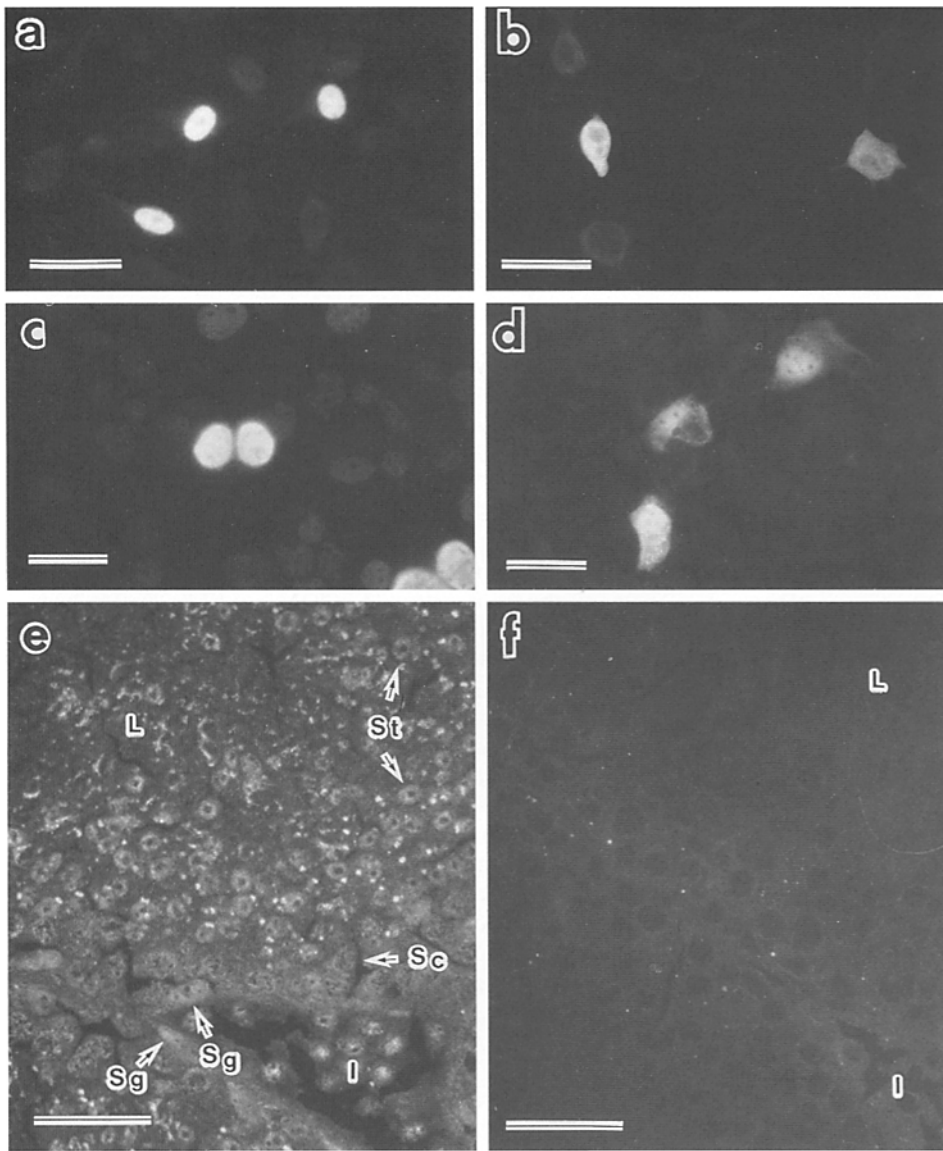


Figure 9. Immunofluorescence analysis of Sox17 and t-Sox17 isoform protein products using anti-recombinant Sox17 antiserum, showing the intracellular localization in the transfected mammalian cells and distribution in the adult testis. *a-d* show the intracellular localization of Sox17 or t-Sox17 in L929 cells (*a* and *b*) and COS cells (*c* and *d*). In the cells transfected with pCDM/Sox17, positive reactions are restricted to their nucleus (*a* and *c*), while reactions are found both in the cytoplasm and nucleus in those with pCDM/t-Sox17 (*b* and *d*). *e* shows the distribution of positive reactions for anti-recombinant Sox17 antiserum in the adult testis, while *f* exhibits a cytochemical control in the testicular section. In the testis, positive reactions are detected in the nucleus of the spermatogonia and the round spermatids (*e*). Moreover, some positive somatic cells are located in the testicular interstitium. Such positive reactions have disappeared by the use of the primary antiserum preincubated with GST-Sox17 (*f*). *L*, lumen of seminiferous tubules; *I*, interstitium; *Sg*, spermatogonium; *Sc*, spermatocyte; *St*, spermatid. Bar, 100 μ m.

