Cloning and Characterization of Testis-Specific Spermatogenesis Associated Gene Homologous to Human SPATA4 in Rat

Shang-Feng LIU, a,b Shan HE,a Bo-Wen LIU,a Ying ZHAO, and Zhao WANG,a,b

a Department of Biological Sciences and Biotechnology, Tsinghua University; b Institute of Biomedicine, Tsinghua University; Beijing 100084, People’s Republic of China. Received August 11, 2004; accepted August 25, 2004

Rat SPATA4 gene, homologue to the human and mouse SPATA4 gene, expressed specifically in the rat testis was cloned by informatics analysis. The cDNA mapped to chromosome 16 in the rat genome is made up of 6 exons and the exon-intron boundaries obey to the AG/GT rule. The gene contains a 972 bp open reading frame encoding 323 amino acid sequences with theoretical molecular weight of 36.64 KD and isoelectric point of 9.65. One CpG island is located in the gene from site –200 to +198. A typical promoter is also predicted from site –630 to +101. According to the computer-aided analysis of the putative protein encoded by the rat SPATA4, no transmembrane region and no signal peptides are found in the protein. Multi-tissue RT-PCR results show that the SPATA4 gene is expressed specifically in the testis only. Moreover, the expression of SPATA4 occurs in a development stage-dependent pattern. According to the RT-PCR results, no expression of SPATA4 is detected until the rat is 30 d old after birth. The amount of SPATA4 mRNA increases from 30-d to 65-d-old rat and then keeps stable after that. In conclusion, this study proves the conservation of SPATA4 in mammalian animals and predicts its important role in spermatogenesis.

Key words SPATA4; testis-specific expression; rat; bioinformatics; spermatogenesis

Apoptosis, which happens selectively to certain spermatogenic cells, keeps the normal spermatogenesis. However, an unbalanced apoptosis may lead to the spermatogenic dysfunction such as infertility. Thus, a profound understanding of the apoptosis mechanism may uncover the causes for many testicular failures and help to find out efficient strategies against these defects.1) So far many testis spermatogenic cell apoptosis related genes were reported to be involved in this process including Mcl-1,2) p53,3) CREM,4) Fas,5) Fas,5) Fas,6) TRAIL,8) c-myc,9) and TR2.10)

However, all of the genes presented above are not testis-specific spermatogenic cell apoptosis-related gene. In previous study, Jiang11) cloned 24 ESTs of mouse testis spermatogenic cell apoptosis-related gene by creating mouse cryptorchidism model and making use of suppression subtractive hybridization, and registered in Genbank. Beginning with the EST BE644542, one of the ESTs mentioned above, a novel mouse gene full-length cDNA sequence-mouse SPATA4 (former name is SRG2, GenBank Accession No. AF395083) was cloned from a mouse testis cDNA library and primary functional study was performed successfully on this gene.12,13)

Based on the mouse SPATA4 sequence, the human testis-specific spermatogenic cell apoptosis related gene homologous to mouse SPATA4 (former name is TSARG2, GenBank Accession No. AY040204) was cloned and identified as candidate oncogene.14,15)

Though many discoveries have been made in exploring the mechanism about the spermatogenic cells apoptosis, many details remain mysteries. Rat, an experimental animal model, can provide much convenience for our research on the mechanism of apoptosis of spermatogenic cells. Thus, in this study, we cloned the SPATA4 homologue in the rat. Besides, we examined the characteristics of the rat SPATA4 gene and its developmental expression pattern.

MATERIALS AND METHODS

Animals Sprague–Dawley rats were purchased from the Laboratory Animal Center of Beijing University (Beijing, China) and maintained in a temperature- and humidity-controlled room on a 12 : 12 h (L: D, light versus dark) photoperiod. The rats had free access to food and water. All measures taken for the rat referred to “Guideline for the Care and Use of Laboratory Animals” established by the Chinese Council on Animal Care.

