CHAPTER I:

Characterization Of The Intron-Containing Human Sperm Protein 17 Gene

CHAPTER I: Characterization of the Intron-Containing Human Sperm Protein 17 Gene

1. Abstract

The human Sp17 was first described as a spliced cDNA (Accession number Z48570), characterized by alternative 5'UTRs. However, sequence analyses revealed inconsistencies in the Sp17 nucleotide sequence. To refute the published Sp17 cDNA sequences the Sp17 transcript was characterized in human testis tissue. RT-PCR and sequence analyses were used to elucidate the complete human Sp17 nucleotide sequence. 5' RLM-RACE PCR revealed a Sp17 cDNA varying by alternative transcription start sites. In addition, alternative polyadenylation was observed at the 3' end of the Sp17 transcript by 3' RACE. Collectively, these findings support a modified view of the Sp17 nucleotide sequence and the identification of the Sp17-1a and Sp17-1b transcripts, which are differentiated by alternative transcriptional start sites. Moreover, sequence analysis of the Sp17-1 gene suggests the existence of an intronless Sp17 gene variant.

2. Introduction

Sp17 is a highly conserved, autoantigenic protein whose primary function is to bind sperm to the zona pellucida of oocytes. Consequently, Sp17 has been targeted in the development of male contraceptives (1).

The published human Sp17 nucleotide sequence was determined by screening a human testis cDNA library with a rabbit Sp17 probe (2). Two human Sp17 cDNAs were detected and characterized by alternative 5'UTRs (Introduction, Figure 1) (2). However, sequence analyses of the published human Sp17 5'UTR sequences revealed inconsistencies in the published Sp17 cDNA sequences. For example, the 1.6kb Sp17 cDNA, cDNA A, exhibits 98% homology to RNA helicase. RNA helicases are known to mediate nucleoside triphosphate-dependent unwinding of double-stranded RNA (3). Although this sequence may not encode Sp17, RNA helicases are prevelant in testis cDNA libraries, suggesting that the human Sp17 5'UTR sequence, may be artifact.

Similarly, alternative 5'UTRs were described for the rabbit Sp17 cDNA sequence (1.1kb and 0.9kb) (4). However, only one of the two published rabbit 5'UTR sequences exhibit homology in a multi-mammal Sp17 cDNA alignment. Thus, whether the proposed 5'UTR sequences for the human and rabbit describe real cDNA variants or represent a library artifact has yet to be established.

Furthermore, extensive sequence analyses, alignment and genomic comparison of the published human Sp17 transcript, suggests the existence of both an intron-containing and an intronless form of the Sp17 gene.

Thus, we hypothesize that the published human Sp17 spliced cDNA represents an incomplete Sp17 nucleotide sequence (Accession number Z48570). To refute the published Sp17 seqeunce, RT-PCR and sequence analyses were used to characterize the human Sp17 transcript and elucidate the complete Sp17 nucleotide sequence in the human testis tissue (2). 5' RLM-RACE and 3' RACE PCR were used to reveal the Sp17 5'UTR and 3'UTR

nucleotide sequences, respectively. Furthermore, to identify conserved Sp17 sequences, a sequence alignment was performed for the known mammalian Sp17 cDNA sequences.

3. Materials and Methods

3.1 RNA isolation

Normal human testes were obtained from the UCD Donated Body Program (UCD biological use authorization #0657). The snap frozen testis (50-100 mg) were thawed and homogenized in TRIzol reagent (Life Technologies, Inc., Frederick, MD) according to manufacturer guidelines. The total RNA was extracted by sequential treatment with chloroform, isopropanol and ethanol. The total RNA precipitate was resuspended in sterile 0.1% diethylpyrocarbonate (DepC; Sigma, St. Louis, MO) treated water. The RNA suspension was then stored in DepC treated 70% ethanol at -70^{o}C.

The integrity of the total RNA samples was observed by formaldehyde gel electrophoresis (1.1% agarose, 2.2M formaldehyde, 0.5ug/mL ethidium bromide) and UV illumination. Sample extracts which visually exhibited both 18S and 28S ribosomal bands were spectrophotometrically analyzed at an absorbance of 260nm and 280nm on a BioSpec-1601 DNA/Protein/Enzyme Analyzer (Shimadzu, Kyoto, Japan). Extracted RNA exhibiting clear ribosomal bands and had a calculated 260:280nm ratio greater than 1.8 were used. The concentration of total RNA was also calculated by spectrophotometrically measuring the optical density of the suspension at absorbance 260nm and 280nm.

3.2 DNase and Reverse transcription (RT)

Total RNA aliquots (4ug) were precipitated with ethanol and 3M NaOac and resuspended in DepC water. Prior to reverse transcription, the RNA samples were treated with 2.5U DNase I Amplification grade reagent (Life Technologies, Inc.) according to the manufacturer's guidelines. To verify the absence of genomic DNA, reverse transcription (RT) was withheld from aliquots of DNased treated RNA and non-DNased RNA. Similarly, a control without RNA template was treated with DNase and reverse transcribed in parallel. These controls are outlined in Table 1.1.

The DNased treated RNA and controls were converted to cDNA by reverse transcription, as required, using Superscript II reagent (Life Technologies, Inc.) according to manufacturer guidelines. Briefly, the samples were heated at 42^{o}C for 50 minutes (final concentration: 4ug RNA, 20mM Tris-HCl (pH 8.4), 50mM KCl, 2.5 mM $MgCl_2$, 10mM DTT, 25uM Oligio (dT) primer, 400uM dATP, 400uM dCTP, 400uM dGTP, 400uM dTTP, 200 units Superscript II). At the completion of the RT step, the RT enzymes were inactivated by incubating the samples at 70^{o}C for 10 minutes. Subsequently, to remove RNA complementary to the cDNA, the samples were incubated with 2 units Rnase H (Life Technologies, Inc.) at 37^{o}C for 30 minutes. The cDNA samples and controls were stored at -20^{o}C.

3.3 Polymerase chain reaction (PCR)

The Sp17 transcript was amplified from cDNA (section 3.2). Sp17 sequence specific primers were designed to amplify the Sp17 transcript and to differentiate between Sp17 variants (Sp17-1 and Sp17-2) and Sp17 transcriptional start sites (Sp17-1a and Sp17-1b) (Table 1.2). All primers were synthesized by Biosynthesis, Inc. (Lewisville, TX) and resuspended to 100pmol in sterile water.

The PCR cocktail contained 200nM sense primer, 200nM anti-sense primer, 4uL cDNA template, 200uM dATP, 200uM dCTP, 200uM dGTP, 200uM dTTP, 1X GeneAmp PCR Buffer II and 1.5units AmpliTaq Gold DNA Polymerase in Buffer II (Applied Biosystems, Foster City, CA). The PCR cocktails were pre-heated at 95^{o}C (10 minutes) and sequentially heated at 94^{o}C (1 minute), 60^{o}C (1 minute) and 72^{o}C (2 minutes) for 35 cycles. After 20 cycles, the 72^{o}C elongation step was extended by 15 seconds for each succeeding cycle followed by a final 72^{o}C (10 minutes) extension cycle.

As needed, a second round of PCR amplification, using identical conditions, was performed to visualize the PCR amplified DNA fragments. The PCR cocktails were prepared with 2uL of the previously amplified PCR reaction substituted for the cDNA template.

