Int. J. Dev. Biol. 51: 241-246 (2007)
doi: 10.1387/ijdb.062220em

Developmental Expression Pattern

Zebrafish *spata2* is expressed at early developmental stages

ENRICO MORO[1,#], CLAUDIO MARAN[3,#], M. LILIANA SLONGO[3], FRANCESCO ARGENTON[1], STEFANO TOPPO[2] and MAURIZIO ONISTO[3,*]

[1]Dept. of Biology, [2]Dept. of Biological Chemistry and [3]Dept. 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

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, digoxygenin; 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, untraslated 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

Printed in Spain
www.intjdevbiol.com

```
   1                                                tgtctctcttttgaggacgatctcagcttgctgattttcactgccacact   50
  51  gcaccattgccccagatcttctggtgtctagtgggagagaggtcttgagatggcaaggggtccttttttaaagaaacacctggattgtgga   140
 141  ataagaggttgcaagctacgcagcccactgccaccaATGGATGCCAAGTTGCGAGAGGATCTGTTCCGCAGGTATGTGGCGTCTCTGGAG   230
                                        M  D  A  K  L  R  E  D  L  F  R  R  Y  V  A  S  L  E
 231  AATCGGCTCGAGGAGGGAGCCGCTGAAGGAGGAGGAGATAGACAAGGAGCTAAGACTGAGGAGGCACTGATTTCTACTGCAACAGCATTG   320
       N  R  L  E  E  G  A  A  E  G  G  G  D  R  Q  G  A  K  T  E  E  A  L  I  S  T  A  T  A  L
 321  CTAGGCTCATACCAGGCTGATCCAGGGCAGCGATTTCGCATGCTTCGCTTCTATGATGTTGTTGAGAACTCCCTGAGGACTCAGAGAGGC   410
       L  G  S  Y  Q  A  D  P  G  Q  R  F  R  M  L  R  F  Y  D  V  V  E  N  S  L  R  T  Q  R  G
 411  ACAAGCTTGCGGACAGTGGGAACAGCCTTTGCAACGCTGGAAACCATCTGCACCAATTTGCTGCTTTTTCCATGGAAGAAGGAGTACCGT   500
       T  S  L  R  T  V  G  T  A  F  A  T  L  E  T  I  C  T  N  L  L  L  F  P  W  K  K  E  Y  R
 501  TGCATAAAGACCTTCACAGGGCCATACGTGTACCAGTTGCAGTCTGTGTTATGCGACTCCGACCTGCGATCTCTACTGCGTTCCATGGGC   590
       C  I  K  T  F  T  G  P  Y  V  Y  Q  L  Q  S  V  L  C  D  S  D  L  R  S  L  L  R  S  M  G
 591  TATTCAAGAGAACAGGAATTGCAGTACAATGTTAGGGACCATCCAGGTGGAGCTTCTCACCTTCGGCAGCTTGCTTTTGAACTGCTTCTG   680
       Y  S  R  E  Q  E  L  Q  Y  N  V  R  D  H  P  G  G  A  S  H  L  R  Q  L  A  F  E  L  L  L
 681  GCTCAGGCAGAGTGCCGTCTGTTGGGCGAGGTGGTTTCCATGTCTCGTGGCTTTGCATCAGAGCTGGAGGCTGTGGAGTTGAGAAGAAAC   770
       A  Q  A  E  C  R  L  L  G  E  V  V  S  M  S  R  G  F  A  S  E  L  E  A  V  E  L  R  R  N
 771  ACCAGAGAAGATGCAGCGGGGTGTGCAGATGCCCTTCGTAGACGAGACAGCCTCACTGGAGATTTGTCCCGCCTGTCTGTGCGACCGCTA   860
       T  R  E  D  A  A  G  C  A  D  A  L  R  R  R  D  S  L  T  G  D  L  S  R  L  S  V  R  P  L
 861  GATATTGATCGTGCCCACCTAAGGCGAAGCGGTCGCCCATCCAAGTCAGTGCGACCGCTAGATATTGATCGTGCCCACCTAAGGCGAAGC   950
       D  I  D  R  A  H  L  R  R  S  G  R  P  S  K  S  V  R  P  L  D  I  D  R  A  H  L  R  R  S
 951  GGTCGCCCATCCAAGTCAGCACTCAGTCTGCGCAAGGAGCCGTTGTTTGTTGACACAGAAGAAGACATGAAAGACGAAATCATCCGACCT  1040
       G  R  P  S  K  S  A  L  S  L  R  K  E  P  L  F  V  D  T  E  E  D  M  K  D  E  I  I  R  P
1041  AGCCCCTCGCTCTATCCAGTGGCAGCGCCACCTTCTTACAGTCCAGTGGCTGACTTCTTCCCAATTCAGTCGCCCCCGTCAGAGCCATAC  1130