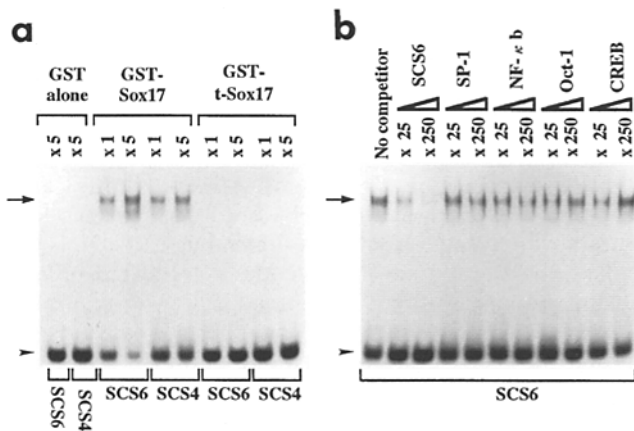


Figure 10. EMSA on the DNA-binding activity of the GST-Sox17 or GST-t-Sox17 proteins to SCS6 (oligonucleotides containing the sequences AACAAT) or SCS4 (oligonucleotides containing the sequences AACAAAG). (*a*) Each purified GST fusion protein ($\times 1$, 50 ng; $\times 5$: 250 ng/10 μ l of total reaction buffer) was incubated with 32 P-labeled SCS6 or SCS4. Neither

Transactivation Activities of Sox17 or Derivatives through Its Binding Motif

To test whether Sox17 could regulate transcription through the Sox 17-binding site, the reporter plasmid containing four copies of the two inverted AACAAT motifs

GST nor GST-t-Sox17 protein showed detectable DNA-binding activity to these oligonucleotides, while GST-Sox17 protein clearly bound to both SCS6 and SCS4. Its binding activity to SCS6 seems to show a high level, compared with SCS4. (*b*) Competition analysis of the binding of GST-Sox17 (50 ng/10 μ l of total reaction buffer) to 32 P-labeled SCS6. Incubation was performed in the absence (no competitor) or presence of unlabeled SCS6 oligonucleotides or a set of oligonucleotides harboring consensus-binding sequences for other transcriptional factors at both $\times 25$ and $\times 250$ molar excesses. The binding of Sox17 to SCS6 is inhibited by the addition of unlabeled SCS6 oligonucleotides, while the nonspecific target oligonucleotides are unable to compete successfully this binding. Arrows indicate DNA-protein complexes, and arrowheads indicate a free probe.

