3982–3990 *Nucleic Acids Research, 2000, Vol. 28, No. 20*

Soggy, a spermatocyte-specific gene, lies 3.8 kb upstream of and antipodal to *TEAD-2*, a transcription factor expressed at the beginning of mouse development

Kotaro J. Kaneko* and Melvin L. DePamphilis

National Institute of Child Health and Human Development, Building 6, National Institutes of Health, Bethesda, MD 20892-2753, USA

Received June 5, 2000; Revised and Accepted August 8, 2000

DDBJ/EMBL/GenBank accession nos AF274312, AF274313

ABSTRACT

Investigation of the regulatory region of *mTEAD-2*, a gene expressed at the beginning of mouse pre-implantation development, led to the surprising discovery of another gene only 3.8 kb upstream of *mTEAD-2*. Here we show that this new gene is a single copy, testis-specific gene called *Soggy* (*mSgy*) that produces a single, dominant mRNA ~1.3 kb in length. It is transcribed in the direction opposite to *mTEAD-2*, thus placing the regulatory elements of these two genes in close proximity. *mSgy* contains three methionine codons that could potentially act as translation start sites, but most mSGY protein synthesis *in vitro* was initiated from the first Met codon to produce a full-length protein, suggesting that mSGY normally consists of 230 amino acids (26.7 kDa). Transcription began at a cluster of nucleotides ~150 bp upstream of the first Met codon using a TATA-less promoter contained within the first 0.9 kb upstream. The activity of this promoter was repressed by upstream sequences between –0.9 and –2.5 kb in cells that did not express *mSgy*, but this repression was relieved in cells that did express *mSgy*. *mSgy* mRNA was detected in embryos only after day 15 and in adult tissues only in the developing spermatocytes of seminiferous tubules, suggesting that *mSgy* is a spermatocyte-specific gene. Since *mTEAD-2* and *mSgy* were not expressed in the same cells, the *mSgy*/*mTEAD-2* locus provides a unique paradigm for differential regulation of gene expression during mammalian development.

INTRODUCTION

Formation of a two-cell mouse embryo marks the transition from maternal to zygotic gene dependence (reviewed in 1–3), a phenomenon characteristic of most, if not all, metazoa. In the mouse, transcription stops when oocytes undergo meiotic maturation to form unfertilized eggs and does not resume until after fertilization. The appearance of proteins whose synthesis is dependent on DNA transcription (i.e. zygotic gene expression) does not occur until after the fertilized egg (one-cell embryo) develops to the two-cell stage. Since only a small number of genes are transcribed at this time, their regulatory regions must exhibit some features that are distinctive. As a first step towards identifying those features, we sequenced the region upstream of the *mTEAD-2* gene, one of the first genes expressed at the beginning of mouse development, and searched for unusual characteristics.

The mouse TEAD/TEF family of transcription factors consists of four genes that are expressed to different extents in a variety of adult tissues, but only *mTEAD-2* is expressed during the first 7 days of embryonic development (4,5 and references therein). mTEAD-specific transcription factor activity is not detected in mouse oocytes, but it is detected as soon as zygotic gene expression begins at the two-cell stage in mouse development and then increases as embryos progress to the four-cell stage (4). These changes in TEAD transcription factor activity are accompanied by changes in *mTEAD-2* RNA, but not by changes in the expression of other *mTEAD* genes. While *mTEAD-1* and *mTEAD-3* RNAs are also detected in oocytes, they are 2- to 5-times less than *mTEAD-2* and they are completely degraded after fertilization. In fact, embryos homozygous for a disruption in mTEAD-1 survive long past the pre-implantation development stage (the first 4 days post-fertilization) (6), consistent with the initial appearance of significant levels of *mTEAD-1* RNA in embryos about day 8 (4). In contrast, ~5000 copies of *mTEAD-2* mRNA are present in each oocyte and this mRNA is recruited to polysomes only when development proceeds to the two-cell stage (7), coincident with the initial appearance of mTEAD transcription factor activity (4). *mTEAD-2* RNA is degraded in two- and four-cell embryos, but then increases 50-fold as pre-implantation embryos progress from the four-cell stage to blastocysts (4). These observations are consistent with degradation of maternally inherited mRNA in two-cell embryos followed by transcription of the *mTEAD-2* gene during pre-implantation development. Thus, *mTEAD-2* is one of the first genes expressed at the beginning of mouse development, where it presumably plays a role in activating transcription of other genes during pre-implantation development.

Efforts to identify sequences that regulate *mTEAD-2* transcription led to the unexpected discovery of a novel spermatocyte-specific gene that is closely linked to *mTEAD-2*.

*To whom correspondence should be addressed. Tel: +1 301 402 0410; Fax: +1 301 480 9354; Email: kjaneko@box-k.nih.gov

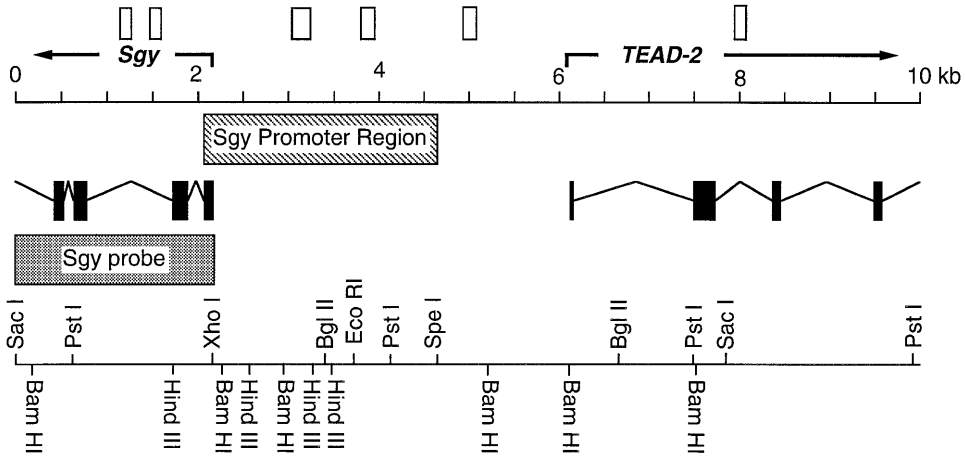

Figure 1. The 10 kb region encompassing the 5′-ends of the mouse *mSgy* and *mTEAD-2* genes. The transcription start sites (arrows), location of *Sgy* probe (grey box), the regions containing promoter (striped box) activity, the first four exons (solid boxes) of each gene, Alu/B1 repeat elements (open boxes) and relevant restriction endonuclease cleavage sites are indicated. The sequence of the 7866 bp *Sac*I–*Sac*I DNA fragment (including the GAGCTC *Sac*I sites at each end) is found under GenBank accession no. AF274313. *mSgy* exon 1 begins at +2166, transcribed in the 'minus' strand. *mTEAD-2* exons are described in Suzuki *et al.* (15), with exon 1 beginning at nt 6031. Thus, the *mSgy* and *mTEAD-2* genes are separated by only 3865 bp.