The PCR products were mixed with 6X DNA gel loading dye (0.25% bromphenol blue, 0.25% xylene cyanole FF, 0.5M EDTA (pH 8.0), 50% glycerol) and were subjected to gel electrophoresis adjacent to a 250bp DNA molecular weight marker (Life Technologies, Inc.) on a 1.5% agarose gel (1.5% agarose, 1X TAE (40mM Tris-acetate (pH 8.0), 2.5mM EDTA (pH 8.0)), 1ug/mL ethidium bromide). The gel was run in 1X TAE at 85 volts for one hour. The results of the PCR gel were visualized with UV light and photographed.

3.4 RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE)

The RLM-RACE procedure was performed according to manufacturer guidelines (GeneRacer™ Kit; Invitrogen, Carlsbad, CA). Briefly, to dephosphorylate mRNA, 1ug of total human testis RNA was treated with 10 U of calf intestinal phosphatase containing 50mM Tris-HCl (pH 8.5), 0.1mM EDTA and 10U of RNase inhibitor. The reaction was incubated at 50^{o}C for 60 minutes. Subsequently, to remove the cap structure from full-length mRNA, the RNA was treated with 0.5U of tobacco acid pyrophosphatase containing 50mM sodium acetate (pH 6.0), 1mM EDTA, 0.1% β-mercaptoethanol, 0.01% Triton X-100 and 10U of RNase inhibitor. The reaction was incubated at 37^{o}C for 60 minutes. A synthetic RNA adapter was ligated to the RNA sample using 5U of T4 RNA ligase containing 0.25ug GeneRacer ™ RNA oligo, 33mM Tris-acetate (pH 7.8), 66mM potassium acetate, 10mM magnesium acetate, 0.5mM DTT, 1mM ATP and 10U of RNase inhibitor. The reaction was incubated at 37^{o}C for 60 minutes. After each enzyme reaction, the RNA was purified by phenol/chloroform extraction and ethanol precipitation.

A random-primed reverse transcription reaction was then performed by adding 0.1ug random primer (N_6) to the RNA. After pre-incubation at 70°C for 5 minutes,

5U of avian myeloblastosis virus reverse transcriptase containing 25mM Tris-HCl (pH 8.3), 5mM $MgCl_2$, 50mM KCl, 2mM DTT, dNTP solution (1.25mM each) and 10U of RNase inhibitor was added to the reaction. The cocktail was incubated at 42^{o}C for 60 minutes. The enzymes were heat-inactivated at 85^{o}C for 15 minutes.

PCR was performed as described in section 3.3 with the following modifications: 1uL treated cDNA cocktail, 200nM GeneRacer ™ 5' sense primer and 200nM Sp17 specific anti-sense primer (Table 1.3). Sp17 specific primers were synthesized by Biosynthesis, Inc. (Lewisville, TX) and resuspended to 100pmol in sterile water. The appropriately sized PCR amplified DNA bands were excised from the agarose gel, purified, subcloned and sequenced as described in section 3.6 and 3.7.

3.5 3' RACE PCR

A 1ug aliqout of total testis RNA was converted into cDNA by reverse transcription using an Oligio (dT) primer as described in section 3.2. The cDNA was amplified according to the guidelines in Section 3.3 with the following modifications: 200nM Shu sense primer, 200nM anti-sense anchor primer and 4uL testis cDNA (Table 1.4). Appropriately sized PCR amplified DNA bands were excised, purified, subcloned and sequenced as described in section 3.6 and 3.7.

3.6 cDNA cloning and sequencing

Appropriately sized PCR amplified DNA bands were excised from the agarose gel and purified using GeneClean (Bio 101, Vista, CA). The purified DNA fragments were subcloned into a pCR 2.1 TOPO vector according to the guidelines of the TOPO TA Cloning Kit (Invitrogen). Potential positive clones were identified by blue/white selection (X-Gal; Life Technologies, Inc.). Positive clones were further verified by PCR amplification using TOPO vector specific sense and anti-sense primers (Invitrogen). The positive clones were purified using a Quiagen Mini-Prep Plasmid Purification Kit (Quiagen, Valencia, CA) according to manufacturer guidelines. The DNA inserts were sequenced from the forward (M13R) and reverse (T7) directions using ABI BigDye Terminator chemistry (Applied Biosystems) by the DBS Automated DNA Sequencing Facility (University of California, Davis, CA).

3.7 Sp17 sequence analysis

cDNA sequence analysis, assembly and predicted translation was performed using Gene Jockey 1.3 software (BioSoft Ltd., Cambridge, UK). DNA sequence comparisons were performed using NCBI's BLAST 2.1 search and BLAST 2 sequence software and ClustalW 1.8 DNA analysis software (5). Promotor analysis was performed using the TFSEARCH program, MatInspector program and TRANSFAC database (6, 7).

3.8 RT-PCR amplification of the Sp17 gene intron 1

The Sp17 intron 1 was amplified from genomic DNA by PCR. Genomic DNA was extracted from human polymorphonuclear blood cells (pmbcs). Briefly, whole

blood (20mL) was phlebotomically drawn from one normal volunteer into two acid citrate dextrose vacuum tubes (ACD solution A; Becton Dickinson Vacutainer Systems, Franklin Lakes, NJ). The whole blood was diluted 1:1 (volume to volume), with Hanks buffered saline solution (containing no Calcium chloride, no Magnesium chloride and no Magnesium sulfate; Life Technologies). The diluted blood was carefully layered on an equal volume of Histopaque 1077 (Sigma) and was density centrifuged (2000 rpm) at room temperature for 30 minutes. The buffy coat was isolated and centrifuged (2000 rpm) at room temperature for 20 minutes.

The pmbc cell pellet (5×10^6 cells) was resuspended and lysed in DNAzol (Molecular Research Center, Inc., Cincinnati, OH). The genomic DNA was extracted by ethanol precipitation and resuspended in sterile deionized water according to manufacturer guidelines. The total DNA concentration was calculated by spectrophotometrically measuring the optical density of the suspension at an absorbance of 260nm and 280nm on a BioSpec-1601 DNA/Protein/Enzyme Analyzer (Shimadzu, Kyoto, Japan). The total genomic DNA sample was stored at 4^oC.

The PCR reactions were performed according to the guidelines in Section 3.3 with the following modifications: 200nM sense primer, 200nM anti-sense primer and 1uL human genomic DNA template (Table 1.5). Appropriately sized PCR amplified DNA bands were excised, purified, subcloned, sequenced and analyzed as described in section 3.6 and 3.7.

4. Results

4.1 Sequence analysis: Rejection of the published human Sp17 cDNA model

RT-PCR and sequence analysis was used to refute the existence of the published Sp17 cDNA sequence (Accession number Z48570) and to elucidate the complete human Sp17 nucleotide sequence (2, 4). RT-PCR amplification was performed, using primers generated from the published Sp17 sequence. Subsequently, the DNA products were sequenced and compared to the published Sp17 cDNA sequence.

