Developmental Expression Pattern

Zebrafish spata2 is expressed at early developmental stages

ENRICO MORO1,#, CLAUDIO MARAN3,#, M. LILIANA SLONGO3, FRANCESCO ARGENTON1, STEFANO TOPPO2 and MAURIZIO ONISTO3,*

1Dept. of Biology, 2Dept. of Biological Chemistry and 3Dept. of Experimental Biomedical Sciences, University of Padua, Padova, Italy

ABSTRACT Spata2 (spermatogenesis-associated protein 2) was originally described as a novel gene involved in the spermatogenic process. In this study, we cloned a potential zebrafish spata2 orthologue. The consensus open reading frame (1650 bp) encodes a polypeptide of 550 amino acids which shares 37% identity with the human SPATA2. Bioinformatic analysis reveals a small pattern PW [KR] KE [YF][RK] which seems to be of particular interest in the light of its strong conservation between SPATA2 and the recently discovered TAMO protein of D. melanogaster. RT-PCR analysis in adult zebrafish tissues revealed that spata2 mRNA has a broad distribution. Whole-mount in situ hybridization demonstrated that spata2 transcripts are maternally derived and becomes strongly localized in the central nervous system at early developmental stages. From 5 dpf, spata2 expression becomes detectable in the gut and pronephric duct epithelium, suggesting a wide tissue function during vertebrate development.

KEY WORDS: spata2, zebrafish (Danio rerio), whole-mount in situ hybridization

pd1 was initially identified and cloned as a novel human gene from testis cDNA library (Graziotto et al., 1999). The rat orthologue of human pd1 was further cloned and, according to the Gene Nomenclature Committee, was renamed spata2 (spermatogenesis-associated protein 2) gene on the basis of expression analysis performed on rat testis at different testicular developmental stages (Onisto et al., 2000). The analysis of the human and rat cDNA sequences disclosed an open reading frame for a protein of 520 and 511 amino acids respectively, with an overall identity of 85%. High expression spata2 levels have been detected in the brain, both in human and rodents, whereas less abundant transcripts were identified by virtue of RT-PCR experiments in skeletal muscle and kidneys (Onisto et al., 2001). Using a bioinformatic approach, it has been predicted that SPATA2 protein is highly hydrophilic, contains a cluster of 17 cysteins at its C-terminus and shares a significant similarity with the recently characterized Drosophila melanogaster TAMO protein at the N-terminus (Minakhina et al., 2003).

The promoter region and the structural organization of the human spata2 gene were also determined showing that this gene is composed of three exons and two introns spanning a region of about 12.1 kb localized at 20q13.13 (Slongo et al., 2003). The transient expression of chimeric SPATA2-GFP in HL-TAT cells (Onisto et al., 2001) revealed that spata2 has a nuclear localization, but a clear function at the cellular level has not elucidated so far.

Having given the major technical drawbacks in using rodents as animal models to address our investigation of SPATA2 function, we decided to identify a potential orthologue in the teleost fish Danio rerio. This animal has become widely used as a genetic tool to uncover specific functions of unknown proteins (Dooley and Zon, 2000; Rubinstein, 2003). Its transparency at early developmental stages together with the easy manipulation and high reproductive capability makes it an ideal animal system for molecular studies. Therefore, we carried out a systematic search of potential spata2 gene in public zebrafish databases.

Here we report the cloning and characterization of a 2905 bp cDNA sequence which represents a candidate zebrafish orthologue of the human spata2 gene. By means of whole-mount in situ analysis we further show its spatiotemporal expression pattern.

Abbreviations used in this paper: aa, aminoacid; BCIP, bromo-4-chloro-3-indolyl phosphate; bp, base pair; CNS, central nervous system; cDNA, complementary DNA; Dig, digoxigenin; dpf, days post fertilization; EDTA, ethylenediaminetetra-acetic acid; hpf, hours post fertilization; kb, kilobase; NBT, nitro blue tetrazolium; nt, nucleotide; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcriptase-polymerase chain reaction; spata, spermatogenesis-associated protein; TAE, tris-acetic acid-EDTA; UTR, untranslated region.