Recently, the same gene has been identified as a relative of the *Dickkopf* family of proteins and given the name *Soggy* (*mSgy*) (8). Here we show that *mSgy* is a single copy gene that lies only 3.8 kb upstream from *mTEAD-2* and is transcribed in the opposite direction. Moreover, we have shown that the expression patterns for these two genes differ dramatically, despite the fact that their regulatory sequences lie in close proximity. Thus, the mouse *mSgy/mTEAD-2* locus provides an excellent paradigm for elucidating mechanisms involved in differential gene expression during mammalian development.

MATERIALS AND METHODS

Cloning the mouse *Soggy* gene

Using PCR and primers (5′-GTTGGAGGTGATGGCAG-CCCT-3′ and 5′-ACAGGATGATCTTGCGACGG-3′) located within exon 2 of *mTEAD-2*, Genome Systems (St Louis, MO) screened a BAC library made from Sv129 mouse embryonic stem cells and identified one clone that contained the *mTEAD-2* gene (Clone address BACM-75i6). This BAC clone was then digested with either *Sac*I restriction endonuclease to release an ~7.8 kb DNA fragment or *Sac*I and *Xho*I restriction endonucleases to release ~5.7 and ~2.1 kb DNA fragments (Fig. 1). These DNA fragments were subcloned into either the *Sac*I site or the *Sac*I and *Xho*I sites of Bluescript KS (Stratagene) and both strands were sequenced using the ABI BigDye Terminator Cycle Sequencing kit (PE Biosystems, AGCT Inc.; GenBank accession no. AF274313). When these sequences were matched against a mouse expressed sequence tag (mEST) database using the BLASTN program (http://www.ncbi.nlm.nih.gov/blast/blast.cgi), five non-repeated sequences from the Stratagene mouse testis library produced significant alignment within the 2.1 kb *Sac*I–*Xho*I region, revealing the presence of a unique gene.

Initially, no significant homologies to known genes were detected using the non-redundant CDS translation databases GenBank, PDB, SwissProt, PIR and PRF, although later this gene was reported and named *Soggy* (*mSgy*) (8). Therefore, in an effort to identify this gene, a complete cDNA was isolated from the Marathon cDNA-ready mouse testis library (Clontech) using rapid amplification of cDNA ends (RACE) (9). The 5′-half of the cDNA was obtained by PCR amplification using a unique primer within *mSgy* (KK281, 5′-GGAACAGTCGA-CTAAAGCTTTGGAG-3′) and a primer (AP1; Clontech) that anneals to the linker-adapter present on both ends of the cDNAs in this library. PCR was performed for 30 cycles (94°C for 30 s, then 68°C for 2 min) using Platinum *Taq* polymerase (Life Technologies) according to the vendor's instructions. A single PCR product was observed. This product was cloned into pCR (Invitrogen) and sequenced. The same sequence was found in three independent clones, thus identifying the 5′-end of the *mSgy* mRNA. Similarly, the 3′-half of the cDNA was obtained using another unique primer (KK277, 5′-ACTGAG-GGTCTTGCTGCTGCT-3′) within *mSgy* and the AP1 primer.

Combinatorial PCR was done in order to obtain a single full-length *mSgy* cDNA clone from the two clones identified by 5′- and 3′-RACE (10). The locations of primers KK277 and KK281 meant that PCR products obtained by 5′- and 3′-RACE would overlap by 140 bp, thus allowing the PCR products to anneal during simultaneous PCR amplification of a mixture of the two DNA fragments. pCR clones containing the 5′- and 3′-RACE products were chosen such that the AP1 primer site was adjacent to a 'reverse' primer (5′-GGAAACAGCTATGAC-CATG-3′) site situated within the pCR vector. The 5′-RACE clone was amplified using the 'reverse' and KK281 primers and the 3′-RACE clone was amplified using the 'reverse' and KK277 primers. The resulting amplicons were purified using a Qiagen PCR purification kit and 1% of each DNA product was combined and amplified using only the 'reverse' primer. The resulting full-length PCR DNA product was digested with *Bam*HI (which cuts outside the cDNA), cloned into Bluescript KS and sequenced (GenBank accession no. AF274312).

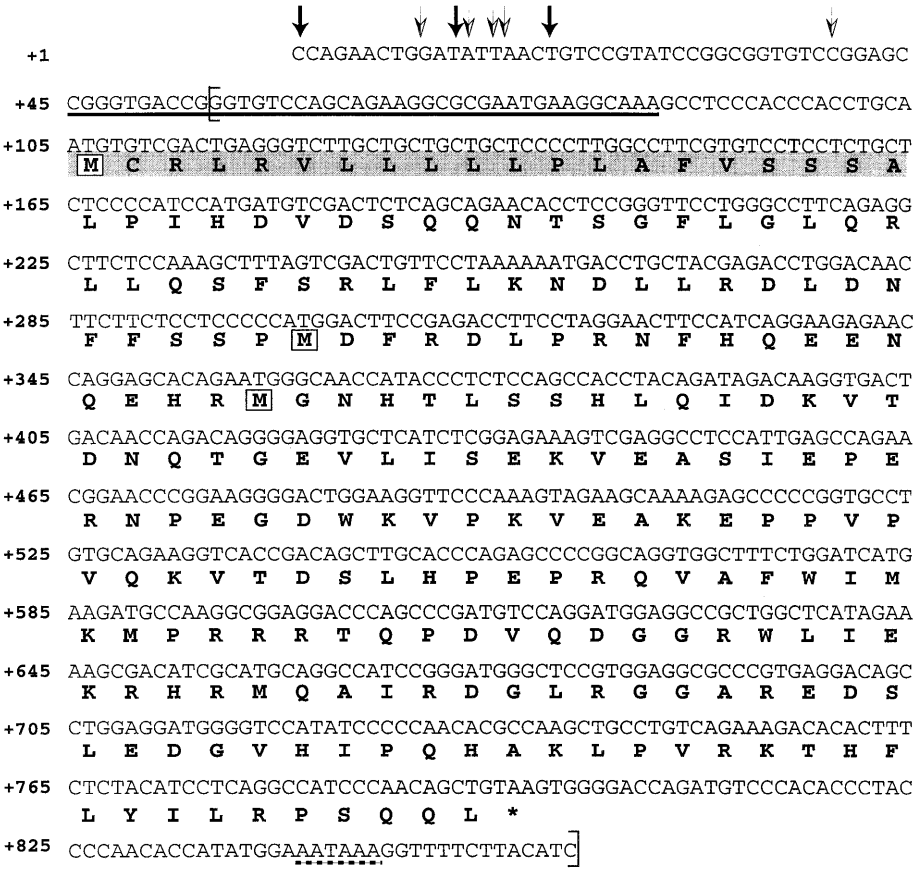

Figure 2. Nucleotide and amino acid sequences of mouse *mSgy* cDNA (GenBank accession no. AF274312). Transcription start sites are indicated by arrows, with the first transcription start site (nucleotide +1 in the cDNA) at nt 2166 in the *Sac*I–*Sac*I map (Fig. 1). The DNA primer used to identify transcription start sites in Figure 5 was complementary to the underlined sequence. Three potential translation start sites (ATG/Met) are indicated by a boxed M, the single termination signal (TAA) by an asterisk and the polyadenylation signal (AATAAA) by dashed underlining. A putative signal peptide is indicated by the shaded amino acids. The cDNA reported by Krupnick *et al.* (8) is identical to the cDNA described here, with three exceptions. First, their sequence begins at nt +55 and proceeds to nt +860 (brackets). Second, the first 8 nt of their cDNA (GGCACGAG) are not identical to the sequence reported here. Third, they report an A instead of a T at nt +427, which would change the codon for amino acid 108 from Leu to His. However, T is most likely correct, because it was present in both our cDNA and genomic clones.