RT-PCR amplification of the Sp17 5'UTR through the coding region was not detected in human testis cDNA (Sp17 specific primer combinations of SUTR1, SUTR2, SUTR3 and ACDS) (Figure 1.1AB). However, RT-PCR amplification within the Sp17 coding region resulted in a 239bp DNA fragment that was identical to the published Sp17 coding region (Sp17 specific primers SCDS and ACDS) (Figure 1.1AB).

RT-PCR and sequence analyses within the 5'UTRs of Sp17 cDNA A and cDNA B, revealed non-Sp17 homology. For example, RT-PCR amplification within the 5'UTR of Sp17 cDNA A, revealed a DNA band (436bp) exhibiting 98% homology to human ATP-dependent RNA helicase mRNA (DDX24; Accession number AF214731) (Sp17 specific primers SUTR1 and AUTR1) (Figure 1.1A) (8). Moreover, sequence analysis of the published Sp17 cDNA A (position 1-717) exhibited 98% homology to DDX24 mRNA. In addition, the published Sp17 cDNA

A sequence was >98% homologous to two genomic BAC clones, clone R-1089B7 (Accession number AL079302) and clone BAC 367P10 (Accession number AC007515) on human chromosome 14 (14q31).

Similarly, RT-PCR amplification within the 5'UTR of Sp17 cDNA B, revealed a DNA band (157bp) identical to a human tyrosine phosphatase mRNA (Accession number AF007118), which encodes the transmembrane phosphatase with tensin (TPTE) (primers SUTR2 and AUTR2) (Figure 1.1A) (9). Sequence analysis of the published Sp17 cDNA B from inverted position 252-432 (180bp) revealed homology to TPTE mRNA (Figure 1.1A). Conversely, sequence homologies were not detected in cDNA B at inverted position 1-251 (Figure 1.1A). However, the published Sp17 cDNA B (180bp fragment) exhibited homology to multiple BAC clones on chromosomes 21 (Accession number AL163201), 13 (Accession number AL139386), 10 (Accession number AC027723) and 3 (Accession number AC021156).

4.2 Sp17 5'UTR sequence

5' RLM-RACE PCR was performed to analyze human Sp17 mRNA transcripts and to identify the Sp17 transcription start site (primers gene racer and ACDS or Ag1). Sequence analysis of 5' RLM-RACE amplified DNA fragments revealed a single mRNA species, which contained a contiguous Sp17 open reading frame (ORF) sequence. The sequenced fragment was identical to the published Sp17 ORF sequence (6). The first nucleotide of the isolated Sp17 cDNA sequence was determined to be just beyond the cap structure defining the transcription start site at position –72 (Figure 1.2). However, Sp17 sequence comparison of mammalian Sp17 5'UTRs found significantly longer Sp17 transcripts in non-human mammals (Figure 1.3).

Subsequently, RT-PCR analysis was performed to elucidate potential longer Sp17 transcription start sites that were not differentiated by 5' RLM-RACE (primers Sg1a, Sg1b, Sg1c, Sg1d, Sg1e and ACDS). RT-PCR amplification revealed DNA bands at 412bp, 451bp, 475bp and 503bp, respectively, suggesting the presence of a second Sp17 mRNA species with a slightly longer 5' end than the one detected by 5' RLM-RACE (Figure 1.4). However, although the exact transcription start site for this second Sp17 mRNA species was not specifically identified through sequencing, it may lie between nucleotide position -164 and -262 (Figure 1.2).

4.3 Sp17 3' UTR sequence

3' RACE PCR was performed to examine alternative polyadenylation at the Sp17 3'UTR (primers shu and anchor). RT-PCR and sequence analysis revealed three alternate 3'UTRs corresponding to the Sp17 gene with lengths of 346, 339 and 127 nucleotides (Figure 1.2). The 346bp 3'UTR sequence contained a conserved (AATAAA) polyadenylation signal located 26 nucleotides upstream from the polyadenylation site. This fragment was identical to the published Sp17 sequence (2). Similarly, the 339bp 3'UTR sequence was 7 nucleotides shorter and exhibited a partially conserved non-consensus polyadenylation signal (AATGAA), 23

nucleotides upstream from the polyadenylation site (10). The 127bp 3'UTR revealed another partially conserved polyadenylation signal (AAATAA) 23 nucleotides upstream of the polyadenylation site.

Sequence comparison of the Sp17 3'UTR in mammals, revealed a highly conserved and significantly shorter Sp17 3'UTR transcript in the sheep, rabbit, wallaby and opossum as compared to the human and baboon (Figure 1.3). Moreover, the aforementioned mammals use a polyadenylation signal (AATAAA) that appears to be mutated in the human and baboon.

4.4 Sp17 promoter sequence analysis

Sequence analysis of the genomic region upstream of Sp17 exon 1, revealed potential cis-acting elements, which may be important in the promoter activity of the Sp17 gene. The existence of two transcription start sites in the human Sp17 5' flanking sequence suggests that these transcripts may represent distinct isoforms which are initiated as a result of alternative promoters.

Sequence analysis of the published Sp17 cDNA revealed that both putative promoter regions lack canonical TATA and CAAT boxes which are characteristic of some mammalian genes (Figure 1.2). However, this region contains GC boxes, which may be recognized by other potential transcription factors. For example, two binding sites for the transcription factor Sp1 were identified close to the start site of two mRNAs (11). Similarly, several consensus sequences were identified corresponding to potential transcription factor binding sites in the 5' flanking region including the myeloid zinc finger gene 1 (MZF1); a nuclear factor of activated T cells (NFAT); the C/EBPbeta enhancer-binding protein; a transcription factor II (TCFII); an upstream stimulating factor; and activator proteins such as AP-1, AP-2 and AP-4 (Figure 1.2) (12, 13, 14, 15, 16, 17, 18, 19).

4.5 Analysis of the full length of Sp17 nucleotide sequence

A full-length Sp17 cDNA sequence was generated by combining the 5' RACE and 3' RACE sequences. The sequences were orientated by aligning the overlapping coding region sequence (239bp). The predicted initiation codon, ATG, was identified and embedded in the sequence, AGAAGATG. Although this sequence is not identical to the eukaryotic Kozak recognition sequence, CCA/GCCATG, additional upstream initiation codons were not identified (20).

The start codon is preceded by two in-frame stop codons found at nucleotide position -21 and –99. The Sp17 cDNA has an ORF that is 453bp in length and ends with a TGA stop codon. The transcript is predicted to encode a 151 amino acid protein of 17,408 Da (6). The combined Sp17 cDNA sequence from nucleotide -135 to 799 with a polyadenylation stretch of 16 nucleotides was submitted to GenBank. (Accession number AF334735) (Figure 1.2).

A complete Sp17 mammalian nucleotide sequence alignment revealed high conservation of the Sp17 sequence with respect to the human sequence. The human

Sp17 sequence exhibited homology to the macaque (98.5%), rhesus (98.1%) and baboon (97.7%), mouse (82.7%), rat (85.6%), cat (81.5%), bovine (87.3%), sheep (84.8%), marmoset (95.3%), rabbit (86.0%), opossum (78.8%) and wallaby (77.3%) (Accession numbers AF334735, AF005551, AF334809, U75209, Z46299, AJ131888, AW520875, BE106945, BF567951, AI177319, AW646882, AW463988, AF179926, AF134585, Z20655, AF054289 and AF054290) (Figure 1.3).