*Address correspondence to: Dr. Maurizio Onisto. Dept. of Experimental Biomedical Sciences, University of Padua, Viale G. Colombo 3, 35121 Padova, Italy. Fax: 0039-049-827-6097. e-mail: maurizio.onisto@unipd.it

# Note: These authors contributed equally to this paper.

0214-6282/2007/$30.00
© UBC Press
Printed in Spain
www.intjdevbiol.com
Since the deposited sequence in Ensembl Database was not retrieved the relative nucleotide sequence (Ensembl Transcript sequence and to obtain the full-length cDNA. cDNA was generated by completed, we sought to confirm the Sequence analysis revealed an ORF of 1650 bp encoding a protein of 550 amino acids. The length of 5' and 3' UTRs is 176 bp and 1076 bp respectively.

Isolation and characterization of zebrafish spata2 gene.

Starting from the amino acid sequence of human SPATA2 we and 1076 bp respectively. protein of 550 amino acids. The length of 5' and 3' UTRs is 176 bp and 1076 bp respectively.

The translation start codon was assigned to the first in frame ATG that appeared downstream at a stop codon TGA at bases 96 to 98. Furthermore, this first ATG position is in agreement with (Minakhina et al., 2003), as will be discussed further. The remaining portion of the protein has a central region from aa 480 to aa 550 it is predicted to be a multidomain protein. Despite the fact that no experimental evidence supports this hypothesis, some details should be pointed out. The first 220 amino acids show an ordered secondary structure topology that is mainly alpha helix and it is the only portion of the protein that shares an evident similarity to TAMO protein

Kozak’s consensus sequence. Compared to the Ensembl transcript ID:ENSDART00000034839, our spata2 sequence contains a 162 bp longer region in the coding sequence. We found that zebrafish SPATA2 protein shares 37% and 36% identity in amino acid sequence with H.sapiens (GenBank Accession No. U28164) and R.norvegicus (GenBank Accession No.AF123651) SPATA2 respectively.

Bioinformatic and amino acid sequence analysis of zebrafish SPATA2.

SPATA2 proteins of D. rerio and T. rubripes do not differ to the general consensus of other proteins belonging to this family (Fig. 2). The N-terminus is slightly longer accounting for a predicted longer alpha helix, whereas the region from 220 to 480 residues varies in length and composition, showing a more manifest divergence and, overall less conservation. Conservation is again seen in the final C-terminal residues that are well aligned. Several in silico analyses suggest that SPATA2 may be a multidomain protein. Despite the fact that no experimental evidence supports this hypothesis, some details should be pointed out. The first 220 amino acids show an ordered secondary structure topology that is mainly alpha helix and it is the only portion of the protein that shares an evident similarity to TAMO protein

Fig. 1. Nucleotide sequence of zebrafish spata2 cDNA. The 5'-UTR and 3'-UTR are shown in lowercase letters and the coding region (nucleotides 177-1830) in uppercase letters. The stop codon is asterisked. A putative polyadenylation signal is underlined. The deduced amino acid lowercase letters and the coding region (nucleotides 177-1830) in uppercase letters. The stop
Fig. 2. Multiple sequence alignment of 13 SPATA2 proteins and their 50% consensus threshold. The final alignment includes SPATA2 from D. rerio and 12 extracted proteins whose accession numbers are reported here. Sequences from Ensembl genome database are: T. rubripes NEWSNRJP0000179181, X. tropicalis ENSXETP00000051798 (gene prediction), M. domestica ENSMODP00000020381 and N. troglodites ENSPTRP00000023412. From GenBank: G. gallus CAG32546, B. taurus XP_608202.2, R. norvegicus AAK61814, M. musculus NP_739862, C. familiaris XP_543047, M. mulatta XP_001098381, P. troglodytes CAG32546, B. taurus XP_608202.2, R. norvegicus AAK61814, M. musculus NP_739862, C. familiaris XP_543047, M. mulatta XP_001098381, P. troglodytes CAG32546, B. taurus XP_608202.2, R. norvegicus AAK61814, M. musculus NP_739862. Secondary structure predictions of D. rerio (top of the alignment) and H. sapiens (bottom of the alignment) are shown.
been demonstrated to be a protein involved in nuclear import machinery during oogenesis in *D. melanogaster*. Nonetheless, no particular functional properties of the N-terminus of TAMO have been ascertained (Minakhina *et al.*, 2003) but strong local similarities (up to 62% highlighted in gray in the alignment in Fig. 3) to SPATA2 suggest a putative important role of this region. Different length patterns, designed on the most conserved portion of this particular region (data not shown), have been tested against protein and domain databanks but no significant hits have been retrieved. The small pattern PW [KR] KE [YF][RK] seems of particular interest due to its strong conservation between SPATA2 and TAMO proteins. It is located in a putative loop region from 100 to 110 residues that connects an alpha helix to a beta strand (Fig. 3). The direct consequence of this investigation indicates that this is a peculiar signature shared between TAMO and SPATA2 protein families.