       S  P  S  L  Y  P  V  A  A  P  P  S  Y  S  P  V  A  D  F  F  P  I  Q  S  P  P  S  E  P  Y
1131  TCTTACCACTTGTCCTCCCTGGACGAGGTGGACCTGTACACCGAGCGGGGGTTGGGTGGCCACCAGACACCCTCTAGGCCACCGAGCCGG  1220
       S  Y  H  L  S  S  L  D  E  V  D  L  Y  T  E  R  G  L  G  G  H  Q  T  P  S  R  P  P  S  R
1221  GAGCCTCGAGACAGTTGGGTTTTGAAAGGACACATGATGAAGTGCCAGGGCTGTGGCATGGGCTGCCCTTCGCTTTCCTCTTGCCAAAGG  1310
       E  P  R  D  S  W  V  L  K  G  H  M  M  K  C  Q  G  C  G  M  G  C  P  S  L  S  S  C  Q  R
1311  TGCGACATGATCTTGTGCTCCTCTTGTCATGCGGTGGACCCAGCACCATGTTGCGGCTTCCAGGATTACACTAAAACCTCCCGCCCTCTG  1400
       C  D  M  I  L  C  S  S  C  H  A  V  D  P  A  P  C  C  G  F  Q  D  Y  T  K  T  S  R  P  L
1401  GATGGCTACATGCCCATCAAGGAAAAATTGTCTGTCTACTCCAGTGCCCACTCGCACACCCACCCTCATCCTCACCCACTTCCCCACGCT  1490
       D  G  Y  M  P  I  K  E  K  L  S  V  Y  S  S  A  H  S  H  T  H  P  H  P  H  P  L  P  H  A
1491  CAGTCCCATTCGCTGTTGCTGGACAAAGCTGTGATGAGCTCCAAGCTGTTTCCCAGTAAACCTGTGGGTTCGGGCCCCAGTCCGGTGGGC  1580
       Q  S  H  S  L  L  L  D  K  A  V  M  S  S  K  L  F  P  S  K  P  V  G  S  G  P  S  P  V  G
1581  AGCTTGGTCAGCAGTGGCAGCAGCAGCAGTGGGGGTGAGCGGCTGAGTGTGGGCGGCTCCCGCTGTGGCTTCTGCAACAAGCCAGGTGCC  1670
       S  L  V  S  S  G  S  S  S  S  S  G  G  E  R  L  S  V  G  G  S  R  C  G  F  C  N  K  P  G  A
1671  TCTCACACCTGTGTGAACTGCTCCAAGGTTTCATGTGACACCTGCATGAGCCTCTACGCCAGTGATTTGTGCACTCGCAAAAATCCCCAT  1760
       S  H  T  C  V  N  C  S  K  V  S  C  D  T  C  M  S  L  Y  A  S  D  L  C  T  R  K  N  P  H
1761  CATAACTTTGTGCCCAATCATCAACTCAATTTCAAATCCAGCACCATATCTCACCTCGTATATCGATAGgaccacgcgtatcgatgtcga  1850
       H  N  F  V  P  N  H  Q  L  N  F  K  S  S  T  I  S  H  L  V  Y  R  *
1851  cttttttttttttttaaatatacaaagcttttgctctaaatttacaataaaaattcatgcatgaactcatcaaaattttaccggcaggca  1940
1941  atatgcggctgccctggcatcatatgttggactgacgcagcgggtgttgcagtagaggtagtggtgtcgtgtagggacgcagtcgaacat  2030
2031  caactccgaaaaacccttcgaacgtggattgcttctcaggacagacctaatggatttaagcggatatgggggacgacttttccacgggcc  2120
2121  tatagccaaatctcagcggaggcgtggtcgaataaggacgatcgtctccgcagcacactaacggcattcatcagacttggccaacaacgg  2210
2211  gacagatgccatatcggattttcaaacattcattcaccacgcctggaaaggaaagaacagaacttcgcgagcaaagggttctggaataca  2300
2301  aactgaaattatgacttaagttacacagaatacctcaacccatacaaaccaaatgccctcctcatgactgtgctatggcaacgacccctg  2390
2391  ataccctctgcttgtgagccctctttctcctgctgcttttgggacacccatcttttggcttgctctgatctagaccagctgagctccccc  2480
2481  tggtggtcaacgtggatctgaagctactctgccttgtagacccggctccccctttctggatactgctgcgcttcaccctgatctgtgccct  2570
2571  ctgatcctcccctggtttctgcctggctcgccaccttgtcctgatttctccctggtccctgctgtagatcttttttcttccatccgacgtcat  2660
2661  ctaagagctatgttgttactggataaactcagtccatttctcctgtcctctgcacatccgtcttctgctgatatctttggatcatgaaaa  2750
2751  caacctacgcttccatttcactgctgttactcaacatgactcagtccatacagcaatcaattacgtggttcagtacaaaaaaaagcatca  2840
2841  ttcttcattgattcttctaggcttgcaaatttaaatcattgaaatactcaattcttcttcttctc-polyA                      2905
```