4.6 Characterization of the Sp17 genomic structure

The Sp17 gene structure was elucidated by searching the National Center for Biotechnology Information High Throughput Genomic Sequences (htgs) database with the human cDNA sequence. Preliminary sequences of BAC clones or cosmids spanning the Sp17 gene were analyzed and aligned to generate a contiguous gene stretch (Figure 1.5A). Sequence comparison of the genomic and Sp17 cDNA sequences revealed that the Sp17 sequence is distributed over region of ~23kb containing 5 exons which are interrupted by 4 introns. In addition, the database search also revealed clones (AC010997 and AL390034) which contain Sp17-like sequences without introns.

Analysis of the intron-containing Sp17 gene found that exon 1 represents the 5' untranslated sequence. The Sp17 exon 2 contains part of the 5'UTR and encodes the first 51 amino acid residues of the human Sp17 protein. Sp17 exons 3 and exon 4 are 71 and 87 bp in length, respectively, and contain the Sp17 coding region. Sp17 exon 5 is characterized by the TGA stop codon followed by the 3'UTR sequence including three polyadenylation signals. The Sp17 introns 2, 3 and 4 were determined to be 5,970, 10,235 and 2,520bp, respectively.

As the Sp17 intron 1 sequence was not known, PCR and sequence analysis were used to identify the sequence from the human genome (primers Sg1b, sInt1, aInt1 and Ai2). The Sp17 intron 1 fragment was determined to be 1920bp in length. The Sp17 intron 1 sequence was submitted to GenBank (Accession number AF334810).

Furthermore, characterization of the exon/intron boundary sequences demonstrated that all exon–intron splice junctions adhere to the GT-AG consensus sequence of splice donor and acceptor sites found in eukaryotic genes (21).

This genomic analysis revealed the location of the Sp17 gene sequence to be on human chromosome 11 (11q24). Interestingly, the sialic acid 9-O-acetylesterase gene was identified opposite the Sp17 gene on human chromosome 11. This finding is of particular importance because the sialic acid 9-O-acetylesterase exons 3A, 2A and 1A are embedded within Sp17 exon 2 and intron 2. Thus, the sialic acid 9-O-acetylesterase sequence may interfere with Sp17 analyses, such as northern blot (Figure 1.5).

5. Discussion

Sp17 specific primers were generated from the published Sp17 cDNA sequence to amplify the Sp17 transcript from human testis by RT-PCR. Amplification of the Sp17

sequence, spanning the 5'UTR through coding region, was not detected. However, independent amplification of the Sp17 coding region and 5'UTR were observed. These results suggest that the Sp17 5'UTR and coding region sequences are disjointed and may not be part of an expressed Sp17 mRNA.

Further analysis of the Sp17 cDNA A and cDNA B sequences suggest that the published Sp17 cDNAs may have been derived from the screening of a defective human testis cDNA library, in which several cDNA fragments may have been inserted into the phage vector at the same position. For example, the bac clones exhibiting homology to Sp17 cDNA A, are located on chromosome 14, whereas the Sp17 gene is located on human chromosome 11. Moreover, the 5'UTR of Sp17 cDNA A exhibits a polyadenylation signal followed by a polyadenylation stretch, which terminates at the beginning of the Sp17 related sequences.

Similarly, the 5'UTR of Sp17 cDNA B exhibits homology to TPTE on chromosome 3, 10, 13 and 21. However, Sp17 coding region homology was not detected to the TPTE sequence. Furthermore, a second Sp17 cDNA was not detected during library screening for the two alternative Sp17 5'UTRs. These findings strongly suggest that the published 5'UTRs for Sp17 cDNA A and cDNA B are not associated with the Sp17 transcript.

In the newly proposed Sp17 nucleotide sequence, the initiating methionine is preceded by two in-frame stop codons found at nucleotide position -21 and -99, thereby, confirming that the upstream sequence represents the true 5'UTR. In addition, two Sp17 mRNA isoforms with alternative transcriptional start sites, Sp17-1a and Sp17-1b, were detected by RT-PCR. However, analyses of expressed sequence tags (EST) suggest the Sp17 mRNA pool is dominated by a single mRNA form, Sp17-1a (22). Similarly, the alternative longer mRNA form, Sp17-1b, was also detected, but is thought to be in low abundance in testes. Thus, although the Sp17-1a and Sp17-1b transcripts can be differentially detected, the role of alternative Sp17 mRNA isoforms is not known. Moreover, whether either of these transcripts are differentially transcribed throughout spermatogenesis or have a role in non-testes tissues is not yet known.

It can be inferred by EST analyses of the Sp17 3'UTR that the AATAAA signal (346bp) may be used with the highest efficiency (12). Moreover, alternative polyadenylation was also described for the baboon and rat ESTs (23). Thus, the conservation of the extended polyadenylation signal in the baboon and human suggest that this signal underwent recent evolutionary changes rendering it an efficient polyadenylation signal (24).

Furthermore, Sp17 gene database searches revealed clones (AC010997 and AL390034) which contain Sp17-like sequences without the presence of introns. This sequence analysis suggests the presence of both intron-containing and intronless human Sp17 gene variants.

6. Conclusion

The published human Sp17 cDNA sequences were refuted. The newly proposed and complete Sp17 nucleotide sequence is characterized by a single 5'UTR containing potential alternative transcriptional start sites, Sp17-1a and Sp17-1b. In addition, the Sp17 3'UTR contains three conserved alternative polyadenylation signals. Moreover, Sp17 sequence analysis suggests the existence of an intronless Sp17 gene variant in the human genome. Furthermore, the Sp17 nucleotide sequence is highly conserved among mammals, suggesting the presence of important regulatory constraints on the evolution of Sp17.

7. Tables and Figures

Table 1.1. Outline of the RT-PCR controls used to amplify the Sp17 transcript. The reagents used and resulting template are indicated for each RT-PCR control.

	Control #			
Reagents	**1**	**2**	**3**	**4**
RNA Template	RNA	RNA	RNA	----
DNase I	DNase	----	DNase	DNase
Reverse Transcribed (RT)	RT	----	----	RT
Rnase H	Rnase	----	----	Rnase

	Control #			
Template	**1**	**2**	**3**	**4**
RNA	----	----	RNA	----
cDNA	cDNA	----	----	----
genomic DNA	----	gDNA	----	----

Table 1.2. Sp17 specific sense and anti-sense primers used in RT-PCR to elucidate the human Sp17 testis transcript and complete nucleotide sequence.

Sense (SCDS)	5’	ATTCTGAGAGAGCAACCGGAC	3’
Sense (SUTR1)	5’	GATGTGGTCAAGGAGCGAATC	3’
Sense (SUTR2)	5’	GTTGTAATGTTATCAAGTACCGAAC	3’
Sense (SUTR3)	5’	CGATGTTGTAGTGACCTTCAGT	3’
Sense (Sg1a)	5'	GAAAAAACGGTTACCCAGCAACTAG	3’
Sense (Sg1b)	5'	CGCCCCTTCTCGGCCGCCGTAG	3’
Sense (Sg1c)	5'	GCGACCTCAGGACTGGGCTGTAC	3’
Sense (Sg1d)	5'	CACCCTTTTGCAGTCCTCGCAGC	3’
Sense (Sg1e)	5'	GATCTGACCGCCGCCTAGGACTG	3’
Anti-sense (ACDS)	5’	GAGTCTAAGATGGTGACTGATG	3’
Anti-Sense (Ag1)	5'	GATAATAGTTAGAACACATAACCTC	3'
Anti-sense (AUTR1)	5’	GCAGTCACTGACACACTTGAC	3’
Anti-sense (AUTR2)	5’	ATCTCCCTCCAAGACGGATACTC	3’

Table 1.3. Sp17 specific sense and nested anti-sense primers used in 5’ RLM-RACE to elucidate the 5'UTR of the human Sp17 testis transcript.

