Genetics and Molecular Biology, 28, 4, 839-842 (2005)

www.sbg.org.br

Research Article

Characterization and copy number of the S27 ribosomal protein gene from amphioxus *Branchiostoma belcheri tsingtauense*

Lifang Ma, Shicui Zhang, Zhenhui Liu, Hongyan Li and Jianjun Xia

Ocean University of China, Department of Marine Biology, PR, China.

Abstract

A cDNA clone encoding ribosomal protein S27 (*AmphiS27*) was identified in the gut cDNA library of amphioxus *Branchiostoma belcheri tsingtauense*. This cDNA consists of 607 bp and contains a 255 bp open reading frame (ORF) corresponding to a deduced protein of 84 amino acids with a calculated molecular mass of 9,488 Da and an isoelectric point (pI) of 9.500. Alignment of the deduced *AmphiS27* amino acid sequence with 12 known S27 protein sequences indicates that *AmphiS27* shares 94-99% homology with its vertebrate homologue, 84-94% with invertebrate homologues and 69-72% with homologues from other eukaryotes, suggesting that *AmphiS27* is more closely related to the vertebrate S27 protein than to its invertebrate counterpart. Southern blot analysis showed a single copy of the S27 gene present in the genome of amphioxus *B. belcheri tsingtauense*, indicating that amphioxus has a genome uncomplicated by extensive gene duplication.

Key words: Amphioxus, *Branchiostoma*, ribosomal protein, S27, copy number.

Received: July 23, 2004; Accepted: May 4, 2005.

Introduction

Ribosomes are RNA-protein complex organelles catalyzing mRNA-directed protein synthesis in all organisms. The structure of ribosome is conserved throughout the prokaryotic and eukaryotic lineages, reflecting the early origin of their essential function. Each ribosome consists of a large (L) and small (S) subunits. In eukaryotes, the large 60S subunit is composed of three ribosomal RNAs (rRNAs) and nearly 50 ribosomal proteins while the small 40S subunit consists of one rRNA and about 30 proteins (Kay and Jacobs-Lorena, 1987). Information contained in the sequences of ribosomal proteins can help in unraveling their evolution and function.

The ribosomal S27 protein appears to be involved in RNA-DNA binding and may influence gene transcription (Revenkova *et al*., 1999; Wool *et al*., 1990). The gene encoding the S27 protein has been identified in all eukaryotes including animals (Chan *et al.*, 1993; Thomas *et al*., 2000; Snyder 1999), plants (Gao *et al*., 1994; Hahn and Kück, 1995) and fungi (Steele and Jacobson, 1986). However, no information is currently available about the S27 protein in the cephalochordate amphioxus, an extant invertebrate most closely related to the proximate ancestor of vertebrates (Strokes and Holland, 1998; Zhang *et al*., 2001).

Send correspondence to Shicui Zhang. Ocean University of China, Department of Marine Biology, Qingdao 266003 PR, China. Email: sczhang@ouc.edu.cn.

Gene and genome duplication has recently been an interesting topic for biologists (Meyer and Schartl, 1999; Sankoff, 2001) and different copy numbers of the S27 protein gene have been detected in several higher eukaryotes such as rats (Chan *et al*., 1993), lobsters and mussels (Snyder, 1999) and also in tobacco plants (Gao *et al*., 1994), but such data are still limited. The aim of this study was to identify the S27 protein gene in the gut cDNA library of amphioxus *Branchiostoma belcheri tsingtauense* and to determine the copy number of this gene in amphioxus.

Material and Methods

A gut cDNA library of adult amphioxus *Branchiostoma belcheri tsingtauense* was constructed using the SMART cDNA Library Construction Kit (CLONTECH, Palo Alto, CA, USA) according to the method described previously by Liu *et al.* (2002) and cDNA clones selected for sequencing. Both strands of all selected clones were sequenced using an ABI PRISM 377XL DNA Sequencer and all sequences were then analyzed for coding probability with the DNATools program (Rehm, 2001).

To determine the identity of the gene we compared our sequences against the GenBank protein database using the BLAST network server at the National Center for Biotechnology Information (Altschul *et al*, 1997). Multiple protein sequences were aligned using the MegAlign program by the CLUSTAL method using the DNASTAR software package (Burland, 2000).