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 sequence (540 amino acids) is also shown below the nucleotide sequence.*

during early developmental stages.

Isolation and characterization of zebrafish spata2 *cDNA sequence*

Starting from the amino acid sequence of human SPATA2 we have identified, by PSI-BLAST alignment, the potential orthologue in zebrafish. Using the Ensembl zebrafish server, we retrieved the relative nucleotide sequence (Ensembl Transcript ID ENSDART00000034839) classified as similar to *spata2* gene. Since the deposited sequence in Ensembl Database was not completed, we sought to confirm the *spata2* nucleotide sequence and to obtain the full-length cDNA. cDNA was generated by means of a RACE-PCR strategy. Using *spata2* gene-specific primers, we obtained 5' and 3' RACE fragments which were assembled to generate a complete cDNA sequence (GenBank Accession No. DQ869310). The total length of the cDNA sequence, as shown in Fig. 1, is 2905 bp excluding the poly (A) tail. 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.

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 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* seem to differ from 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 (Minakhina *et al.*, 2003), as will be discussed further. The remaining portion of the protein has a central region from aa 220 to aa 480 that does not seem to be conserved. This lack of conservation is both due to mismatches and deletions that are reported as dots in the multiple alignment in Fig. 2. From aa 480 to aa 550 it is possible to observe again a high sequence conservation along this protein family. Both secondary structure and disorder predictions have also detected a wide region starting from residue 220 up to the end of the protein with few secondary structures and highly disordered segments confirmed by the presence of prolines and charged amino acids. In addition, this region, if compared with the N-terminus, contains a high content of cysteines that may play an important role in either keeping the tertiary structure of this domain tight or interacting with other proteins in signal pathways forming inter/intra chain disulfide bonds. Long disorder regions are, generally, involved in communication and interaction processes (Dyson and Wright, 2005) suggesting that the C-terminus of SPATA2 may play a similar functional role.