Silicon Cloning of Rat SPATA4 Based on the human and mouse SPATA4 sequence we used Blast algorithms through the National Center of Biotechnology Information and Ensembl Genome Browser (http://wwwensembl.org/) found rat genomic SPATA4 sequence. Genscan program (http://genes.mit.edu/GENSCAN.html) was used to predict the full-length cDNA of rat SPATA4.

RNA Extraction Total cellular RNA was isolated from brain, heart, pancreas, small intestine, lung, liver, kidney, eye, ovary and testis and was also extracted from rat embryo, 14-d-, 30-d-, 40-d-, 45-d-, 50-d-, 65-d- and 2-year-old rats. RNA was isolated using the RNA isolation kit (cell and tissue kit, Gentra, U.S.A.) according to the protocol.

Identification of SPATA4 Full Length Sequence by RT-PCR First-strand cDNA was synthesized from total RNAs of testis. 2 µg total RNAs were reversely transcribed to synthesize first-strand cDNA using the MMLV First Strand cDNA Synthesis Kit (Sangon, Shanghai) according to the protocol. 1 µl of product was used as template for the subsequent PCR amplification with gene specific primers. For rat SPATA4 mRNA identification, the primers used were: RAT1, 5’-CGG AAT TCT CAT GGC TGC CGC CGG CCG GG-3’ and RAT2, 5’-CGG TCG ACG TGT CAT TAT ATA CCC AGA TTG TGG-3’. Following an initial denaturation step at 95°C for 1 min 30 s, the reaction was subjected to 35 cycles of amplification at 94°C for 40 s, 58°C for 30 s, 72°C for 40 s, and a final extension for 5 min. 2 µl of each reaction product was size-fractioned by 2% (w/v) agarose gel electrophoresis and stained with ethidium bromide and photographed using Imagemaster VDS (Amersham Pharmacia biotech, U.S.A.). The reactions were carried out in a total...
The RT-PCR for identification of the rat condition and reaction mixture components as those used in traction part. The RT-PCR was performed using the same tissues and testis at different stage listed in the RNA Extraction. Putative proteins were also obtained from the Compute pI/Mw Program at the ExPASy molecular biology WWW server of the Swiss Institute of Bioinformatics (http://cn.expasy.org/tools/protscale.html). Motifs of the putative protein were predicted by the Prosite Program at the ExPASy molecular biology WWW server (http://ex.expasy.org/prosite/). Molecular weight and isoelectric point predictions and TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html) for transmembrane domain prediction. The hydrophobicity of the amino acid sequences was analyzed by the ProtScale at the ExPASy molecular biology WWW server (http://cn.expasy.org/tools/protscal.html). Motifs of the putative protein were predicted by the Prosite Program at the ExPASy molecular biology WWW server (http://ex.expasy.org/prosite/). Molecular weight and isoelectric point were also obtained from the Compute pI/Mw Program at the ExPASy molecular biology WWW server (http://cn.expasy.org/tools/pi_tool.html). Finally, we used the Conserved Domain Database at the NCBI WWW server (http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/BLAST/cdd_search.html) to search for any conserved important domains in the putative proteins.

**Bioinformatics Analysis**

The CpG island was predicted using the CpG Island Finder program (http://cctn.hsc.usc.edu/cpgislands) and the promoter location was obtained using the Gene2promoter program (http://www.genomatix.de/cgi-bin/eldorado/main.pl). ClustalW program was used for multiple DNA sequence alignment (http://www.ebi.ac.uk/clustalw/).

The amino acid sequence was analyzed using various bioinformatics tools as follows: SignalP for signal peptide predictions (10) and TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html) for transmembrane domain prediction. The hydrophobicity of the amino acid sequences was analyzed by the ProtScale at the ExPASy molecular biology WWW server of the Swiss Institute of Bioinformatics (http://cn.expasy.org/tools/protscal.html). Motifs of the putative protein were predicted by the Prosite Program at the ExPASy molecular biology WWW server (http://ex.expasy.org/prosite/). Molecular weight and isoelectric point were also obtained from the Compute pI/Mw Program at the ExPASy molecular biology WWW server (http://cn.expasy.org/tools/pi_tool.html). Finally, we used the Conserved Domain Database at the NCBI WWW server (http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/BLAST/cdd_search.html) to search for any conserved important domains in the putative proteins.