Sense (GeneRacer)	5’	GACCACGCGTATCGATGTCGAC	3’
Anti-Sense (ACDS)	5'	GAGTCTAAGATGGTGACTGATG	3'
Anti-Sense (Ag1)	5'	GATAATAGTTAGAACACATAACCTC	3'

Table 1.4. Sp17 specific sense and nested anti-sense primers used in 3’ RACE PCR to elucidate the 3'UTR of the human Sp17 testis transcript.

Oligio (dT) primer	5’	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTV	3'
Sense (Shu)	5’	CGCGGATCCATGTCGATTCCATTCTCC	3’
Anti-sense (anchor)	5’	GACCACGCGTATCGATGTCGAC	3’

Table 1.5. Sp17 specific sense and anti-sense primers used in PCR to amplify the Sp17 intron 1 region from human genomic DNA.

Sense (Sg1b)	5'	CGCCCCTTCTCGGCCGCCGTAG	3'
Sense (sInt1)	5’	TCAGTAAGTAAGCTCAGCCCGC	3’
Anti-Sense (Ai2)	5’	TCATCTCAGGAATAACAGTCAGCTG	3’
Anti-Sense (aInt1)	5’	TCAGTAAGTAAGCTCAGCCCGC	3’

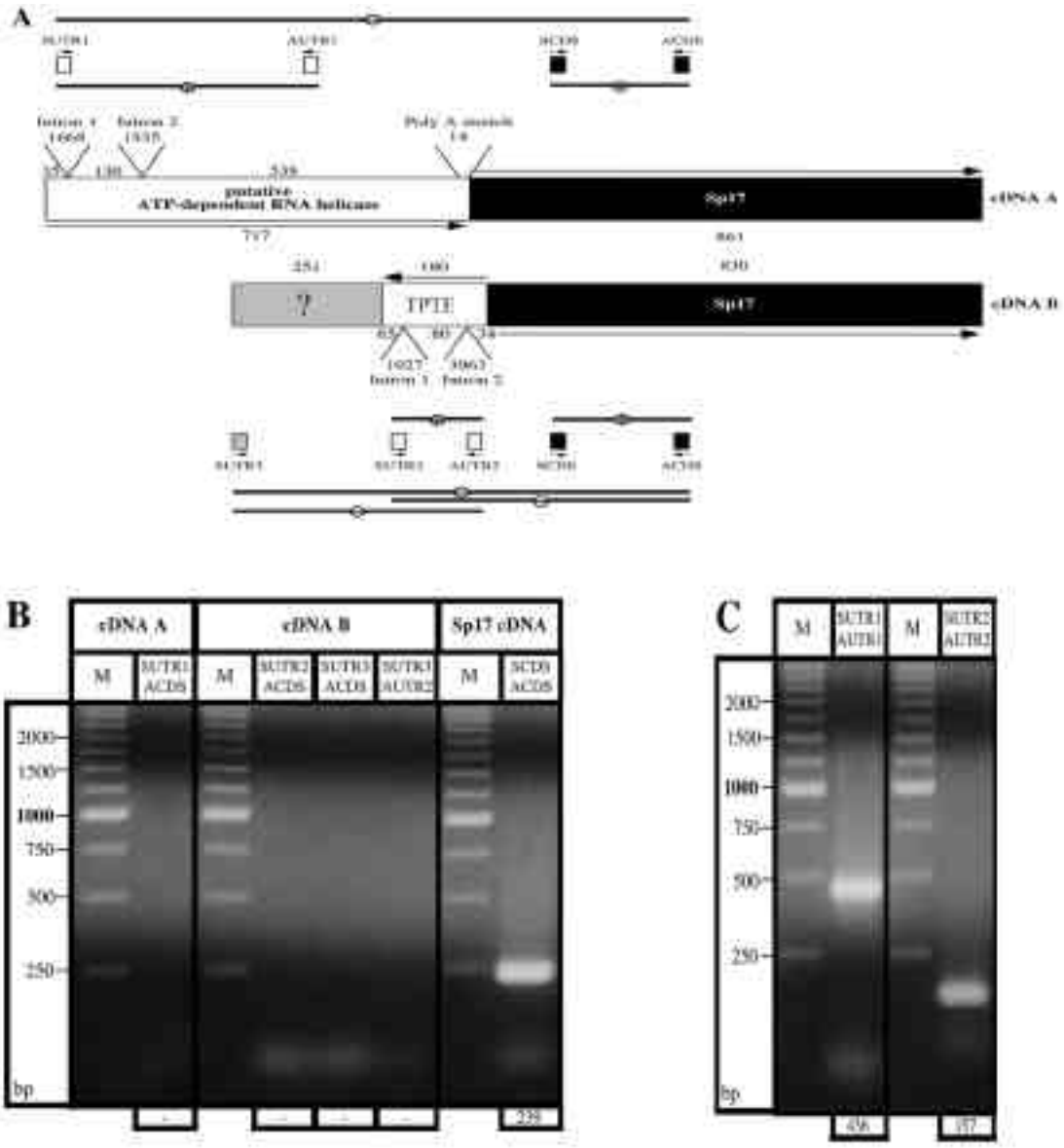


Figure 1.1. Graphic diagram refuting the published human Sp17 spliced cDNA model. **A.** The published alternative Sp17 cDNAs, cDNA A and cDNA B, are shown with characteristic features noted. An arrow indicates the sequence orientation with sequence length in bp. Sp17 specific primers and potential PCR product lengths are presented above and below the cDNAs. **B and C.** RT-PCR analysis of the published Sp17 nucleotide sequence. PCR amplification of the continuous Sp17 5'UTR through coding region sequence was not detected (B). However, independent amplification of the Sp17 coding region and the 5'UTR cDNA A region was detected (C). These results suggest that the Sp17 5'UTR and coding region are disjointed. Primer names, combinations and PCR products (as measured (bp) by a DNA marker (M)) are indicated.