As already detected, the only evident similarity of SPATA2 is with TAMO protein. SPATA2 shows a similarity of over 43% in the first 174 residues of the N-terminal region to the N-terminus of TAMO protein (Fig. 3) which, in addition, shares the same predicted secondary structure topology with SPATA2. TAMO has

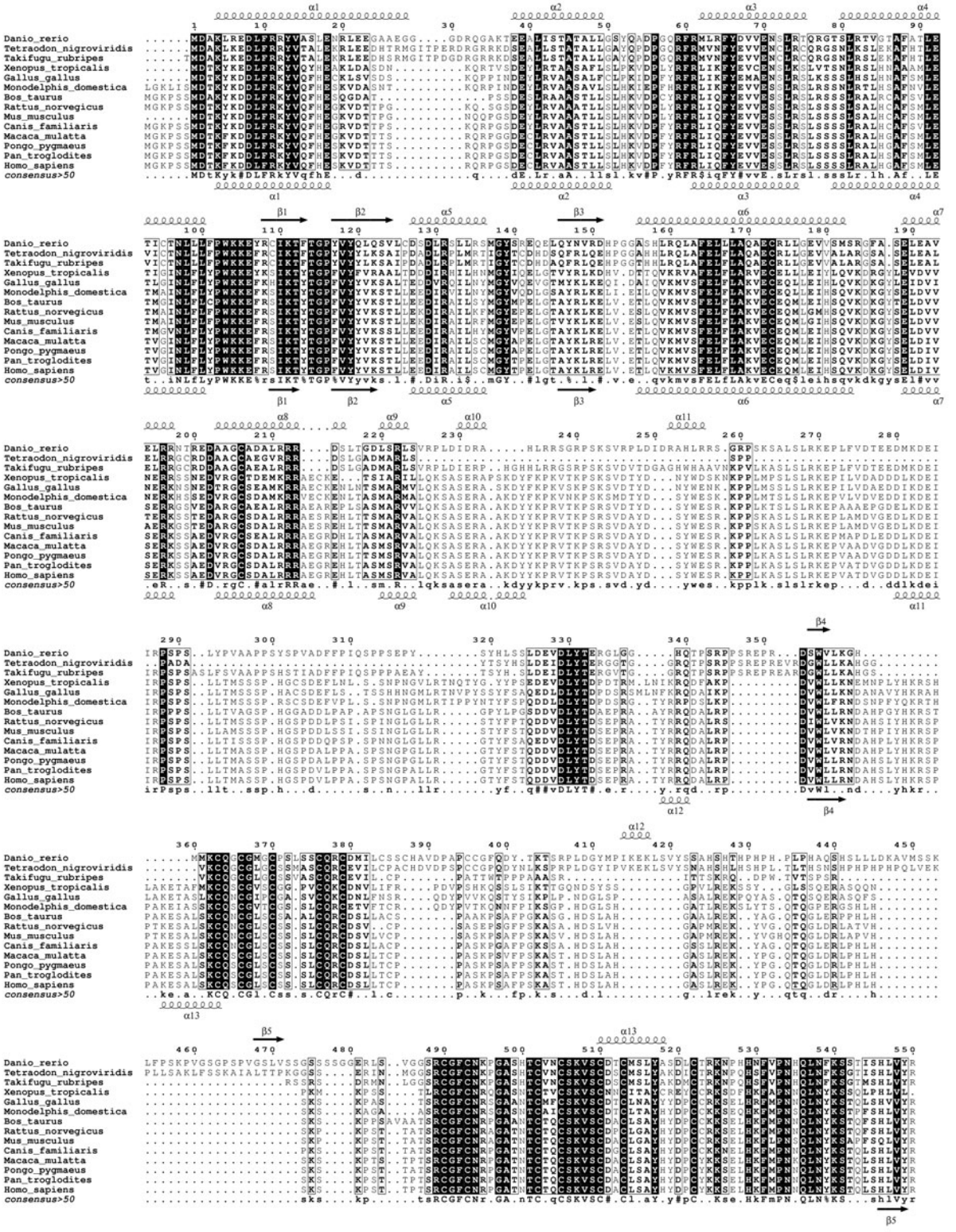


Fig. 2. Multiple sequence alignment of 13 SPATA2 proteins and their 50% consensus threshold. *The final alignment includes SPATA2 from D. rerio and 12 extracted proteins whose accession numbers are reported here. Sequences from Ensembl genome