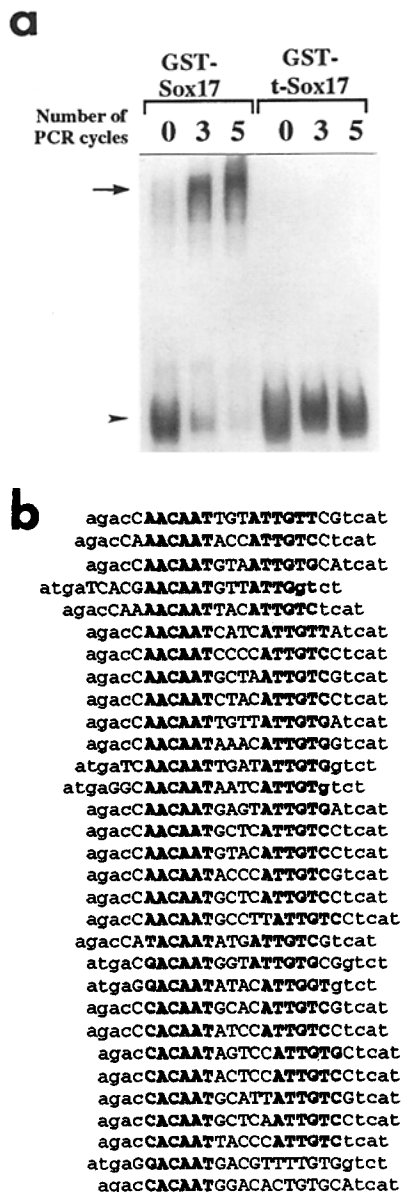


Figure 11. Random selection analysis from a pool of random DNA sequences using each recombinant GST-Sox17 or GST-t-Sox17 protein. (a) EMSA of ^{32}P -labeled random oligonucleotides from alternate cycles of selection by GST-Sox17 or GST-t-Sox17. The number of cycles is shown in each lane (0 indicates the random oligonucleotides). PCR products at each cycle were labeled by T4 polynucleotide kinase with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$, and were incubated with each recombinant protein (250 ng/10 μl of total reaction buffer). In a selection using GST-Sox17, appreciable enrichment for Sox17-binding oligonucleotides was found by the third cycle of selection, while no DNA-protein complex was observed in the case of using t-Sox17 protein, even in the fifth cycle of selection. Arrows indicate DNA-protein complexes, and arrowheads indicate a free probe. (b) DNA sequences of oligonucleotides isolated after seven rounds of selection for GST-Sox17 protein binding sites. Uppercase letters represent the random 18-bp region of the original random oligonucleotide. Lowercase letters denote the common sequences flanking the random core. Sequences were aligned about their core-binding sequence, AACAAT.

(pH4 \times 4-Luci) was constructed and cotransfected with each effector plasmid (pCDM, pCDM/t-Sox17, or pCDM/Sox17) into L929 cells (Fig. 12 a). Transfection of pCDM/t-Sox17 into L929 cells gave no detectable stimulation of luciferase, while the expression of Sox17 was able to activate transcription of the reporter gene about six- to sevenfold (Fig. 12 b). Moreover, such transactivation by Sox17 was dependent on the presence of the AACAAT sites, since Sox17 failed to promote luciferase expression from a reporter gene lacking this sequence, pGL2-Luci (Fig. 12 c). The similar results were also obtained in the transfection experiments using HeLa cells (not shown). To define the domain involved in transactivation, the deletion mutants of Sox17 were constructed (Fig. 12 a) and cotransfected with pH4 \times 4-Luci into L929 cells. As a result, transfection of pCDM/Sox17(d343-419), which lacks the sequence corresponding to amino acid residues 343-419, resulted in a significantly low level of luciferase activity, which shows a similar level to that of pCDM8 (Fig. 12 d). pCDM/Sox17 (d295-419) also exhibited the lack of the activation activity, but the transfection of pCDM/Sox17 (d286-346) significantly induced reporter gene expression by $\sim 75\%$ of that of pCDM/Sox17, suggesting that the COOH-terminal region of amino acid residues 347-419 may be involved in the transactivation. In addition, the transfection of pCDM/Sox17 (d174-346) failed to promote luciferase expression. Such a result may be possibly caused by the structural failure of this deletion mutant protein.

In conclusion, Sox17 protein is a transcriptional activator through the AACAAT motif, while t-Sox17 is a truncated protein that contains the activation domain but lacks both the DNA-binding domain and the nuclear localization signal. Thus, Sox17 may function as a transcriptional activator during the early stage of the testicular germ cells, and thereafter, Sox17 may be functionally inactivated by the splicing switch into t-Sox17 during the peri- and post-meiotic stages.

Discussion

In the present study, two types of Sox17 cDNA clones were isolated from an adult mouse testis cDNA library. Genomic structural analysis revealed that the mouse *Sox17* gene contains two introns at the positions of the 5' UTR and the HMG box region that forms a 391-bp exon containing the NH₂-terminal 102 amino acids (from the initiation methionine to the upper half of HMG box). By alternative splicing of this exon, we showed that the *Sox17* gene produces at least two mRNA isoforms: one encodes a protein with a single HMG box domain near the NH₂ terminus (Sox17), and another encodes a truncated protein lacking most parts of the HMG box domain (t-Sox17). The *Sox17* gene is therefore the first member of Sox gene family that produces multiisoforms, one of which encodes a truncated protein that lacks most of the DNA-binding domain.