**Spatiotemporal expression of spata2**

To analyze the spatiotemporal expression of *spata2* during early embryonic development, whole-mount in situ hybridization was performed from two-cell stage to five-day-old embryos using the antisense probe. *spata2* transcripts were already detected at two-cell stage, thus pointing to a maternal origin of the transcript (Fig. 4A). From the shield stage to 5-prim (24 hpf), *spata2* mRNAs were ubiquitously expressed (Fig. 4 B,C,D), however by the hatching period (48 hpf) a characteristic pattern was displayed in the head with marked staining in the midbrain–hindbrain boundary and in the otic vesicles (Fig. 4 E,F). Restricted signal localization was evident at 120 hpf in the oral cavity and in a dorsal portion of the periorbital region (Fig. 4 G,H). More caudally *spata2* expression was visible in the gut and pronephric duct luminal epithelium (Fig. 4 I,J). By means of a detailed examination on transverse section we showed that *spata2* mRNA is mainly present in luminal epithelia, such as gut and pronephric duct as well as in the oral cavity. To assess the specificity of the antisense probe, a sense

---

**Fig. 3. The preliminary domain structure of SPATA2.** Two distinct regions are apparent: the first one shares a strong similarity to TAMO protein and its composition is mainly alpha, whereas the second is mainly unstructured and does not have significant similarities to other proteins (see text for details). The highly conserved region between TAMO and SPATA2 is highlighted in light grey.

---

**Fig. 4 (Left). Expression of zebrafish spata2 during embryonic development.** *spata2* mRNA was already detected at the 2-cell stage (A) and was found to be distributed ubiquitously at shield (B), 20 somite (C) and 5-prim (D) stages. At the hatching period (48 hpf), a strong signal was detected in the head, particularly at the midbrain–hindbrain boundary and in the otic vesicles (E,F). Expression became more restricted at 5 dpf in the oral cavity and in the periorbital cartilage (G,H). More caudally *spata2* expression was visible in the gut and pronephric duct luminal epithelium (J). Abbreviations: mbh, midbrain–hindbrain boundary; e, eye; oc, oral cavity; gt, gut; pd, pronephric duct.

**Fig. 5 (Right). RT-PCR analysis of spata2 mRNA extracted from several adult zebrafish tissues and at different stages of development.** Total RNA from each tissue was reverse-transcribed into cDNA and subsequently amplified using zebrafish spata2 and β-actin (positive control) specific primers. The expected 614 bp band was observed in all tissues showing a ubiquitous spatial expression in adult stages. *spata2* was also detected at different early developmental stages (2-cell stage; 24 hpf; 6 dpf).
probe was used in a parallel control experiment at all stages but no staining was detected in any embryo (data not shown).

The consistency of hybridization experiments was confirmed by RT-PCR expression analysis performed on cDNAs from whole zebrafish embryos at various early developmental stages (Fig. 5).

Since in adult rats spata2 orthologue exhibits a restricted pattern of tissue distribution, we further addressed the question as to whether zebrafish spata2 maintain its ubiquitous spatial expression in adult stages. As shown in Fig. 5, RT-PCR based analysis demonstrated that almost all analyzed tissues of adult fish do display a high content of spata2-transcripts.

In conclusion, these results of whole-mount in situ hybridization and RT-PCR performed both in zebrafish embryos and adult tissues provide evidence that spata2 may have a broader function than previously described. The localization of its transcripts at early stages in the CNS as well as in peculiar epithelial-restricted domain in later stages suggests a function for spata2 in zebrafish development. Further analysis is ongoing in order to improve knowledge on the role of spata2.