database are:* T. rubripes *NEWSINFRUP00000179181,* X. tropicalis *ENSXETP00000051789 (gene prediction),* M. domestica *ENSMODP00000020381 and* P. troglodites *ENSPTRP00000023412. From GenBank:* G. gallus *CAG32546,* B. taurus *XP_608202.2,* R. norvegicus *AAK61814,* M. musculus *NP_739562,* C. familiaris *XP_543047,* M. mulatta *XP_001098381,* P. pigmaeus *CAH91094 and* H. sapiens *CAB46029. Secondary structure predictions of* D. rerio *(top of the alignment) and* H. sapiens *(bottom of the alignment) are shown.*

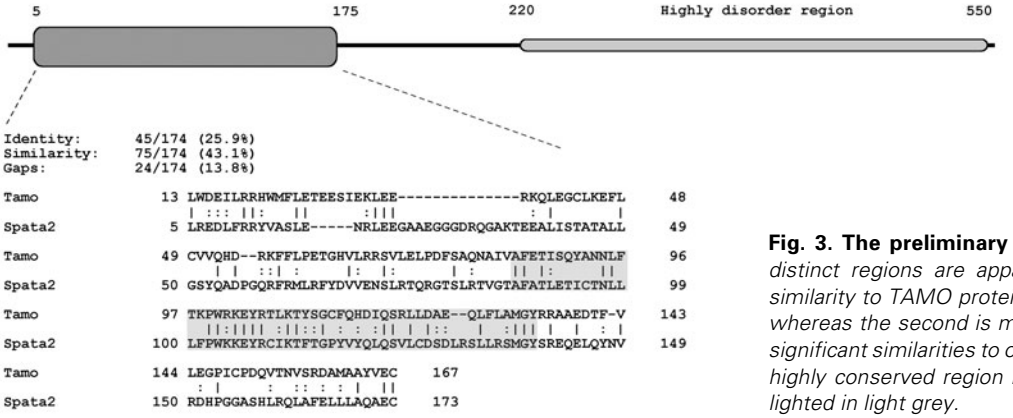
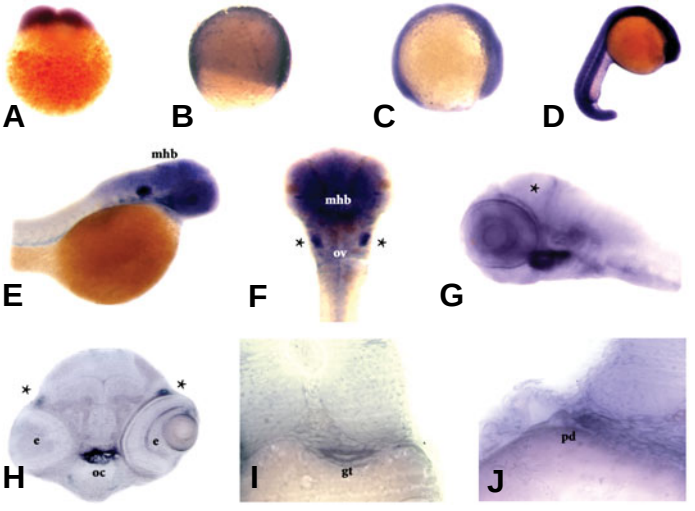