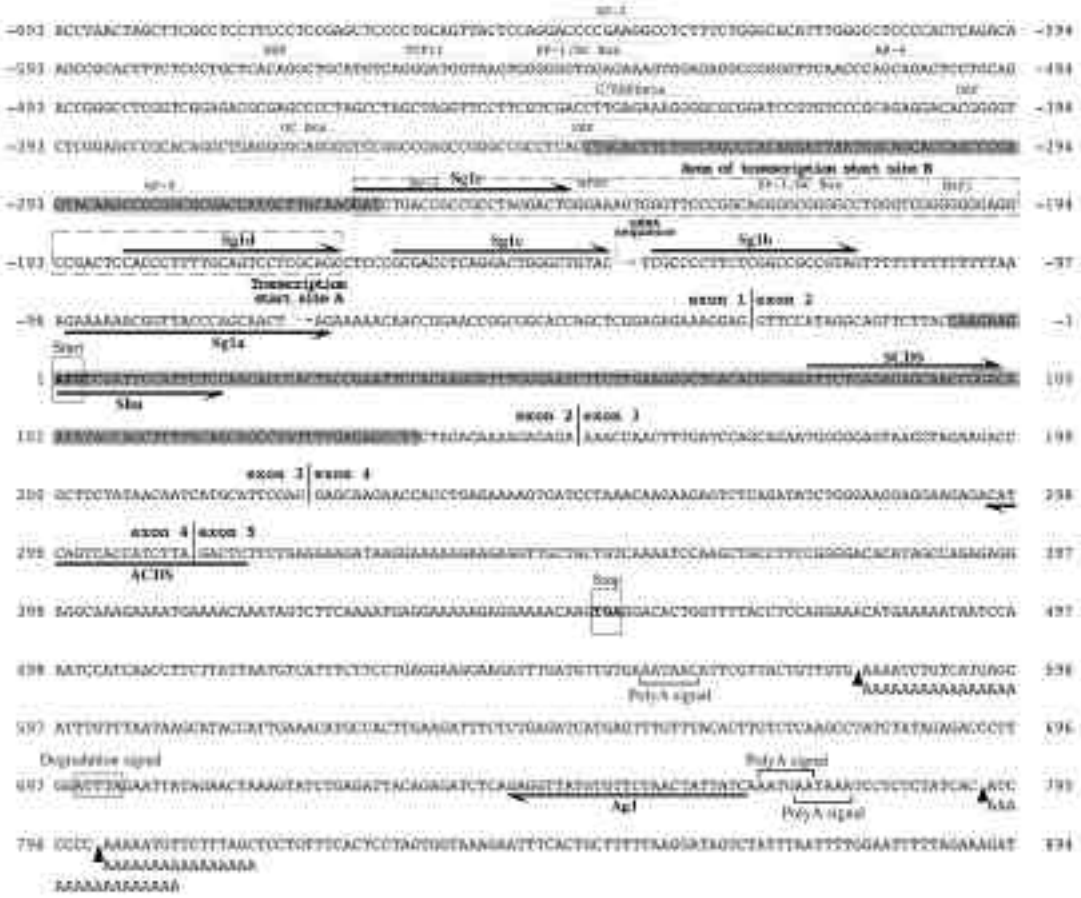
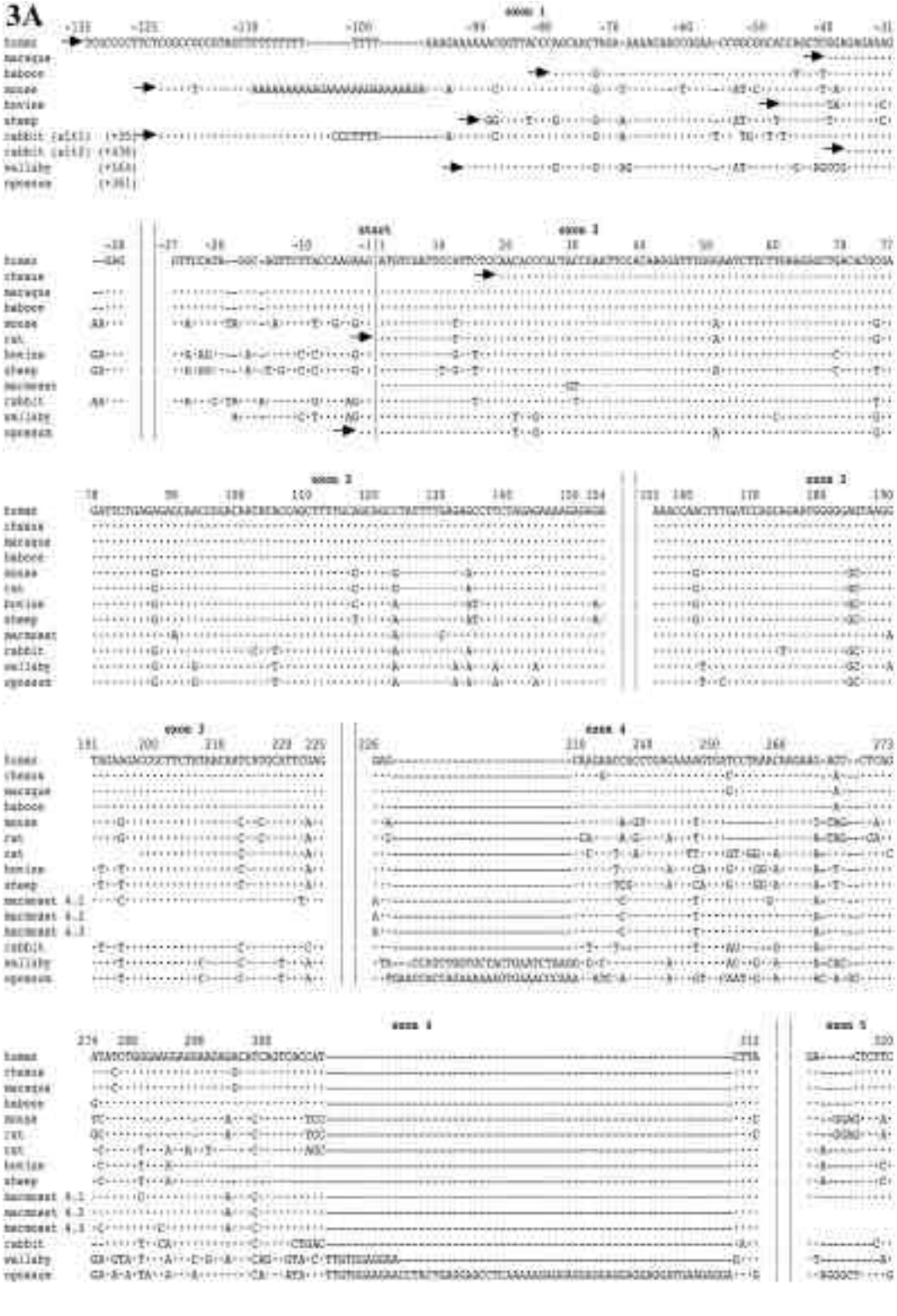


Figure 1.2. The complete human Sp17-1 nucleotide sequence (Accession number AF334810). The human Sp17-1 nucleotide sequence was generated by combining 5' RLM-RACE and 3' RACE sequences. The start (ATG, position 1), stop (TGA) codons and transcriptional start sites (position -72 and position -186 to -136), are indicated by boxes. Exon and intron boundaries, alternative poly(A) tracts, polyadenylation signals, degradation signals and common regulatory signals are indicated. Primer names and primer sequence directions are indicated by arrows. The shaded nucleotides show similarity to the sialic acid 9-O-acetylesterase exons on the opposite chromosome strand.