**Spatio-temporal Expression Pattern Detection by RT-PCR**

RNA used in the detection was extracted from various tissues and testis at different stage listed in the RNA Extraction part. The RT-PCR was performed using the same condition and reaction mixture components as those used in the RT-PCR for identification of the rat SPATA4 gene. β-actin was used as internal control. The primers for its detection were: forward primer, 5'-GGT ACC GCA GCG GCC GCC CCG GGT T -3'; and reverse primer, 5'-GGC TCG ACA GCC GAG GAG G -3'. The final product was detected in the same way mentioned in the Identification of SPATA4 full length sequence by RT-PCR part. Besides, equal volumes of the PCR products were analyzed by densitometry with Gel-Pro Analyzer version 3.1.

**RESULTS AND DISCUSSION**

The rat SPATA4 cDNA we cloned in silicon was submitted to the Genebank and the Genebank accession number is AY653229. The complete cDNA sequence of rat SPATA4 contains 1086 bp, with a 972 bp open reading frame which encodes a polypeptide of 323 amino acid residues with molecular weight of 36.6 kDa and pI 9.65 (Fig. 1). To prove the validity of the SPATA4 gene sequence cloned in silicon by us, we performed RT-PCR using the RNA extracted from adult rat testis. The specific primers were designed according to the sequences we submitted to the NCBI to cover the open reading frame. The sequences of expected fragment amplified during RT-PCR were identical to what we submitted to the Genebank (data not shown). Thus, the sequences we cloned using the bioinformatics methods really exist in the rat genome and are valid.

To get more information about the characteristics about the gene, we analyze the DNA sequences using different bioinformatics programs. With the CpG Island Finder, a 398 bp CpG island with 55.5% GC value from -200 to +198 was predicted in the rat SPATA4 gene (Fig. 2). Besides, a typical 731 bp promoter was located from -630 to +101 in the CpG island. When comparing the rat SPATA4 cDNA sequence with the genome sequence, the exon/intron organization was determined. The rat SPATA4 gene is split into 6 exons (Fig. 3a) and the exon/intron junction obeys to the AG/GT rule (Fig. 3b). We searched for rat SPATA4 in the rat genome (http://www.ncbi.nlm.nih.gov-genome/seq/RnBlast.html) and the rat SPATA4 gene was mapped to the rat chromosome 16.

To predict the characteristics, we also analyze the putative protein using different bioinformatics method. The amino acid sequence analysis of the encoded protein using the SignalP (V 3.0) predicts no signal peptide. No transmembrane domain is predicted from ProtScale at the ExPASy molecular biology WWW server, the protein has two hydrophobic regions. Amino Acid Sequences

**Polyadenylation signal is underlined. Stop codon is indicated with an asterisk (+).**

Fig. 1. Nucleotide Sequence of Rat SPATA4 cDNA and the Deduced Amino Acid Sequences

```plaintext
ATG GGT CCA GAC GAC AAT GGC GGC GGC GGC GGC
```

Vol. 27, No. 11

---

**REFERENCES**

309—311), two N-myristoylation sites (60—65, 131—136), three N-glycosylation sites (69—72, 99—102, 165—168), six Casein kinase II phosphorylation sites (95—98, 152—155, 158—161, 243—246, 277—280, 309—312) and one cAMP- and cGMP-dependent protein kinase phosphorylation site (220—223) are found in the putative protein using the Prosite Program.