Fig. 3. The preliminary domain structure of SPATA2. *Two distinct regions are apparent: the first one shares a strong similarity to TAMO protein and its composition is mainly alpha, whereas the second is mainly unstructured and does not have significant similarities to other proteins (see text for details).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

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 vescicles (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



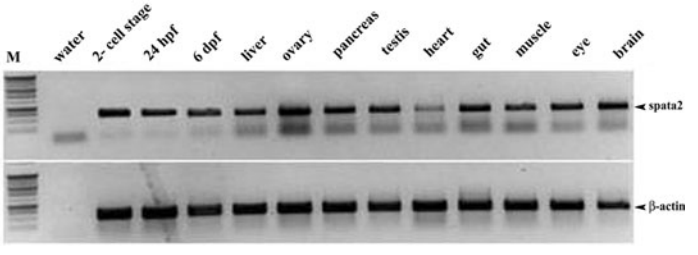


Fig. 4 (Left). Expression of zebrafish *spata2* during embryonic development. spata2 *mRNA was already detected at the 2-cell stage* **(A)** *and was found to be distributed ubiquitously at shield* **(B)**, *20 somite* **(C)** *and 5-prim* **(D)** *stages. At the hatching period (48 hpf), a strong signal was detected in the head, particularly at the midbrain-hindbrain boundary and in the otic vescicles* **(E,F)**.*Expression become more restricted at 5 dpf in the oral cavity and in the periorbital cartilage* **(G,H)**. *More caudally, vibratome sections allowed the identification of* spata2 *mRNA in the gut* **(I)** *and pronephric duct 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 28ºC (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 $MgCl_2$), 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'-TGTGTGAACTGCTCCAAGG-3') and a universal anchor primer (5'–CTAATACGACTCACTATAGGGC-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 Na_2EDTA) and ethidium bromide stained. The PCR product was cloned directly into pCR®II-TOPO® Vector (TOPO TA Cloning® system, Invitrogen) and sequenced on both strands using Big Dye terminator v3.1 protocol on ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

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 Espript (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). Polyadenylations site 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 a *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 (Thisse *et al.,* 1993). In brief, embryos were permeabilized with Proteinase K (10 μg/ml, Sigma) and hybridized overnight at 65˚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, vibratome 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'-CCTTGGAGCAGTTCACACA-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 Na_2EDTA) and ethidium bromide stained.

Positive control primers were designed according to zebrafish *β-actin* (GenBank Accession No. AF057040) as d-β-Actin: 5'-TGTTTTCCCCTCCATTGTTGG-3' and r-β-Actin: 5'-TTCTCCTTGATGTCACGGAC-3' to amplify a 558 nucleotide fragment. All primers were selected from two exons separated by one or more intronic sequences, to identify possible amplicons from contaminating genomic DNA.

Acknowledgments

This work was supported by University of Padova (ex 60% Grant) and by the European Commissions as part of the ZF-MODELS Integrated Project in the Sixth Framework Program (Contract no. LSHG-CT-2003-503496).

References

ALBRECHT, M., TOSATTO, S.C., LENGAUER, T. and VALLE, G. (2003). Simple consensus procedures are effective and sufficient in secondary structure

prediction. *Protein Eng.* 16(7): 459-462.

ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W. and LIPMAN, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17): 3389-3402.

DE CASTRO, E., SIGRIST, C.J., GATTIKER, A., BULLIARD, V., LANGENDIJK-GENEVAUX, P.S., GASTEIGER, E., BAIROCH, A. and HULO, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.*34: W362-365.

DOOLEY, K. and ZON, L.I. (2000). Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* 10: 252-256.

DYSON, H.J. and WRIGHT, P.E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell. Biol.* 6(3): 197-208.

EDGAR, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5: 113.