3A


exon 1
start
exon 2
exon 3
exon 4
exon 5
human
chanas
macaque
baboon
mouse
rat
bovine
sheep
rabbit (aLG1)
rabbit (aLG2)
wallaby
opossum
macaque 4.1
macaque 4.2
macaque 4.3

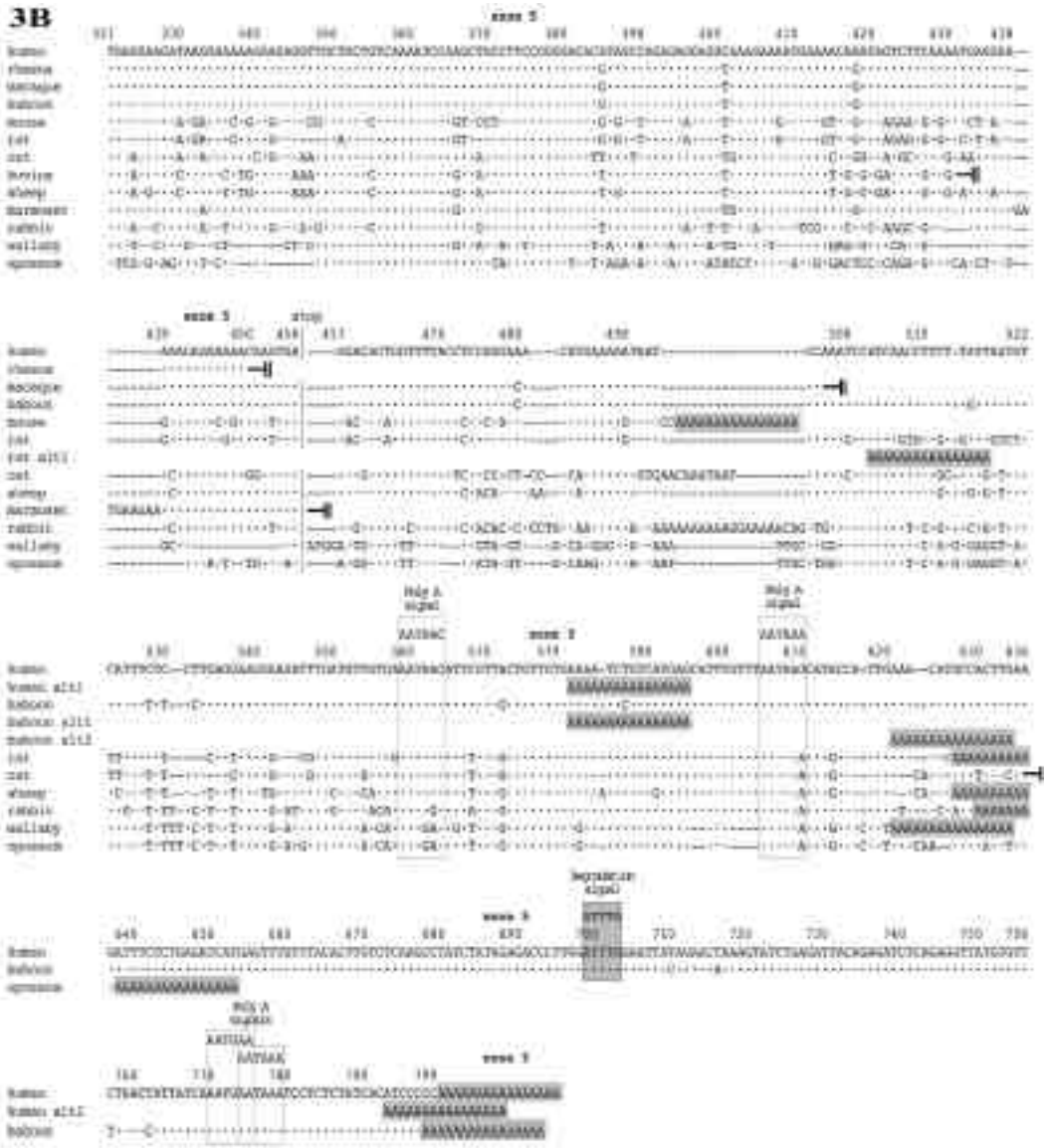


Figure 1.3. Alignment of the known mammalian Sp17 nucleotide sequences. The mammalian Sp17 cDNA sequences including human, rhesus macaque, cynomologous macaque, baboon, mouse, rat, cat, bovine, sheep, marmoset, rabbit, wallaby and opossum were aligned by Clustal algorithm with respect to the human Sp17 nucleotide sequence. Sequence homology is represented by dots. Dashes were used to maintain alignment. The sequence start (arrows), sequence end (double-bars), poly A stretch (shaded), start and stop codons (vertical line), exons (double vertical line), poly A signals (box), degradation signals (shaded box), and alternative polyadenylation (alt1 or alt2) are indicated. The rabbit, wallaby and opossum 5'UTR did not exhibit Sp17 homology and were omitted. In addition, the marmoset multiple exon 4s are numbered 4.1 to 4.3.

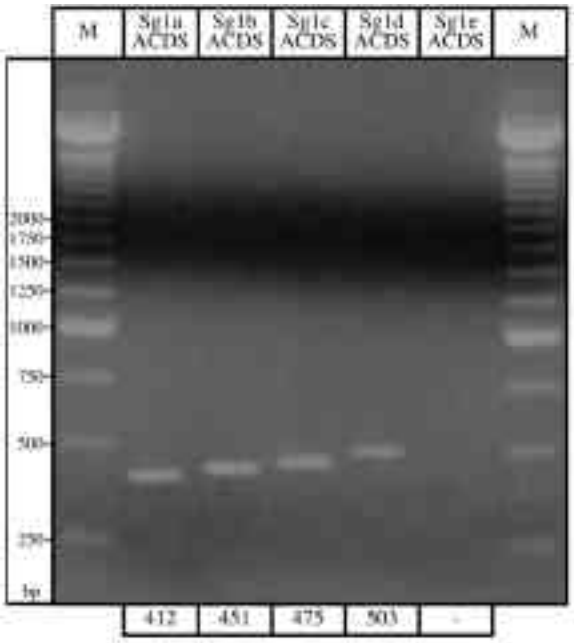


Figure 1.4. RT-PCR analysis of Sp17-1 alternative transcriptional start sites. RT-PCR was used to elucidate a more upstream Sp17 transcription start site not detected by 5' RLM-RACE PCR. RT-PCR analysis was performed using Sp17 sense primers with increasing distances from the first transcription start site (Figure 1.2). PCR products were detected for Sg1a, Sg1b and Sg1c. However, DNA products were not detected for Sg1d and Sg1e, suggesting the position of the second transcription start site between Sg1c and Sg1d. Sp17 specific primer sets are indicated above the running lane. PCR product sizes (bp) are shown below the running lane. The DNA marker (M) is shown at the left.