Genomic DNAs for Southern blotting were isolated from adult amphioxus. A total of 30 amphioxus were ground in liquid nitrogen and the powder suspended in 15 mL of lysis buffer containing 10 mM Tris-HCl (pH8.0), 100 mM EDTA and 0.5% SDS. After treatment with proteinase K (100 mg/mL, final concentration) at 55 °C for 3 h the mixture was cooled to room temperature and mixed with equal volume of saturated phenol (pH8) before being centrifuged at 5000 g at 4 °C for 20 min, the supernatant being pooled and then mixed with an equal volume of 1:1 (v:v) phenol-chloroform and then centrifuged as above and the supernatant collected from which the DNA was precipitated by ethanol and digested with the *EcoR*, *Hind*, *Bgl* and *Bst*X restriction enzymes at 37 °C for 20 h. The digested DNA was separated on 1% agarose gel using 1 X TBE (89 mM Tris-borate and 2 mM EDTA), transferred to a nylon membrane (Osmonics Inc.) and hybridized with digoxigenin (DIG)-labeled DNA probes produced using the DIG DNA labeling kit (Roche) corresponding to cDNA clone 094. Hybridized bands were visualized according to the instruction of the detection kit.

Results and Discussion

The nucleotide sequence obtained from the randomly selected 094 clone contained 607 base pairs made up of a 32 bp 5' untranslated region (UTR), a 255 bp open reading frame (ORF) and a 320 bp 3' UTR (GenBank accession number: AY168455). The ORF encoded a protein with a calculated molecular mass of 9,488 Da made up of 84 amino acids (Figure 1) while the 5' UTR had an in-frame stop codon upstream of the first ATG start codon and a CTTTTC polypyrimidine sequence which has been found at the 5' end of many eukaryotic protein mRNAs (Wool *et al*, 1996), the 3' UTR possessed an AATTAA polyadenylation signal 16 bases upstream of the poly(A) site. It was clear that the cDNA contained the full-length sequence for a protein.

```
GCTTTTCGGCCGACGTGAGAAAGTCAGCCACCATGCCTTTAGCAAAGGACTTGCTGCATC  60
                                  M  P  L  A  K  D  L  L  H  P  10
CCTCCCCAGAGGAGGAGAAGAGGAAACACAAGAAGAAGAGGCTGGTTCAGAGCCCCAATA 120
  S  P  E  E  E  K  R  K  H  K  K  K  R  L  V  Q  S  P  N  S  30
GTTACTTCATGGACGTCAAATGTCCTGGTTGCTATAAGATCACCACAGTCTTCAGCCATG 180
  Y  F  M  D  V  K  C  P  G  C  Y  K  I  T  T  V  F  S  H  A  50
CCCAGACAGTAGTTCTGTGTGTGGGGTGTTCAACAGTGCTGTGTCAGCCAACAGGGGGCA 240
  Q  T  V  V  L  C  V  G  C  S  T  V  L  C  Q  P  T  G  G  R  70
GGGCTAGGCTTACAGAAGGCTGCTCGTTCAGACGGAAGCAGCACTAGAGTCAAGGAAAGG 300
  A  R  L  T  E  G  C  S  F  R  R  K  Q  H  *                 84
AGAGGAGAAAGCCGGCCACCAGCCTGTTGACTTCAGCAGACACTAGAGCGAAAACAAGAG 360
GCAATGCCACGTCCTGGTCCCATTCCCCGCCTTCAAACACGCTTGAGAATGGCTCCCCCA 420
CGGATGTGTGCCTTCAACTTAGTCAAGCTATACTCGGTCATCATAAAATTGCCTGCAAAA 480
TCTGCATTAAATCTCCACTTAGTGTATTGCTTGTAAAACACGTTGGTGGCAATTAAAATC 540
CTATCTCAAGCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 600
AAAAAAA                                                        607
```

Figure 1 - Nucleotide and deduced amino acid sequences of the amphioxus S27 protein gene (*AmphiS27*; GenBank accession number AY168455). The presumed translational start and terminal sites are underlined and the asterisk represents the stop codon. The polyadenylation signal at the 3' end is boxed and the polypyrimidine sequence at the 5' end is double underlined. The basic amino acid cluster in *AmphiS27* is marked by a heavy bar.

Our initial BLAST search revealed that the amphioxus cDNA-encoded protein had high homology with the S27 protein and had 85% homology with its human homologue and 84% with its counterpart from zebrafish and the fish *Ictalurus punctatus*. We also found that the amphioxus S27 protein is a basic protein with an isoelectric point (pI) of 9.500, a common feature of many ribosomal proteins (Wool *et al*, 1995), and has a zinc finger $CX_2CX_{14\text{-}15}CX_2C$ motif characteristic of the S27 protein (Chan *et al*, 1993). Taken together these factors indicate that the cloned amphioxus gene codes for the S27 ribosomal protein and we thus designated the gene as *AmphiS27*. Eight of the 18 basic amino residues (6 arginyl, 9 lysyl and 4 histidy) at positions 16-23 were clustered together in *AmphiS27*, this amino acid combination being thought to function as a nuclear localization signal (Revenkova *et al*, 1999).

Comparison of the deduced *AmphiS27* amino acid sequence with 12 known S27 protein sequences (Table 1) revealed that *AmphiS27* shares 94-99% homology with its

Table 1 - Representative members of the S27 ribosomal protein family.