Sox family genes are divided into two groups, intronless and multiexon genes. It is known that *Sry*, *Sox3*, and *Sox4* are all intronless genes (Clepét et al., 1993; Stevanović et al., 1993; Schilham et al., 1993), while the *Sox9* gene contains at least two introns, one of which is located in the HMG box region (Foster et al., 1994; Wagner et al., 1994;

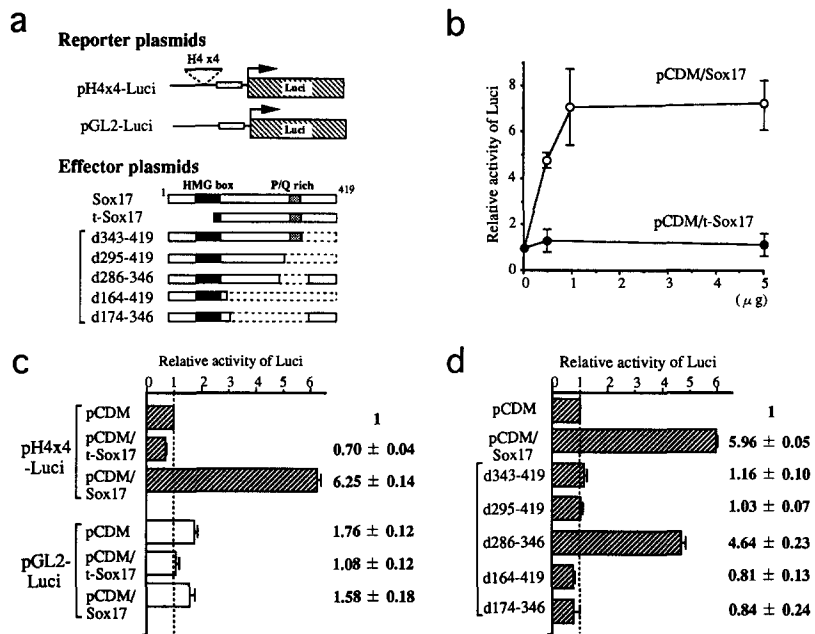


Figure 12. Transactivation activities of Sox17 and its derivatives through its binding motif in a transient luciferase assay. (a) Schematic representation of reporter and effector plasmids. pH4x4-Luci were constructed by insertion of four copies of the H4 site (two inverted AA-CAAT motifs separated by 4 bp) into pGL2-Luci immediately upstream of the SV40 promoter. Effector plasmids were constructed by cloning the Sox17 or t-Sox17 cDNA or its derivatives into pCDM8 downstream of the cytomegalovirus promoter. (b) pH4x4-Luci (1 μg) was transfected into L929 cells together with the indicated amount of pCDM/Sox17 (○) or pCDM/t-Sox17 (●). Total amounts of plasmid were kept constant by adding the vector without its insert. (c) L929 cells were transfected with a reporter plasmid (pH4x4-Luci or pGL2-Luci, each 1 μg) and effector plasmids (pCDM8, pCDM/Sox17, or pCDM/t-Sox17, 5 μg each). (d) pH4x4-Luci (1 μg) was transfected into L929 cells together with pCDM/Sox17 or its deletion derivatives, pCDM/Sox17[d343-419], pCDM/Sox17[d295-419], pCDM/Sox17[d286-346], pCDM/Sox17[d164-419], or pCDM/Sox17[d174-346] (5 μg each). Transfection efficiency was normalized to β-galactosidase activity. Error bars represent SE from six (a) or four (c and d) independent transfection trials.

Wright et al., 1995). Although *Sox9* and *Sox17* show no significant homology in the sequences of the region outside the HMG box, the COOH terminal halves of both predicted proteins show a high proportion of proline residues, and they contain proline- and glutamine-rich regions near the COOH terminus (Foster et al., 1994; Wright et al., 1995). The similarity between these two genes suggests that the *Sox17* gene is more closely related to the *Sox9* gene than the intronless Sox family genes that contain *Sry*. The Sox17 HMG box domain also shows the most similarity to Sox18 cDNA (90.6%), and a stretch of nine amino acids, SDASSAVYY, at positions 7–15 from the COOH terminus, is also conserved in these two genes (Dunn et al., 1995). Although we have no information about whether or not the *Sox18* gene is a multiexon gene that contains an intron positioned at the HMG box region, the *Sox17* gene seems likely to be most closely associated with the *Sox18* genes than other Sox family genes. Further information about the *Sox18* gene will shed light on the close relationship among the *Sox9*, *Sox17*, and *Sox18* genes.