**Experimental Procedures**

**Zebrafish embryo maintenance**

Zebrafish were raised and maintained under standard laboratory conditions at 28°C (Westerfield et al., 1995). Embryos at different stages were collected and stored in 0.003% phenyl-thiourea (PTU) (Sigma). This solution was added to the embryos to block pigment formation in order to improve visualization of RNA during in situ hybridization. The embryo stages were identified by morphological features and the corresponding embryos were fixed in 4% paraformaldehyde.

**Cloning and sequencing of the zebrafish spata2 gene**

Total RNA was isolated from pooled embryos at various developmental stages using the “TRIzol® Reagent” (Invitrogen) according to the manufacturer’s instructions. By means of the SMART RACE cDNA amplification kit (Clontech) we isolated the full-length of zebrafish spata2 cDNA. Using 1µg of total mRNA, the reverse transcription reaction was performed for 90 min at 42°C in a final volume of 10 µl in the presence of SMART II™ A Oligonucleotide and 200 U of PowerScript™ reverse transcriptase, 5 x First-Strand Buffer (250 mM Tris-HCl, pH 8.3; 375 mM KCl; 30 mM MgCl₂), 10 mM each of dNTPs and 20 mM dithiothreitol. RACE-PCR reactions were performed using two spata2 gene-specific primers (forward 5’-AGTGGCTGACTTCTTCCCA-3’ and reverse 5’-TGTTGTGAACTGCTCCAAGG-3’) and a universal anchor primer (5’-CTAATAGCTACATAGGGC-3’) complementary to the adaptor ligated to both ends of the ds cDNA, according to the manufacturer’s instructions. Amplification was performed for 25 cycles as follows: 30 sec denaturation at 94°C, 30 sec primer annealing at 68°C and 3 min extension at 72°C. The PCR product was electrophoresed on a 1% agarose gel in 1 x TAE buffer (50 mM Tris-HCl, pH 8.0; 20 mM sodium acetate; 2 mM Na₂EDTA) and ethidium bromide stained.

The gene-specific primers d-SPATA2zebra (5’-AGTGGCTGACTTCTTCCCA-3’) and r-SPATA2zebra (5’-TGTTGTGAACTGCTCCAAGG-3’) were designed to amplify 614 bp fragment (from base 909 to base 1523) of spata2 gene. Amplification was performed for 35 cycles as follows: 30 sec denaturation at 94°C, 30 sec primer annealing at 58°C and 1 min extension at 72°C. The PCR product was electrophoresed on a 1% agarose gel in 1 x TAE buffer (50 mM Tris-HCl, pH 8.0; 20 mM sodium acetate; 2 mM Na₂EDTA) and ethidium bromide stained.

Positive control primers were designed according to zebrafish β-actin (GenBank Accession No. AF057040) as d-β-Actin: 5’-TGTTTTCCCCCTCATTGGTG-3’ and r-β-Actin: 5’-TTCCTCCTGATGTCACGGAC-3’ to amplify a 588 nucleotide fragment. All primers were selected from two exons separated by one or more intronic sequences, to identify possible ampiclons from contaminating genomic DNA.

**Bioinformatic analysis of zebrafish spata2**

Sequences of SPATA2 were searched for using PSI-BLAST (Altschul et al., 1997) vs. non-redundant database. Sequences annotated as putative spata2 genes were extracted and enriched with sequences coming from Ensembl database of genome projects (http://www.ensembl.org). The multiple alignment was performed by using MUSCLE (Edgar, 2004) and manual edited to identify blocks of identities among sequences. The final version of the alignment figure was prepared using ESPRIT (Gouet et al., 2003). The secondary structures of proteins were predicted using the consensus method (Albrecht et al., 2003) whereas prediction of disorder regions was performed using a similar approach applying state-of-the-art methods such as DISOPRED (Ward et al., 2004) and SPRITZ (Vullo et al., 2006). The presence of functional domains was searched against Pfam database (Finn et al., 2006), PROSITE (Hulo et al., 2006) and InterPRO (Mulder et al., 2005). Pattern search of putative functional signatures extracted from the conserved regions of the multiple alignment was performed using ScanProsite (de Castro et al., 2006). Polyadenylation sites was predicted by means of Hcpolya (http://www.itb.cnr.it/sun/webgene/) and POLYAH (www.softberry.com) programs.