RNA blotting–hybridization assays

The *mSgy* probe was the ~2.1 kb *Sac*I–*Xho*I genomic fragment containing the first four exons (Fig. 1). The *mTEAD-2* probe was the ~1.3 kb *Eco*RI fragment containing the entire coding region (4). Human β-actin cDNA was obtained from Clontech. The probes were radiolabeled with [α-^{32}P]dCTP using a Prime-It RmT labeling kit (Stratagene).

In vitro translation

Plasmid +1/+834, containing nt 1–834 of the *mSgy* cDNA (Fig. 2), was constructed by digesting the full-length *mSgy* cDNA clone with *Nde*I, filling in the ends using the Klenow DNA polymerase I fragment, digesting with *Not*I (Bluescript KS contains a single *Not*I site) and cloning into the *Not*I and *Sma*I sites of Bluescript KS. Plasmid +233/+834 was constructed by digesting plasmid +1/+834 with *Hin*dIII and cloning the released DNA fragment into the *Hin*dIII site of Bluescript KS. Plasmid +318/+834 was constructed by digesting plasmid +1/+834 with *Bln*I and *Hin*dIII and then cloning the released fragment into the *Spe*I and *Hin*dIII sites of Bluescript KS. These plasmids were translated *in vitro* using the TNT T7 Coupled Reticulocyte Lysate System (Promega) in the presence of [^{35}S]methionine (*in vivo* cell labeling grade; Amersham) according to the manufacturer's instructions. The reticulocyte lysate mix was fractionated by electrophoresis in a 12% Tris–HCl polyacrylamide gel (Bio-Rad) according to the manufacturer's instructions. Radiolabeled proteins were detected by fluorography using Enlightning (Dupont) according to the manufacturer's instructions.

Primer extension assay

Primer extension analysis was carried out essentially as described previously (11). The deoxyribonucleotide primer shown in Figure 2 was synthesized by Sigma-Genosys and purified by denaturing PAGE. Ten picomoles of primer was radiolabeled at its 5′-end using phage T4 polynucleotide kinase (Promega) and [γ-^{32}P]ATP. ^{32}P-labeled primer (1 pmol) was mixed with 5 μg total RNA, precipitated with ethanol in the presence of 0.3 M sodium acetate and redissolved in Hybridization Buffer, denatured at 85°C for 10 min and then re-annealed overnight at 35°C. The RNA:primer complex was precipitated with ethanol, resuspended in 20 μl of 20 mM Tris–HCl, pH 8.4,

50 mM KCl, 5 mM $MgCl_2$, 0.5 mM each dNTP, 10 mM dithiothreitol, 50 µg/ml actinomycin D and 200 U Superscript II reverse transcriptase (Life Technologies). DNA synthesis was carried out for 2 h at 42°C. RNase One (10 U; Promega) was added and the incubation continued for 30 min at 37°C. Nucleic acids were precipitated with ethanol, resuspended in loading buffer and then fractionated by electrophoresis in an 8% polyacrylamide gel containing 7 M urea. DNA sequencing was done (OmniBase DNA Cycle Sequencing System; Promega) using the same labeled oligonucleotide as used in the primer extension assay and a DNA sequence ladder was run in parallel with the [^{32}P]DNA products from primer extension.

Cell lines and transient assays for promoter activity

The following cell lines were obtained from the American Type Culture Collection: F9 (CRL-1720), TM3 (CRL-1714), TM4 (CRL-1715) and EL4 (TIB-39). Plasmid pGL3-Basic is the parent luciferase expression vector without a promoter and pGL3-promoter contains the firefly luciferase reporter gene driven by the SV40 promoter (Promega). DNA fragments from the *mSgy* promoter region were obtained by PCR using *pfu* Taq polymerase from Stratagene and appropriate DNA primers. An *Asc*I restriction site was added to the 5′-end of the downstream primer and a *Xba*I restriction site was ligated to each of the upstream primers. PCR DNA products were then digested with *Asc*I and *Xba*I, purified by gel electrophoresis and cloned into the *Mlu*I and *Nhe*I sites of pGL3-Basic. Transfection of adherent cells (TM3, TM4 and F9) was done as follows. About 4×10^5 cells were seeded in 6-well plates and transfected using Lipofectamine Plus (Life Technologies) according to the vendor's recommended protocol. The final DNA concentration was 1 µg DNA/2 ml Dulbecco's modified Eagle's medium (DMEM) and it contained 20 ng luciferase plasmid, 100 ng pCMV-SPORT-βgal (Life Technologies) and 880 ng Bluescript KS. For EL4 cells, 10 µg reporter plasmid along with 2.5 µg pCMV-SPORT-βgal (the total amount of DNA was adjusted to 40 µg with salmon sperm DNA) was electroporated. Approximately 1×10^7 cells in 300 µl of DMEM (Life Technologies) were combined with DNA mix and electroporated at 250 V, 1180 µF using a Cell-Porator apparatus (Life Technologies). After the cells were allowed to recover at room temperature for 5 min, the cells were transferred to a 100 × 20 mm tissue culture dish. After 48 h, the cells were harvested and lysed in 500 µl/well of 1× Reporter Lysis Buffer (Promega) and 50 µl aliquots were assayed for luciferase activity (12) and 30 µl aliquots were assayed for β-galactosidase activity (11).

RESULTS

The mouse *Soggy* (*mSgy*) gene lies upstream of the mouse *mTEAD-2* gene

In an effort to identify upstream sequences that regulate transcription of the mouse *TEAD-2* gene, ~10 kb of genomic DNA was sequenced that included the 5′-end of *mTEAD-2*. A novel gene was discovered ~3.8 kb upstream of *mTEAD-2* (Fig. 1) and cloned as described in Materials and Methods. The 860 bp cDNA (Fig. 2) contained three potential ATG translation initiation codons, one TAA termination codon, a signal for polyadenylation (AATAAA) and an open reading frame of 230 amino acids that encoded a protein with a theoretical size of 26.7 kDa and a pI of 7.98. Subsequent searches of the GenBank, PDB, SwissProt, PIR and PRF databases revealed that this gene had recently been identified as a member of the *Dickkopf* gene family and named *Soggy* (*mSgy*) (8).