The most interesting feature of the *Sox17* gene in this study is the expression patterns of each isoform. Our RNA analyses revealed that both Sox17 and t-Sox17 isoforms are expressed in the spermatogenic cell lineage, and the expression of each isoform is differently and developmentally regulated in spermatogenesis. In mammalian spermatogenesis, male germ cells undergo three developmental events: the mitotic proliferation of spermatogonia (mitotic phase), the meiotic division of spermatocytes (meiotic phase), and the subsequent formation of haploid spermatids and their morphogenic changes into sperm (postmeiotic phase). According to the present analysis, the Sox17 form is expressed in spermatogonia, and such expression is clearly reduced at the early pachytene sperma-

toocyte stage and onward; therefore, this leads us to the postulation that the Sox17 form may be associated with the mitotic phase of spermatogenesis. The recombinant Sox17 protein synthesized in vitro exhibits DNA-binding activity with a similar sequence specificity to other members of the Sox gene family. Moreover, by cotransfection experiments using a luciferase reporter gene, Sox17 could stimulate reporter gene expression through its binding site. These results therefore suggest that Sox17 may function as a transcriptional factor, activating (or repressing) the expression of other genes involved in the premeiotic phase of spermatogenesis, such as the spermatogonial proliferation and induction of meiosis.

On the other hand, form t-Sox17 shows a deletion of the region containing the upper half of HMG box region, and its product shows a truncated protein lacking most parts of the HMG box domain and the nuclear localization signal. Since the *Sox17* gene contains a single HMG box, as do *Sry* and the other Sox family genes, the t-Sox17 isoform missing the DNA-binding domain may have no sequence-specific DNA-binding activity. This fact is also supported by no appreciable enrichment of t-Sox17-binding sequences in the random selection assay (Fig. 11 a). The functional significance of t-Sox17 isoform products is elusive at the present time. However, the expression of form t-Sox17 clearly shows a reciprocal proportion to that of form Sox17 in spermatogenesis. The t-Sox17 form begins expression in the pachytene spermatocyte, and its transcripts are highly accumulated in the round spermatids, in contrast with the appreciable reduction in Sox17 isoform expression at the meiotic phase. These findings suggest that Sox17 pre-mRNA is likely to be expressed in the spermatogenic cells from the mitotic through postmeiotic phases, and they further suggest that some changes in the

pre-mRNA-processing machinery at the meiotic phase may induce the switch from the Sox17 isoform into t-Sox17 isoform.

In fact, the meiotic and postmeiotic germ cells are known to express aberrant forms of various substances such as Oct-2, CREB (cAMP-responsive element-binding factor), CREM (cAMP-responsive element modulator), and fer-T by alternative splicing (see review of Erickson, 1990; Eddy et al., 1993). This pattern suggests that mechanisms specific to meiotic and postmeiotic germ cells may alter pre-mRNA processing, which results in the production of a functionally different substance during these phases. For example, it is known that during the meiotic phase, a significant switch in the splicing of CREB transcripts leads to higher levels of the two isoforms that lack the leucine zipper domain and the nuclear translocation signals (Waeber et al., 1991; Ruppert et al., 1992). Moreover, CREM is expressed in the antagonist form in premeiotic germ cells, but an agonist isoform that acquires two glutamine-rich transcriptional activation domains by alternative splicing is produced exclusively at the pachytene spermatocyte stage (Foulkes et al., 1992). It is therefore conceivable that similar mechanisms may lead to a significant switch from Sox17 to the t-Sox17 isoform during the meiotic phase, and this developmental switch may alter the function of Sox17 at the meiotic and postmeiotic phases.