FINN, R.D., MISTRY, J., SCHUSTER-BOCKLER, B., GRIFFITHS-JONES, S., HOLLICH, V., LASSMANN, T., MOXON, S., MARSHALL, M., KHANNA, A., DURBIN, R., EDDY, S.R., SONNHAMMER, E.L. and BATEMAN, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res.* 34: D247-251.

GOUET, P., ROBERT, X. and COURCELLE, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 31(13): 3320-3323.

GRAZIOTTO, R., FORESTA, C., SCANNAPIECO, P., ZEILANTE, P., RUSSO, A., NEGRO, A., SALMASO, R. and ONISTO, M. (1999). cDNA cloning and characterization of PD1: a novel human testicular protein with different expressions in various testiculopathies. *Exp. Cell Res.* 248: 620-626.

HULO, N., BAIROCH, A., BULLIARD, V., CERUTTI, L., DE CASTRO, E., LANGENDIJK-GENEVAUX, P.S., PAGNI, M. and SIGRIST, C.J. (2006). The PROSITE database. *Nucleic Acids Res.* 34: D227-230.

MINAKHINA, S., YANG, J. and STEWARD, R. (2003). Tamo selectively modulates nuclear import in Drosophila. *Genes Cells.* 8: 299-310.

MULDER, N.J., APWEILER, R., ATTWOOD, T.K., BAIROCH, A., BATEMAN, A., BINNS, D., BRADLEY, P., BORK, P., BUCHER, P., CERUTTI, L., COPLEY, R., COURCELLE, E., DAS, U., DURBIN, R., FLEISCHMANN, W., GOUGH, J., HAFT, D., HARTE, N., HULO, N., KAHN, D., KANAPIN, A., KRESTYANINOVA, M., LONSDALE, D., LOPEZ, R., LETUNIC, I., MADERA, M., MASLEN, J., MCDOWALL, J., MITCHELL, A., NIKOLSKAYA, A.N., ORCHARD, S., PAGNI, M., PONTING, C.P., QUEVILLON, E., SELENGUT, J., SIGRIST, C.J., SILVENTOINEN, V., STUDHOLME, D.J., VAUGHAN, R. and WU, C.H. (2005). InterPro, progress and status in 2005. *Nucleic Acids Res.* 33: D201-205.

ONISTO, M., GRAZIOTTO, R.,SCANNAPIECO, P., MARIN, P., MERICO, M., SLONGO, M.L. and FORESTA, C. (2000). A novel gene (PD1) with a potential role on rat spermatogenesis. *J. Endocrinol. Invest.* 23: 605-608.

ONISTO, M., SLONGO, M.L., GRAZIOTTO, R., ZOTTI, L., NEGRO, A., MERICO, M., MORO, E. and FORESTA, C. (2001). Evidence for FSH-dependent upregulation of SPATA2 (Spermatogenesis-Associated Protein 2). *Biochem. Biophys. Res. Comm.* 283: 86-92.

RUBINSTEIN, A.M. (2003). Zebrafish: from disease modeling to drug discovery. *Curr. Opin. Drug Discov. Dev.* 6(2): 218-223.

SLONGO, M.L., ZOTTI L. and ONISTO, M. (2003). Cloning and characterization of the promoter region of human SPATA2 (Spermatogenesis-Associated Protein 2) gene. *Bioch. Biophys. Acta.* 1625: 192-196.

THISSE, C., THISSE, B., SCHILLING, T.F. and POSTLETHWAIT, J.H. (1993). Structure of the zebrafish snail1 gene and its expression in wild-tipe, spadetail and no tail mutant embryos. *Development.* 119: 1203-1215.

VULLO, A., BORTOLAMI, O., POLLASTRI, G. and TOSATTO, S.C. (2006). Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucleic Acids Res.* 34: W164-168.

WESTERFIELD, M. et al. (1995). The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish. Eugene Press.

WARD, J.J., SODHI, J.S., MCGUFFIN, L.J., BUXTON, B.F. and JONES, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337(3): 635-645.

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