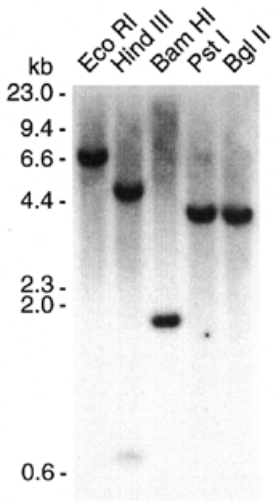

Figure 3. *mSgy* is a single copy gene. Mouse genomic DNA, digested with the indicated restriction endonuclease, fractionated by gel electrophoresis and then transferred to a membrane, was purchased from Clontech and hybridized with the *mSgy* ^{32}P-labeled probe indicated in Figure 1, as previously described (5).

Although *mSgy* mRNA was clearly expressed *in vivo* (see below; 8), it was possible that the *mSgy* gene existed at multiple genomic loci and that the copy located upstream of *mTEAD-2* was a pseudogene. Therefore, to determine whether or not *mSgy* exists as a single copy gene, mouse genomic DNA was digested with various restriction endonucleases and the DNA products were fractionated by gel electrophoresis, transferred to a membrane and then hybridized with a ^{32}P-labeled probe (Fig. 1) containing the first four exons of *mSgy*. As expected for a single copy gene with the restriction sites indicated in Figure 1, each of the five enzymes tested released a single intense band (Fig. 3). Digestion with either *Bgl*II or *Eco*RI released a single intense band that must have resulted from the *Bgl*II and *Eco*RI restriction sites upstream of *mSgy* and unidentified *Bgl*II and *Eco*RI sites downstream of the probe. Digestion with either *Bam*HI or *Pst*I released the expected fragments of ~2 and ~3.6 kb, respectively, resulting from one internal restriction site and one upstream external site. Additional fragments that might have been expected with *Pst*I and *Bam*HI from downstream unidentified restriction sites were not observed, presumably because the fragments produced lacked sufficient overlap with the probe. For example, *Hin*dIII released a strong ~5 kb fragment that resulted from the *Hin*dIII site within *mSgy* and the next unidentified *Hin*dIII site downstream of *mSgy*, as well as a faint ~0.8 kb fragment that resulted from the internal *Hin*dIII site and the next *Hin*dIII site upstream. This band was faint, presumably because it only partially overlapped with the probe. These results revealed that *mSgy* is a single copy gene that lies just upstream of *mTEAD-2*.

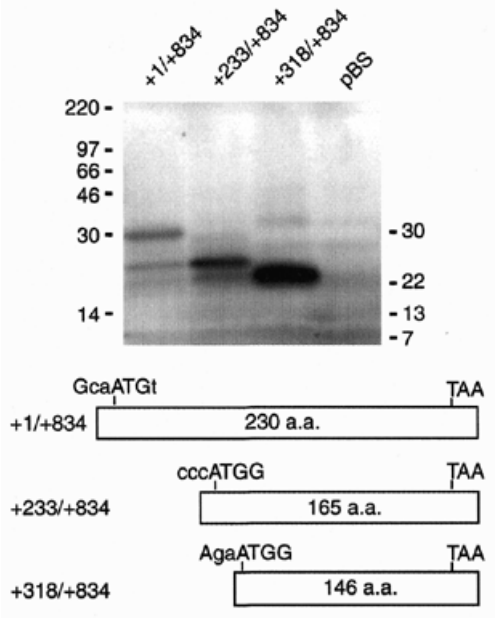

Figure 4. *mSgy* mRNA translation start sites. Plasmid DNA containing the indicated fragments of *mSgy* cDNA was translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. The protein products were fractionated by SDS gel electrophoresis in parallel with ^{14}C-labeled protein molecular weight standards (kDa scale; Amersham Pharmacia Biotech) and the radiolabeled proteins were detected by fluorography. pBS is the Bluescript parental plasmid control. Translational start sites are indicated and nucleotides in upper case match the Kozak consensus sequence.

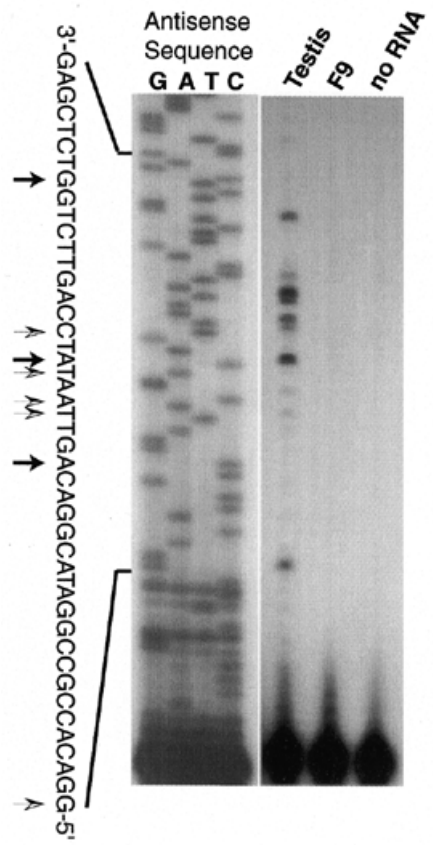

Figure 5. *mSgy* transcription start sites. Total RNA from adult mouse testis and from mouse F9 cells was hybridized to the [5′-^{32}P]DNA primer indicated in Figure 2 and extended towards the 5′-end of *mSgy* RNA by reverse transcriptase, as described in Materials and Methods. The nucleotide positions of the 5′-ends of *mSgy* mRNA were determined by direct comparison with the sequence of the antisense DNA. Dark arrows indicate strong start sites; open arrows indicate weak start sites.

Translation start sites for *mSgy*

Analysis of the *mSgy* cDNA sequence using the ATG program (http://www.hri.co.jp/atgpr/) selected the first methionine of the open reading frame as the most likely translation start site, although sequences surrounding this ATG deviated from the Kozak consensus sequence 4 nt downstream (+108), where a T was found instead of a G (Fig. 2). Therefore, it was possible that translation of this gene actually began further downstream. In fact, the ATG program suggested that Met66 and Met85 could also act as translation start sites. Therefore, to determine which Met codon is the primary translation initiation site, different fragments of the *mSgy* cDNA were translated in a rabbit reticulocyte lysate system. Nucleotides 1–834, containing the complete *mSgy* coding sequence, produced a protein of ~30 kDa as well as two minor proteins of ~23 and ~21 kDa (Fig. 4). The 30 kDa protein, initiated at Met1, accounted for ~75% of the total mSGY protein synthesized. Nucleotides 233–834 (beginning with Met66) efficiently produced only the two smaller proteins and nt 318–834 (beginning with Met85) produced only the smallest of the three proteins, confirming that all three Met codons could act as translational start sites. The amount of protein produced when the first two Met codons were deleted increased dramatically, consistent with the fact that the third Met codon matched a perfect Kozak consensus sequence and therefore resulted in efficient translation in the absence of upstream Met codons. Thus, in agreement with the scanning model of translation initiation (13), the primary translation initiation site is at the first Met codon, although secondary translation initiation events can also occur at either of the next two downstream Met codons, but at a reduced frequency.