It is most likely that a splicing switch at the meiotic phase leads to the loss of function of Sox17. At the present time, however, we can not rule out a possible role of truncated Sox17 protein, which lacks a DNA-binding domain, in postmeiotic germ cells. Among the HMG box family proteins, there are several reports showing that this family has important regulatory interactions with the other transcriptional factor; for example, the upstream binding factor (UBF) is reported to interact with SL-1 in a way that DNA binding by these factors is cooperative (Jantzen et al., 1992). Moreover, an HMG box protein, SSRP1, interacts specifically with the c-Myc oncoprotein, and SSRP1 affects c-Myc function in transfection experiments (Bunker and Kingston, 1995). The present immunofluorescence analysis, in combination with the transfection experiment, showed that a part of t-Sox17 protein was transported into the nucleus of L929 and COS cells in spite of t-Sox17 cDNA missing the region that corresponds to the nuclear localization signal. Moreover, in the testicular sections positive reactions were detected in the nucleus of the round spermatids by anti-Sox17 staining. These results suggest that t-Sox17 protein may interact with some nuclear proteins, even in the COS and L929 cells, and that t-Sox17 may be transported into the nucleus together with these proteins. Testing this hypothesis would require identifying the nuclear protein(s) that can interact specifically with t-Sox17 by the yeast two-hybrid system and etc.

The Sox17 form is not only expressed in the spermatogenic cell lineage, but in some other somatic cell lines as well. The present Northern and in situ hybridization analyses showed that form Sox17 is expressed in the lung and in some somatic cells located in the testicular interstitium. Moreover, RT-PCR analysis revealed that the t-Sox17 isoform is also expressed in the lung, although its expression level is quite low compared with that of form Sox17. The

expression of both isoforms in somatic cell lines also stresses the biological importance of t-Sox17 isoform.

In conclusion, we found that a Sox17 gene is the first member of Sox gene family to encode functionally different mRNA isoforms by alternative splicing, and the expression of each isoform is regulated differently in spermatogenesis. We believe that identification of the Sox17 gene not only provides information about the transcriptional networks in spermatogenesis, but also a novel insight to the functional regulation of Sox family genes at the mRNA-processing level.

The authors wish to thank Dr. Hirokazu Fujimoto (Department of Molecular Reproduction and Development, Mitsubishi Kagaku Institute of Life Sciences) and Associate Professor Dr. Yoshimitsu Kanai (Department of Anatomy and Cell Biology, Faculty of Medicine, University of Tokyo) for their kind and helpful advice and discussions regarding this work; Dr. Peter Koopman for his helpful discussions and suggestions; and Dr. George Muscat for providing information about the Sox18 gene before publication; department director Dr. Hiroshi Hamada, Dr. Yukio Saijoh (Department of Developmental Biology and Cancer Prevention, The Tokyo Metropolitan Institute of Medical Science), Associate Professor Dr. Masamichi Kurohmaru, and Dr. Kentaro Tanemura (Department of Veterinary Anatomy, Faculty of Agriculture, University of Tokyo) for their helpful support; Mr. Iwao Tsugiyama (Department of Veterinary Anatomy, University of Tokyo) for his expert care of laboratory mice; and Mrs. Hiroe Hayashi (Department of Ultrastructural Research, The Tokyo Metropolitan Institute of Medical Science) for her secretarial assistance.

This work was supported by financial grants from the Ministry of Education, Science, and Culture of Japan to Y. Kanai (Nos. 06660394 and 07660412), and the Japan Society for the Promotion of Science to M. Kanai-Azuma (No. 2786).

Received for publication 10 March 1995 and in revised form 23 January 1996.

References

- Bunker, C.A., and R.E. Kingston. 1995. Identification of a cDNA for SSRP1, an HMG-box protein, by interaction with the c-Myc oncoprotein in a novel bacterial expression screen. *Nucleic Acids Res.* 23:269-276.
- Capel, B., A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow, and R. Lovell-Badge. 1993. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell.* 73:1019-1030.
- Chardard, D., A. Chesnel, C. Goz , C. Dournon, and P. Berta. 1993. Pw Sox1: the first member of the Sox gene family in Urodeles. *Nucleic Acids Res.* 21:3576.
- Chirgwin, J.M., A.E. Przybyla, J.R. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294-5299.
- Clep t, C., A.J. Schafer, A.H. Sinclair, M.S. Palmer, R. Lovell-Badge, and P.N. Goodfellow. 1993. The human *SRY* transcript. *Hum. Mol. Genet.* 2:2007-2012.
- Connor, F., E. Wright, P. Denny, P. Koopman, and A. Ashworth. 1995. The *Sry*-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res.* 23:3365-3372.
- Denny, P., S. Swift, N. Brand, N. Dabhade, P. Barton, and A. Ashworth. 1992a. A conserved family of genes related to the testis determining gene, *SRY*. *Nucleic Acids Res.* 20:2887.
- Denny, P., S. Swift, F. Connor, and A. Ashworth. 1992b. An *SRY*-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO (Eur. Mol. Biol. Organ.) J.* 11:3705-3712.
- Dunn, T.L., L. Mynen-Johnson, E.M. Wright, B.M. Hosking, P.A. Koopman, and G.E.O. Muscat. 1995. Sequence and expression of Sox18 encoding a new HMG-box transcription factor. *Gene (Amst.)* 161:223-225.
- Eddy, E.M., J.E. Welch, and D.A. O'Brien. 1993. Gene expression during spermatogenesis. In *Molecular Biology of the Male Reproductive System*. D. de Krestser, editor. Academic Press, CA 181-232.
- Erickson, R.P. 1990. Post-meiotic gene expression. *Trends Genet.* 6:264-269.
- Foster, J.W., M.A. Dominguez-Steglich, S. Guioli, C. Kwok, P.A. Weller, M. Stevanovi , J. Weissenbach, S. Mansour, I.D. Young, P.N. Goodfellow, J.D. Brook, and A.J. Schafer. 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an *SRY*-related gene. *Nature (Lond.)* 372:525-530.