**Whole-mount in situ hybridization**

spata2 antisense and sense riboprobes were synthesized from spata2 cDNA, corresponding to the entire coding sequence region (1650 bp) of zebrafish spata2 gene, using SP6 and T7 RNA polymerase respectively. Whole-mount in situ hybridization was carried out as previously described (Thisset et al., 1993). In brief, embryos were permeabilized with Protease K (10 µg/ml, Sigma) and hybridized overnight at 66°C with the DIG-labeled antisense or sense riboprobes. After several washes at high stringent temperature, NBT/BCIP (Roche) staining was performed according to the manufacturer’s instructions.

Images were obtained by a Nikon SMZ 1500 digital camera system. For a more detailed analysis, vibrotome sections were prepared from gelatine-albumine embedded embryos after the staining.

**Multi-tissue RT-PCR**

To reveal the tissue distribution and expression of zebrafish spata2 gene total RNA was extracted from various developmental stages and several fully developed organs and tissues of zebrafish. Total RNA was extracted from the tissues using the “TRIzol® Reagent” (Invitrogen) according to the manufacturer’s instructions. Two micrograms of total RNA from each sample were used for synthesizing the first strand cDNA by means of reverse transcriptase.

The gene-specific primers d-SPATA2zebra (5’-AGTGGCTGACTTCTTCCCA-3’) and r-SPATA2zebra (5’-CTCGGAGCTTCACACA-3’) were designed to amplify 614 bp fragment (from base 909 to base 1523) of spata2 gene. Amplification was performed for 35 cycles as follows: 30 sec denaturation at 94°C, 30 sec primer annealing at 58°C and 1 min extension at 72°C. The PCR product was electrophoresed on a 1% agarose gel in 1 x TAE buffer (50 mM Tris-HCl, pH 8.0; 20 mM sodium acetate; 2 mM Na₂EDTA) and ethidium bromide stained.

Positive control primers were designed according to zebrafish β-actin (GenBank Accession No. AF057040) as d-β-Actin: 5’-TGTTTTCCCCCTCATTGGTG-3’ and r-β-Actin: 5’-TTCCTCCTGATGTCACGGAC-3’ to amplify a 588 nucleotide fragment. All primers were selected from two exons separated by one or more intronic sequences, to identify possible ampiclons from contaminating genomic DNA.

**References**

ALBRECHT, M., TOSATTO, S.C., LENGAUER, T. and VALLE, G. (2003). Simple consensus procedures are effective and sufficient in secondary structure
DYSON, H.J. and WRIGHT, P.E. (2005). Intrinsically unstructured proteins and

GRAZIOTTO, R., FORESTA, C., SCANNAPIECO, P., ZEILANTE, P., RUSSO, A.,


FINN, R.D., MISTRY, J., SCHUSTER-BOCKLER, B., GRIFFITHS-JONES, S.,


ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., ZHANG, J., ZHANG, Z.,


DOOLEY, K. and ZON, L.I. (2000). Zebrafish: a model system for the study of

MULDER, N.J., APWEILER, R., ATTWOOD, T.K., BAIROCH, A., BATEMAN, A.,

DE CASTRO, E., SIGRIST, C.J., GATTIKER, A., BULLIARD, V., LANGENDEJK-


Received: 13th September 2006
Reviewed by Referees: 31st October 2006
Modified by Authors and Accepted for Publication: 15th December 2006
Published Online: 4th April 2007

Related previously published Int. J. Dev. Biol. articles

See our recent Special Issue on Mammalian Reproduction and Development edited by B. Hogan at: http://www.ijdb.ehu.es/web/contents.php?v=45&issue=3


mgm 1, the earliest sex-specific germline marker in Drosophila, reflects expression of the gene esg in male stem cells. Adrian Streit, Luca Bernasconi, Pavel Sergeev, Alex Cruz and Monica Steinmann-Zwicky. Int. J. Dev. Biol. (2002) 46: 159-166