Transcription start sites for *mSgy*

Transcription initiation sites for *mSgy* were mapped by primer extension analysis of RNA using the 5′-end of the cDNA sequence as a reference point. A 5′-^{32}P-labeled DNA primer complementary to the coding sequence was annealed to RNA with its 3′-end 44 bp downstream from the 5′-end of the mRNA originally detected by RACE (Fig. 2). This primer was then extended by reverse transcriptase in the direction of the 5′-end of the RNA under conditions that minimized secondary structure. No [^{32}P]DNA fragments were detected when the primer extension reaction was carried out either in the absence of RNA or with RNA from mouse F9 cells (Fig. 5), which do not express *mSgy* (Fig. 8). In contrast, eight [^{32}P]DNA fragments were detected when mouse testis RNA was used as the substrate (Fig. 5). One fragment began at the same G residue in the antisense strand that had been detected by the 5′-RACE reaction that generated this cDNA. However, most initiation events occurred 86–95 bp upstream of the first Met codon (Fig. 2).

Regulatory sequences for *mSgy*

As the first step towards identifying sequences that regulate *mSgy* transcription, sequences upstream of *mSgy* were tested for their ability to promote expression of a plasmid-encoded reporter gene following transfection of mouse cells. Various DNA fragments from the *mSgy* upstream region were cloned 67 bp upstream of the transcription start site of the firefly luciferase gene. These luciferase plasmids were then co-transfected into mouse cells together with another plasmid containing the β-galactosidase gene driven by the CMV promoter. The level of luciferase activity in each assay was then corrected for differences in transfection efficiency by normalizing it to a constant level of β-galactosidase activity. The results were expressed as stimulation above the level of luciferase activity observed in the absence of a promoter. SV40 T antigen promoter activity observed in each experiment (also normalized to β-galactosidase activity) was used as a standard for comparing results from several independent experiments. The effects of these sequences on luciferase gene activity was examined in four established mouse cell lines, of which only EL4 expressed its endogenous *mSgy* gene (described below).

The ~2.5 kb region (+77 to –2485) upstream of the *mSgy* gene stimulated luciferase expression 4- to 6-fold in the *mSgy* non-expressing cell lines F9, TM3 and TM4, consistent with the presence of a promoter in this region (Fig. 6). However, as sequences were deleted progressively from the –2485 terminus, promoter activity increased to a level of 15- to 25-fold above background (Fig. 6). These levels were within ±2-fold of the levels observed with the SV40 T antigen promoter in the same cells using the same assay (data not shown), suggesting that *mSgy* promoter elements lie between nt +77 and –894. Sequences further upstream repress this promoter activity. In support of this conclusion, this repression was relieved in mouse EL4 cells when sequences in the vicinity of –1344 to –2485 were present (Fig. 6). Since EL4 cells express their endogenous *mSgy* gene (Fig. 8), they appear to contain a transacting factor that can bind to this region and relieve the repression induced by sequences in the vicinity of –875 to –894.

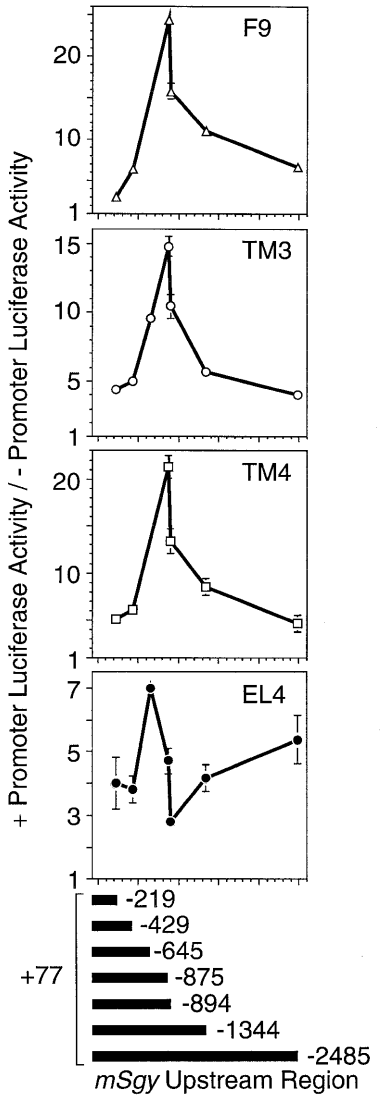

Figure 6. *mSgy* regulatory sequences. Putative *mSgy* promoter sequences from nt +77 to nt –219, –429, –645, –875, –894, –1344 and –2485 were tested for their ability to stimulate expression of the firefly luciferase gene following transfection of either F9 (triangle), TM3 (open circle), TM4 (square) or EL-4 (closed circle) cells, as described in Materials and Methods. Results were expressed as the ratio of luciferase activity observed from the promoter-driven luciferase plasmid relative to the luciferase activity produced by pGL3-Basic, the parent plasmid containing the luciferase gene without a promoter.

Expression of *mSgy* mRNA

The general expression pattern for *mSgy* in the adult mouse was determined by blotting–hybridization of RNA from various adult mouse tissues using ^{32}P-labeled DNA probes specific either for *mSgy* or β-*actin* mRNA. In adult mouse tissues *mSgy* mRNA was detected as a single major species of ~1.3 kb that was present only in the testis (Fig. 7A and B). In mouse embryos *mSgy* mRNA was not detected until day 15 (Fig. 7). Since sexual differentiation in mice begins around day 13, *in situ* hybridization was carried out on 13-day-old embryos in an effort to identify cells that first express the *mSgy* gene. However, although *mTEAD-2* mRNA was detected in this analysis, *mSgy* mRNA was not (data not shown). These data suggest that expression of *mSgy* begins after sexual differentiation occurs and that it is expressed only in the testis, consistent with the fact that the *mSgy* gene was identified from an expressed sequence tag library prepared from mouse testis.

Blotting–hybridization for *mSgy* mRNA was also used to determine if *mSgy* was expressed in established mouse cell lines. F9 embryonal carcinoma cells are pluripotent stem cells that can be stimulated to differentiate into parietal endoderm. TM3 cells were derived from testicular Leydig cells. TM4 cells were derived from testicular Sertoli cells. EL4 cells were established from a carcinogen-induced lymphoma. *mSgy* RNA was detected only in EL4 cells, where the level of *mSgy* expression was ~100-fold less than in mouse testis, as judged by the relative amounts of RNA applied to each gel lane and by the ratio of *mSgy* mRNA to β-*actin* mRNA (Fig. 8).