- Foulkes, N.S., B. Mellström, E. Benusiglio, and P. Sassone-Corsi. 1992. Developmental switch of CREM function during spermatogenesis: from antagonist to activator. *Nature (Lond.)* 355:80–84.
- Fujita, T., G.P. Nolan, S. Ghosh, and D. Baltimore. 1992. Independent modes of transcriptional activation by the p50 and p65 subunits of NF- κ B. *Genes Dev.* 6:775–787.
- Gozé, C., F. Poulat, and P. Berta. 1993. Partial cloning of SOX-11 and SOX-12, two new human SOX genes. *Nucleic Acids Res.* 21:2943.
- Gubbay, J., J. Collignon, P. Koopman, B. Capel, A. Economou, A. Münsterberg, N. Vivian, P. Goodfellow, and R. Lovell-Badge. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature (Lond.)* 346:245–250.
- Harley, V.R., D.J. Jackson, P.J. Hextall, J.R. Hawkins, G.D. Berkovitz, S. Sockanathan, R. Lovell-Badge, and P.N. Goodfellow. 1992. DNA binding activity of recombinant SRY from normal males and XY females. *Science (Wash. DC)* 255:453–456.
- Harley, V.R., R. Lovell-Badge, and P.N. Goodfellow. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.* 22:1500–1501.
- Jantzen, H.-M., A. Admon, S.P. Bell, and R. Tjian. 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG box proteins. *Nature (Lond.)* 344:830–836.
- Jantzen, H.-M., A.M. Chow, D.S. King, and R. Tjian. 1992. Multiple domains of the RNA polymerase I activator hUBF interact with the TATA-binding protein complex hSL1 to mediate transcription. *Genes Dev.* 6:1950–1963.
- Jay, P., C. Gozé, C. Marsollier, S. Taviaux, J.-P. Hardelin, P. Koopman, and P. Berta. 1995. The human SOX11 gene: cloning, chromosomal assignment and tissue expression. *Genomics* 29:541–545.
- Kamachi, Y., S. Sockanathan, Q. Liu, M. Breitman, R. Lovell-Badge, and H. Kondoh. 1995. Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO (Eur. Mol. Biol. Organ.) J.* 14:3510–3519.
- Koopman, P., J. Gubbay, N. Vivian, P. Goodfellow, and R. Lovell-Badge. 1991. Male development of chromosomally female mice transgenic for *Sry*. *Nature (Lond.)* 351:117–121.
- Noce, T., Y. Fujiwara, M. Ito, T. Takeuchi, N. Hashimoto, M. Yamanouchi, T. Higashinakagawa, and H. Fujimoto. 1992a. A novel murine zinc finger gene mapped within the *r¹⁸* deletion region expresses in germ cells and embryonic nervous system. *Dev. Biol.* 155:409–422.
- Noce, T., Y. Fujiwara, M. Sezaki, H. Fujimoto, and T. Higashinakagawa. 1992b. Expression of a mouse zinc finger protein gene in both spermatocytes and oocytes during meiosis. *Dev. Biol.* 153:356–367.
- Poulat, F., F. Girard, M.-P. Chevron, C. Gozé, X. Rebillard, B. Calas, N. Lamb, and P. Berta. 1995. Nuclear localization of the testis determining gene product SRY. *J. Cell Biol.* 128:737–748.
- Rossi, P., S. Dolci, C. Albanesi, P. Grimaldi, and R. Geremia. 1993. Direct evidence that the mouse sex-determining gene *Sry* is expressed in the somatic cells of male fetal gonads and in the germ cell line in the adult testis. *Mol. Reprod. Dev.* 34:369–373.
- Ruppert, S., T.J. Cole, M. Boshart, E. Schmid, and G. Schütz. 1992. Multiple mRNA isoforms of the transcription activator protein CREB: generation by alternative splicing and specific expression in primary spermatocytes. *EMBO (Eur. Mol. Biol. Organ.) J.* 11:1503–1512.