To examine the spatial expression pattern of rat SPATA4, we performed RT-PCR, using the total RNA extracted from different tissues including brain, heart, pancreas, intestine, lung, liver, kidney, eye, ovary and testis. The RT-PCR results show that the SPATA4 transcript of about 1.0 kb was detected exclusively in the testis (Fig. 4). To investigate developmental change of the SPATA4 gene expression in the testis, total testicular RNAs were isolated from the prenatal embryo and at the age of 14, 30, 40, 45, 50, 65 d and two years. The mRNA transcript was detected first at the age of 30 d. Besides, RT-PCR results showed that mRNA transcription increased gradually from the age of 30 to 65 d and the transcription keeps stable from the age of 65 d (Fig. 5).

Comparing the rat SPATA4 with the other two homologous genes in human and mouse, high degree similarity can be found among these. The three homologous genes show very similar genomic organizations. They all have 6 exons and even the sizes of exon 2, 3, and 5 are identical. The results of the sequence alignment between the 6 exons of the three ho-
mologues are presented below. For exon 1 of SPATA4, the rat/human and rat/mouse similarity are 76% and 91%. For exon 2, the rat/human and rat/mouse similarity are 88% and 95% respectively. For exon 3, the rat/human and rat/mouse similarity are 76% and 91%. For exon 6 of rat SPATA4, the rat/human and rat/mouse similarity are 74% and 83%. But exon 6 of rat SPATA4 has 49% and 87% similarity to the exon 6 of the homologous gene in human and mouse. For exon 5, the rat/human and rat/mouse similarity are 74% and 83%. But exon 6 of rat SPATA4 has 85% and 93% similarity to that of the homologous gene in human and mouse. For exon 5, the rat/human and rat/mouse similarity are 74% and 83%. But exon 6 of rat SPATA4 has 85% and 93% similarity to that of the homologous gene in human and mouse. It is rather possible that these exon 1, 2, 3, 4, and 5 especially exon 2 and 3 which have high similarity among the three homologues in human, mouse and rat may encode key domains of the proteins. Besides, all three genes have a CpG island respectively. Considering the expression pattern, all of the three only have expression in testis and expression of the three genes after the spermatogenesis happens.\textsuperscript{15,15} From all the data shown above, evolutionary conservation of SPATA4 in human, mouse and rat can be detected and this conservation may suggest a functional importance in the mammalian spermatogenesis.

The computer-aided analysis shows that no signal peptides and no transmembrane regions are found. Rat SPATA4 encodes a protein containing only 323 amino acid residues, a small size for a protein. Comparing the amino acid sequences of the putative protein with the conserved domain database at the NCBI WWW server, we found no homologous significant domains known.

The results of RT-PCR expression show that the rat SPATA4 is only expressed in the testis and its expression in the testis can only be detected in the rats elder than 30-d-old. Moreover, increase of expression can be found from the age of 30 to 65 d and the increase stops in the rats elder than 65 d. Gametogenesis, a cyclic process which generates spermatocytes, consists of two different steps: spermatogenesis, in which spermatogonia undergo two consecutive meiotic divisions and generates haploid cells and spermogenesis in which haploid cells are committed to drastic remodeling to become spermatooza.\textsuperscript{13} For normal development procedures of male rats, the testis begins to fall 23—25 d after birth; then during the age from 30 to 35 d, the testis comes into the scrotum. When being 45 to 60-d-old, the rats can generate the spermatozoa. Comparing the stage of the rat SPATA4 expression with the development process, we propose that the rat SPATA4 functions from the spermatogonia stage because the expression begins to happen from 30 d after birth, a time just in the process of spermatogonia.

A profound mechanism of the apoptosis of spermatogenic cells is essential for getting more information for the strategies against the diseases that involve the spermatogenic apoptosis. The strong similarity between the three homologous genes may form the bases for building a knock-out model to study the SPATA4's function in human.

**Acknowledgements** This work was supported by research grants from the Tsinghua-Yue-Yuen Medical Sciences Fund (Grant THYY20040008) and the China Postdoctoral Science Foundation (No. 2004035052).

**REFERENCE**