To identify the specific cell types that express the *mSgy* gene within the testis, *mSgy* mRNA was detected by *in situ* hybridization using the entire *mSgy* cDNA as probe. *mSgy* mRNA was abundant in all of the seminiferous tubules in the testis (Fig. 9A), where it was specifically associated with the developing spermatocytes (Fig. 9B and C). *mSgy* mRNA was absent

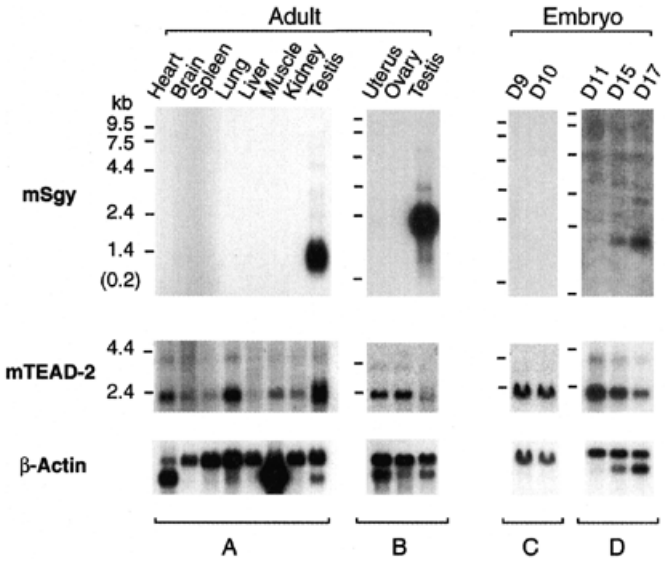

Figure 7. Expression of *mSgy* and *mTEAD-2* mRNA in mouse adult tissues and embryos. RNA from the indicated mouse tissue was fractionated by gel electrophoresis, transferred to a membrane and then hybridized with the indicated probe (*mSgy*, *mTEAD-2* or *β-actin*). Blots were probed first with the *mSgy* probe and then stripped and reprobed with the *mTEAD-2* probe and then with the *β-actin* probe. Each sample in (**A**), (**B**) and (**D**) contained 2 µg poly(A)$^+$ RNA. Each sample in (**C**) contained 20 µg total RNA. Uterus, ovary, day 9 and day 10 RNA samples were prepared as previously described (4,5). Mouse tissue RNAs in (A) are Clontech MTN Blot no. 7762-1. Mouse 11, 15 and 17 day embryo RNAs in (D) are Clontech MTN Blot no. 7763-1. Mouse testis RNA in (B) was purchased from Clontech (no. 6612-1).

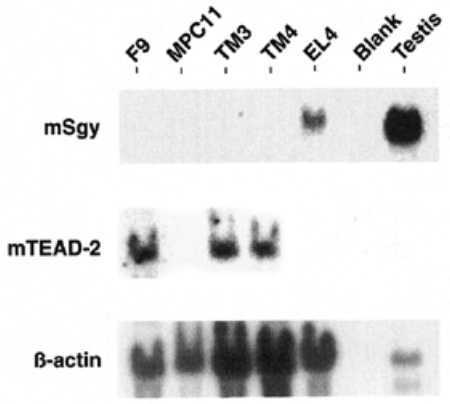

Figure 8. Expression of *mSgy* and *mTEAD-2* mRNA in established mouse cell lines. Blotting–hybridization analysis of RNA isolated from F9, TM3, TM4 and EL4 mouse cell lines was compared directly with the same sample of mouse testis RNA used in Figure 7B, as described in Figure 7. Each lane contained 20 µg total RNA from the indicated cell line, and 2 µg total RNA from testis. The blank lane contained no RNA.

from the darkly staining spermatogonia that line the basement membrane of the tubules, as well as from the mature spermatids that lie in the center of the tubules. Although Sertoli and Leydig cells cannot be identified with certainty in these sections, *mSgy* RNA was not detected by blotting–hybridization

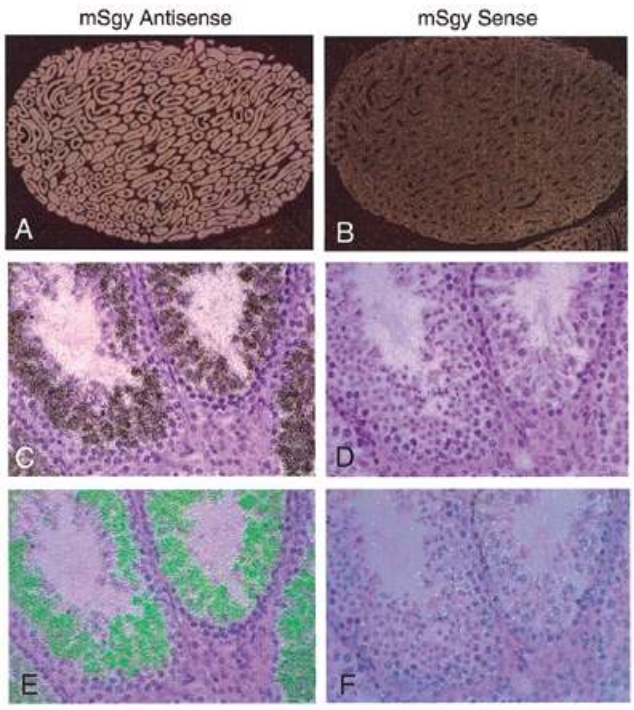

Figure 9. Expression of *mSgy* RNA in the cells of adult mouse testis. *In situ* RNA:RNA hybridization of sections through adult mouse testis was carried out using ^{35}S-labeled antisense (**A**, **C** and **E**) or sense (**B**, **D** and **F**) RNA probes specific for *mSgy* mRNA. (A and B) 5× dark field; (C and D) 40× bright field; (E and F) 40× bright field with AP polarization. Synthesis of ^{35}S-labeled riboprobes and hybridization protocols have been described (22) and were carried out by Molecular Histology Laboratory Inc. (Rockville, MD). Probes were synthesized from a Bluescript plasmid containing *mSgy* cDNA sequences +1 to +834 (Fig. 2).

in either TM3 cells (derived from Leydig cells) or TM4 cells (derived from Sertoli cells) (Fig. 8). Therefore, *mSgy* appears to be expressed exclusively in developing spermatocytes.

Expression of *mTEAD-2* mRNA

The general expression pattern for *mTEAD-2* in the adult mouse was determined by blotting–hybridization of the same RNA samples used above to detect *mSgy* and β-*actin*. Both the major ~2.4 kb and the minor ~4.2 kb *mTEAD-2* RNA species were present in all of the adult tissues examined, albeit at significantly different levels (Fig. 7A and B). In contrast to *mSgy* RNA, *mTEAD-2* RNA was easily detected in F9, TM3 and TM4 cells, but not in EL4 cells or mouse testis (Fig. 8), suggesting that only one of these two genes was expressed in a particular cell type.