- Saido, T.C., T. Iwatsubo, D.M.A. Mann, H. Shimada, Y. Ihara, and S. Kawashima. 1995. Dominant and differential deposition of distinct β -amyloid peptide species. A $\beta_{N3(pE)}$ in senile plaques. *Neuron* 14:457–466.
- Saido, T.C., S. Nagao, M. Shiramine, M. Tsukaguchi, H. Sorimachi, H. Murofushi, T. Tsuchiya, H. Ito, and K. Suzuki. 1992. Autolytic transition of μ -calpain upon activation as resolved by antibodies distinguishing between the pre- and post-autolysis forms. *J. Biochem.* 111:81–86.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
- Schilham, M.W., M. van Eijk, M. van de Wetering, and H.C. Clevers. 1993. The murine Sox-4 protein is encoded on a single exon. *Nucleic Acids Res.* 21:2009.
- Sinclair, A.H., P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A.-M. Frichauf, R. Lovell-Badge, and P.N. Goodfellow. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature (Lond.)* 346:240–244.
- Smith, D.B., and K.S. Johnson. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene (Amst.)* 67:31–40.
- Stevanovic, M., R. Lovell-Badge, J. Collignon, and P.N. Goodfellow. 1993. *SOX3* is an X-linked gene related to *SRY*. *Hum. Mol. Genet.* 2:2013–2018.
- Takamatsu, N., H. Kanda, I. Tsuchiya, S. Yamada, M. Ito, S. Kabeno, T. Shiba, and S. Yamashita. 1995. A gene that is related to *SRY* and is expressed in the testes encodes a leucine zipper-containing protein. *Mol. Cell Biol.* 15:3759–3766.
- Uwanogho, D., M. Rex, E.J. Cartwright, G. Pearl, C. Healy, P.J. Scotting, and P.T. Sharpe. 1995. Embryonic expression of the chicken *Sox2*, *Sox3* and *Sox11* genes suggests an interactive role in neuronal development. *Mech. Dev.* 49:23–36.
- van de Wetering, M., M. Oosterwegel, D. Dooijes, and H. Clevers. 1991. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO (Eur. Mol. Biol. Organ.) J.* 10:123–132.
- van de Wetering, M., M. Oosterwegel, K. van Norren, and H. Clevers. 1993. Sox4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO (Eur. Mol. Biol. Organ.) J.* 12:3847–3854.
- Waeber, G., T.E. Meyer, M. LeSieur, H.L. Hermann, N. Gérard, and J.F. Habener. 1991. Developmental stage-specific expression of cyclic adenosine 3',5'-monophosphate response element-binding protein CREB during spermatogenesis involves alternative exon splicing. *Mol. Endocrinol.* 5:1418–1430.
- Wagner, T., J. Wirth, J. Meyer, B. Zabel, M. Held, J. Zimmer, J. Pasantes, F.D. Bricarelli, J. Keutel, E. Hustert, U. Wolf, et al. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *Sox9*. *Cell* 79:1111–1120.
- Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan. 1993. Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.* 7:2120–2134.
- Wright, E., M.R. Hargrave, J. Christiansen, L. Cooper, J. Kun, T. Evans, U. Gangadharan, A. Greenfield, and P. Koopman. 1995. The *Sry*-related gene *Sox9* is expressed during chondrogenesis in mouse embryos. *Nature Genet.* 9:15–20.
- Wright, E.M., B. Snopek, and P. Koopman. 1993. Seven new members of the *Sox* gene family expressed during mouse development. *Nucleic Acids Res.* 21:744.