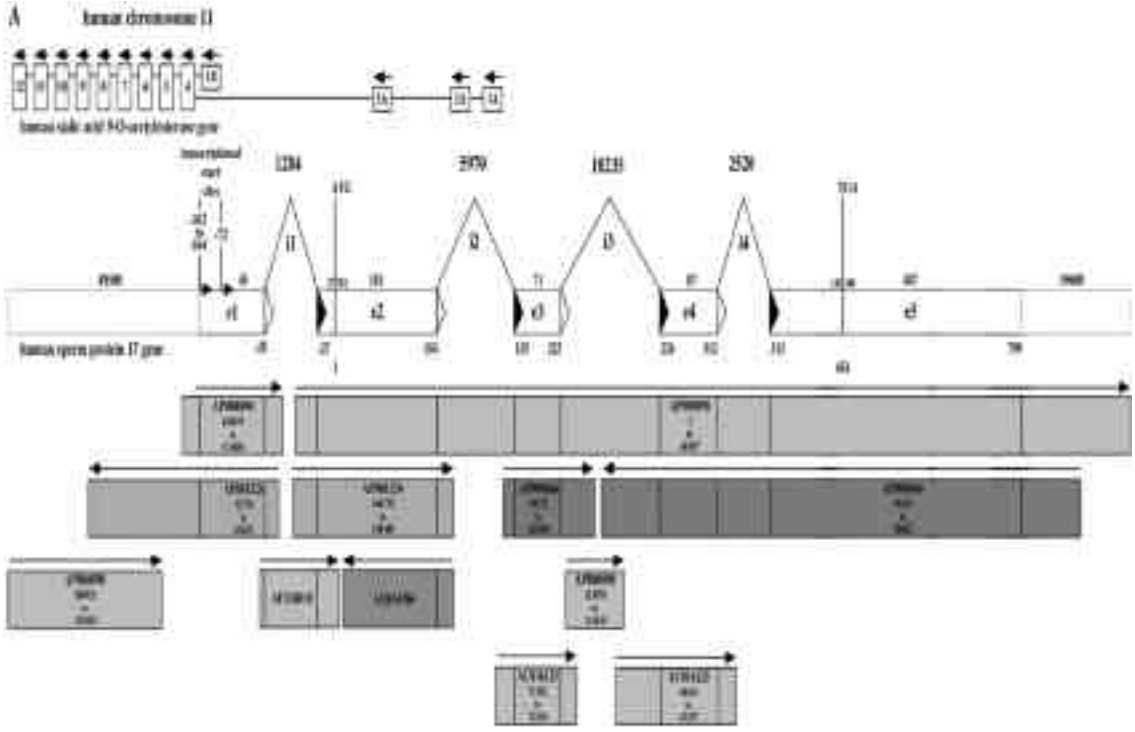


Figure 1.5. A schematic diagram of the genomic organization of the human Sp17 gene. The five exons (e1-e5) (open boxes), 4 introns (i1-14) (peaks) and start (ATG; position 1) and stop (TGA) codons (vertical bars) are indicated. The length of each region is expressed in nucleotides (nt) above boxes and peaks. Identified BAC clones or cosmids (boxes) covering the Sp17 genes are presented with the Accession number and sequence position of the clones. Sequence orientation is indicated by arrows. In addition, the sialic acid 9-O-acetylesterase gene on the opposite chromosome strand indicated by rectangles which correspond to identified exons. The sequence of Sp17 intron 1 (primers Sg1b/Ai2) is available under accession number AF334810.

8. References

1 O'Rand, M. G. and E. E. Widgren. 1994. Identification of sperm antigen targets for immunocontraception: B-cell epitope analysis of Sp17. Reprod. Fertil. Dev. 6:289-296.

2 Lea, I. A., R. T. Richardson, E. E. Widgren and M. G. O'Rand. 1996. Cloning and sequencing of cDNAs encoding the human sperm protein, Sp17. Biochim. Biophys. Acta. 1307:263-266.

3 de la Cruz, J., D. Kressler and P. Linder. 1999. Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem. Sci. 24:192-198.

4 Richardson RT, N. Yamasaki and M. G. O'Rand. 1994. Sequence of a rabbit sperm zona pellucida binding protein and localization during the acrosome reaction. Dev Biol. 165:688-701.

5 Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

6 Heinemeyer, T., E. Wingender, I. Reuter, H. Hermjakob, A. E. Kel, O. V. Kel, E. V. Ignatieva, E. A. Ananko, O. A. Podkolodnaya, F. A. Kolpakov, N. L. Podkolodny and N. A. Kolchanov. 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res. 26:362-367.

7 Quandt, K., K. Frech, H. Karas, E. Wingender and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23:4878-4884.

8 Zhao, Y., L. Yu, Q. Fu, W. Chen, J. Jiang, J. Gao and S. Zhao. 2000. Cloning and characterization of human DDX24 and mouse Ddx24, two novel putative DEAD-Box proteins, and mapping DDX24 to human chromosome 14q32. Genomics 67:351-355.

9 Chen, H., C. Rossier, M. A. Morris, H. S. Scott, A. Gos, A. Bairoch and S. E. Antonarakis. 1999. A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. Hum. Genet. 105:399-409.

10 Sheets, M. D., S. C. Ogg and M. P. Wickens. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. 18:5799-5805.

11 Thiesen, H.J. and C. Bach. 1990. Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res. 18:3203-3209.

12 Morris, J. F., R. Hromas and F. J. Rauscher. 1994. Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. Mol. Cell. Biol. 14:1786-1795.

13 Rao, A., C. Luo and P. G. Hogan. 1997. Transcription factors of the NFAT family: regulation and function. Annu. Rev. Immunol. 15:707-747.

14 Muller, C., M. Alunni-Fabbroni, E. Kowenz-Leutz, X. Mo, M. Tommasino and A. Leutz. 1999. Separation of C/EBPalpha-mediated proliferation arrest and differentiation pathways. Proc. Natl. Acad. Sci. USA. 96:7276-7281.

15 Johnsen, O., P. Murphy, H. Prydz and A. B. Kolsto. 1998. Interaction of the CNC-bZIP factor TCF11/LCR-F1/Nrf1 with MafG: binding-site selection and regulation of transcription. Nucleic Acids Res. 26:512-520.

16 Qyang, Y., X. Luo, T. Lu, P. M. Ismail, D. Krylov, C. Vinson. and M. Sawadogo. 1999. Cell-type-dependent activity of the ubiquitous transcription factor USF in cellular proliferation and transcriptional activation. Mol. Cell. Biol. 19:1508-1517.

17 Foletta, V. C.. 1996. Transcription factor AP-1, and the role of Fra-2. Immunol. Cell. Biol. 74:121-133.

18 Garcia, M. A., M. Campillos, S. Ogueta, F. Valdivieso and J. Vazquez. 2000. Identification of amino acid residues of transcription factor AP-2 involved in DNA binding. J. Mol. Biol. 301:807-816.

19 Hu, Y. F., B. Luscher, A. Admon, N. Mermod and R. Tjian. 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. 4:1741-1752.

20 Kozak, M.. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12:857-872.

21 Shapiro, M. B. and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15:7155-7174.

22 Boguski, M. S., T. M. Lowe and C. M. Tolstoshev. 1993. dbEST--database for "expressed sequence tags" [letter]. Nat. Genet. 4: 332-333.

23 Adoyo, P.A., I. A. Lea, R. T. Richardson, E. E. Widgren and M. G. O'Rand. 1997. Sequence and characterization of the sperm protein Sp17 from the baboon. Mol. Reprod. Dev. 47:66-71.

24 Alberta, J. A., K. Rundell and C. D. Stiles. 1994. Identification of an activity that interacts with the 3'-untranslated region of c-myc mRNA and the role of its target sequence in mediating rapid mRNA degradation. J. Biol. Chem. 269:4532-4538.