Previously reported blotting–hybridization of RNA samples from adult mouse tissues concluded that *mTEAD-2* was expressed in the adult testis (4). These data were collected using a pre-made northern blot of mouse tissue RNA samples purchased from Clontech; the same results were obtained using a new blot purchased from the same company (Fig. 7A). However, the amount of *mTEAD-2* RNA in a sample of mouse testis RNA purchased from Clontech was ~10-fold less (Fig. 7B). This was not a hybridization artifact, because it was reproducible in several independent trials and the ratio of *mSgy*

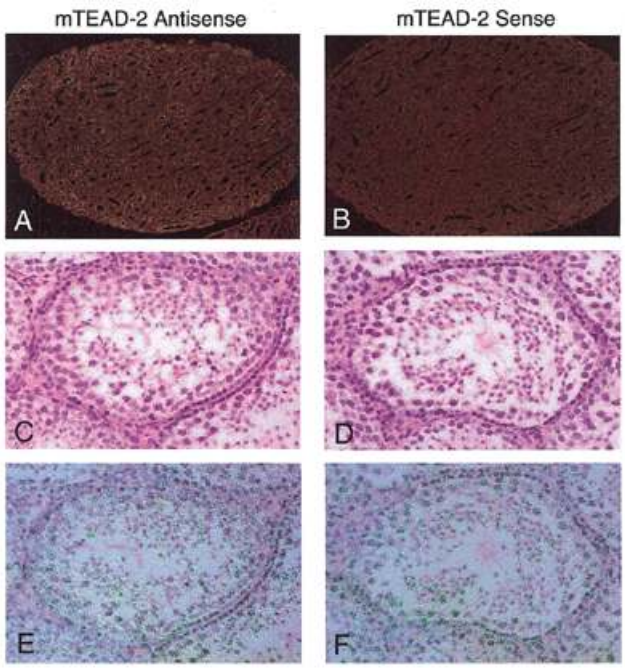

Figure 10. Expression of *mTEAD-2* RNA in the cells of adult mouse testis. *In situ* RNA:RNA hybridization of sections through adult mouse testis was carried out using ^{35}S-labeled antisense (**A**, **C** and **E**) or sense (**B**, **D** and **F**) RNA probes specific for *mTEAD-2* mRNA (described in ref. 4) as described in Figure 9.

RNA to *β-actin* RNA was essentially the same in all testis RNA samples. Therefore, the testis RNA used in the pre-made tissue blots from Clontech appeared to be contaminated with *mTEAD-2* RNA from another tissue.

To test this hypothesis, tissue slices through mouse testis were subjected to *in situ* hybridization using a probe specific for *mTEAD-2* mRNA. The results confirmed that *mTEAD-2* was expressed poorly in testis (Fig. 10A), particularly when compared with *mSgy* (Fig. 9A). Moreover, closer examination of seminiferous tubules failed to identify *mTEAD-2* expression in any particular cell type (Fig. 10C and E). Therefore, we conclude that *mTEAD-2* is expressed poorly, if at all, in those cells that strongly express the *mSgy* gene.

DISCUSSION

Several promoters have been identified in mammals that function bi-directionally (14 and references therein). In each case, the two antipodal genes are coordinately expressed and their gene products either have similar functions or participate in the same metabolic pathway. The data presented here reveal the first example, to our knowledge, of two antipodal mammalian genes whose regulatory sequences lie in close proximity but whose expression patterns differ dramatically. *mSgy* is a single copy gene that lies 3.8 kb upstream from *mTEAD-2*, which has been mapped to chromosome 7 (15). *mSgy* is transcribed in the opposite direction from *mTEAD-2* and its regulatory sequences occupy >50% of the intergenic region (Fig. 1). Remarkably, results from RNA blotting–hybridization and *in situ* hybridization analyses revealed that *mSgy* and *mTEAD-2* were differentially regulated both temporally and spatially. *mSgy* was expressed late during embryonic development and then only in the spermatocytes of the adult testis. *mTEAD-2* was expressed at the very beginning of mouse development and then later in a variety of adult tissues, but little, if at all, in the testis. Moreover, cell lines that expressed one of these genes did not express the other. Thus, the *mSgy/mTEAD-2* locus provides a unique paradigm for differential regulation of gene expression during mammalian development.

This regulation likely involves positive as well as negative transcriptional modulators that specifically affect each gene. The promoter/enhancer region for the *mSgy* gene was localized to nt +77 to –894. Levels of promoter activity of 4- to 6-fold above background were observed with sequences proximal to the transcription start sites (+77 to –429). However, addition of sequences from –429 to –875 increased promoter activity to 15- to 25-fold above background (Fig. 6), comparable to the activity observed using the SV40 T antigen promoter. Surprisingly, in cells that do not express their endogenous *mSgy* gene (F9, TM3 and TM4), further extension of this region resulted in decreasing levels of promoter activity, suggesting that the region between –875 and –2485 contained sequences that repressed the *mSgy* promoter. In contrast, in cells that do express their endogenous *mSgy* gene (EL4), promoter activity resided between +77 and –645. Extension of this region to –894 resulted in repression of promoter activity, while further extension relieved this repression. Therefore, cells that express the *mSgy* gene appear to harbor the same repression mechanism for *mSgy* that exists in non-producing cells, except that *mSgy*-producing cells appear to relieve this repression by producing a factor that interacts with sequences in the vicinity of –1344 to –2485.

Previous analysis of *mTEAD-2* transcription by S1 nuclease protection and primer extension revealed multiple transcription initiation sites and the absence of a TATAA consensus sequence (15). Similarly, transcription of *mSgy* produced a single ~1.3 kb RNA that began at several different nucleotide locations (Fig. 5). Some initiation events occurred at nt +1 in the cDNA sequence, as expected, and some as far away as nt +39, but most initiation events occurred from nt +12 to +19, consistent with the fact that a TATAA consensus sequence was not present in or around this region. Therefore, both the *mTEAD-2* and *mSgy* genes are driven by a TATA-less promoter.

Since these two genes are transcribed in opposite directions, their regulatory sequences must lie in close proximity. In fact, results from transient expression assays, like the ones used here to define the *mSgy* promoter region, reveal that the *mTEAD-2* regulatory region lies within the *Spe*I–*Pst*I DNA fragment that overlaps the transcription start site for *mTEAD-2* (Fig. 1; K.Kaneko, unpublished results). Therefore, one might expect *mSgy* and *mTEAD-2* to share a common pattern of expression, similar to previous reports where two neighboring genes that are driven by a single bi-directional promoter appear to be regulated coordinately. However, the expression patterns for *mSgy* and *mTEAD-2* were strikingly different.

Taken together, the results from RNA blotting–hybridization and RNA *in situ* hybridization reveal that the spatial and temporal patterns of expression for *mSgy* differ dramatically from those for *mTEAD-2* and suggest that mouse cells can express either *mSgy* or *mTEAD-2*, but not both. *mSgy* expression

does not begin until about day 15 of embryonic development (2 days after sex differentiation begins) and then appears to be confined exclusively to developing spermatocytes in the seminiferous tubules of the adult testis. Several studies of mammalian spermatogenesis have shown that RNA synthesis occurs primarily in primary spermatocytes during the pachytene stage of meiotic prophase (16). In contrast, *mTEAD-2* transcription begins at the four- to eight-cell stage of pre-implantation development (2 days after fertilization) and then continues to be expressed in various adult tissues such as heart, lung, muscle and kidney. The amount of *mTEAD-2* mRNA in testis was minimal to non-detectable. The original RNA blotting–hybridization results using a commercial blot (Fig. 7A; 4) revealed significant levels of *mTEAD-2* expression in testis, but subsequent RNA blotting–hybridization of independent samples of testis RNA revealed little *mTEAD-2* expression when poly(A)$^+$ RNA was used (Fig. 7B) and no *mTEAD-2* expression when total RNA was used (Fig. 8). These conclusions were confirmed by *in situ* RNA hybridization of thin sections through testis in which cell-specific *mTEAD-2* expression was not detected (Fig. 10), particularly in comparison with *mSgy* expression (Fig. 9). Therefore, we conclude that *mSgy* and *mTEAD-2* are not expressed in the same cell types. Since *mTEAD-2* is expressed strongly in ovaries where most of its mRNA is confined to the granulosa cells within the ovarian follicles (Fig. 7B; 4), it is possible that *mSgy* is involved in the production of sperm while *mTEAD-2* is involved in the production of oocytes.

The close proximity of the regulatory regions of *mSgy* and *mTEAD-2* underscores the fact that regulators of their respective transcriptions do not interfere with one another. This is dramatically revealed in developing spermatocytes, where *mSgy* is strongly expressed while *mTEAD-2* expression is barely detectable (Figs 9 and 10). One possible mechanism of how this is achieved is through the existence of insulator-like sequences between the two genes. Insulator sequences isolate genes from controlling elements in adjacent genes (reviewed in 17). For example, the ~8 kb region that separates the *iab*-7 and *iab*-8 regulatory regions within the bithorax complex allow those regulatory regions to function autonomously in flies (18). Recently, Alu/B1 elements have been reported to exert position-independent and copy number-dependent expression on heterologous transgenes in mice (19), suggesting that Alu/B1 elements can act as insulators. Since several Alu/B1 elements are found in the 3.8 kb region between the *mSgy* and *mTEAD-2* genes (Fig. 1), they may allow these two closely linked genes to be differentially expressed.

Although the function of the mSGY protein is not known, its partial homology to the *Dickkopf* (*Dkk*) gene family suggests that it may function as an inhibitor of Wnt proteins. Wnt proteins are a vertebrate family of secreted glycoproteins that mediate cell proliferation and fate determination in both adults and embryos (20). Dkk is a family of secreted proteins that antagonize Wnt function by binding to Wnt proteins (8). Interestingly, *mWnt-1*, like *mSgy*, is expressed only in the developing spermatocytes of testis in adult mice (21), suggesting that *mSgy* may be involved in regulation of Wnt function during spermatogenesis. However, if the same pattern of *mSgy* translation occurs *in vivo* as it does *in vitro*, *mSgy* mRNA may produce both secreted and non-secreted forms of the protein, each of which may have a distinct function.

REFERENCES

1. Schultz,R.M., Davis,W.,Jr, Stein,P. and Svoboda,P. (1999) *J. Exp. Zool.*, **285**, 276–282.
2. Latham,K.E. (1999) *Int. Rev. Cytol.*, **193**, 71–124.
3. Nothias,J.Y., Majumder,S., Kaneko,K.J. and DePamphilis,M.L. (1995) *J. Biol. Chem.*, **270**, 22077–22080.
4. Kaneko,K.J., Cullinan,E.B., Latham,K.E. and DePamphilis,M.L. (1997) *Development*, **124**, 1963–1973.
5. Kaneko,K.J. and DePamphilis,M.L. (1998) *Dev. Genet.*, **22**, 43–55.
6. Chen,Z., Friedrich,G.A. and Soriano,P. (1994) *Genes Dev.*, **8**, 2293–2301.
7. Wang,Q. and Latham,K.E. (2000) *Biol. Reprod.*, **62**, 969–978.
8. Krupnik,V.E., Sharp,J.D., Jiang,C., Robison,K., Chickering,T.W., Amaravadi,L., Brown,D.E., Guyot,D., Mays,G., Leiby,K., Chang,B., Duong,T., Goodearl,A.D., Gearing,D.P., Sokol,S.Y. and McCarthy,S.A. (1999) *Gene*, **238**, 301–313.
9. Zhang,Y. and Frohman,M.A. (1997) *Methods Mol. Biol.*, **69**, 61–87.
10. Higuchi,R. (1989) In Erlich,H.A. (ed.), *PCR Technology, Principles and Applications for DNA Amplification*. Stockton Press, New York, NY, pp. 61–70.
11. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. Majumder,S., Zhao,Z., Kaneko,K. and DePamphilis,M.L. (1997) *EMBO J.*, **16**, 1721–1731.
13. Kozak,M. (1996) *Mamm. Genome*, **7**, 563–574.
14. Orii,K.E., Orii,K.O., Souri,M., Orii,T., Kondo,N., Hashimoto,T. and Aoyama,T. (1999) *J. Biol. Chem.*, **274**, 8077–8084.
15. Suzuki,K., Yasunami,M., Matsuda,Y., Maeda,T., Kobayashi,H., Terasaki,H. and Ohkubo,H. (1996) *Genomics*, **36**, 263–270.
16. Eddy,E.M. and O'Brien,D.A. (1998) *Curr. Top. Dev. Biol.*, **37**, 141–200.
17. Bell,A.C. and Felsenfeld,G. (1999) *Curr. Opin. Genet. Dev.*, **9**, 191–198.
18. Barges,S., Mihaly,J., Galloni,M., Hagstrom,K., Muller,M., Shanower,G., Schedl,P., Gyurkovics,H. and Karch,F. (2000) *Development*, **127**, 779–790.
19. Willoughby,D.A., Vilalta,A. and Oshima,R.G. (2000) *J. Biol. Chem.*, **275**, 759–768.
20. Orsulic,S. and Peifer,M. (1996) *Curr. Biol.*, **6**, 1363–1367.
21. Shackleford,G.M. and Varmus,H.E. (1987) *Cell*, **50**, 89–95.
22. Fox,C.H. and Cottler-Fox,M. (1993) In Coligan,J.E., Kruisbeek,A.M., Margulies,D.H., Shevach,E.M. and Strober,W. (eds), *Current Protocols in Immunology*. Greene Publishing and Wiley-Interscience, New York, NY, pp. 12.8.1–12.8.21.