Protein	Organism (abbreviation)	Accession number	Amino acid	Source
S27Hs	*Homo sapiens* (Hs)	CAC36086	84	EMBL
S27-1Rn	*Rattus norvegicus* (Rn)	AAD56582	84	GenBank
S27Xl	*Xenopus laevis* (Xl)	P47904	84	SwissProt
S27-1Ip	*Ictalurus punctatus* (Ip)	AAK95210	84	GenBank
S27-2Ip	*Ictalurus punctatus* (Ip)	AAK95211	84	GenBank
S27Ec	*Epinephelus coioides* (Ec)	AAM27204	84	GenBank
S27Bb	*Branchiostoma belcheri tsingtaunese*(Bb)	AAN86980	84	GenBank
S27Ha	*Homarus americanus* (Ha)	P55833	84	SwissProt
S27-1Ai	*Argopecten irradians* (Ai)	AAN05598	84	GenBank
S27Lr	*Lumbricus rubellus* (Lr)	CAB58439	84	EMBL
S27Sc	*Saccharomyces cerevisiae* (Sc)	CAA81997	81*	EMBL
S27At	*Arabidopsis thaliana* (At)	AAM66954	84	EMBL
S27Cr	*Chlamydomonas reinhardtii* (Cr)	P47903	86	SwissProt

*The initial methionine is missing.

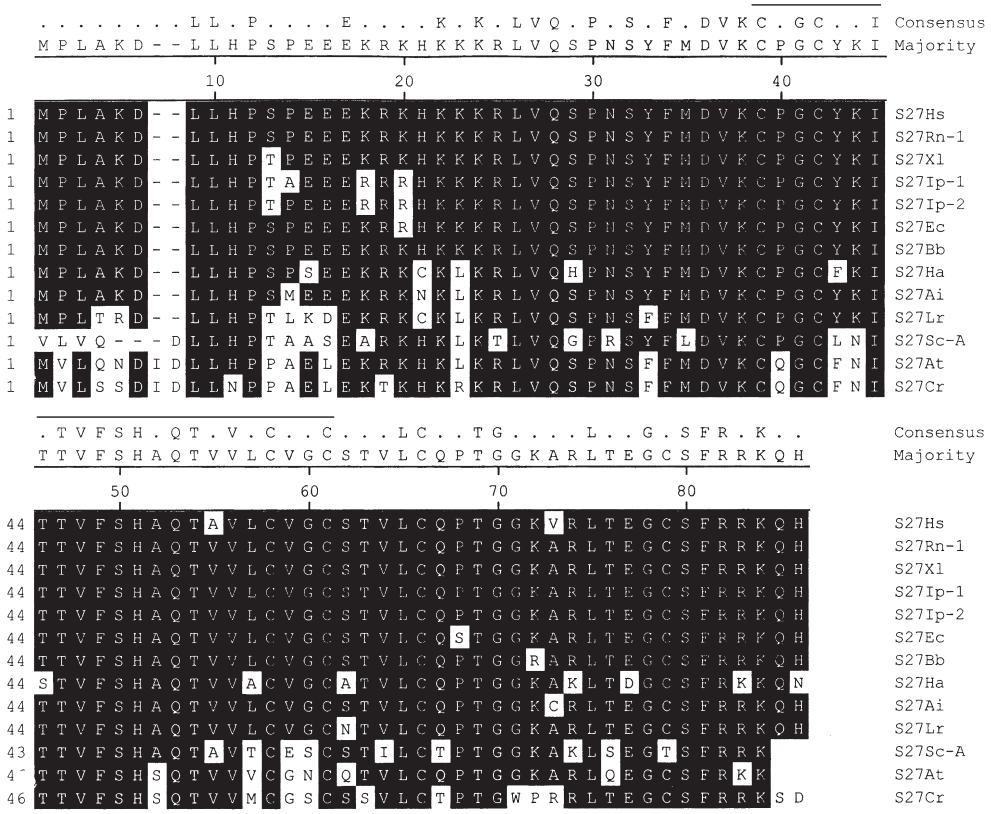

Figure 2 - Amino acid sequence alignment of representative S27 proteins using the MegAlign (DNASTAR) and CLUSTAL programs. The amino acid residues that match the consensus sequence are shaded solid black. Gaps introduced into sequences to optimize the alignments are represented by (-). The consensus sequence is provided above the multiple alignments and the zinc finger-like motif is underlined.

homologues in vertebrates (humans, *Xenopus* and fish), 84-94% with its invertebrate (annelids, mollusks and crustaceans) homologues and 69-72% with its counterparts from other eukaryotes (plants and yeast) (Figure 2). Our results show that the S27 protein is highly conserved among eukaryotes and that the *AmphiS27* gene appears to be more closely related to the gene responsible for vertebrate S27 protein than to the corresponding invertebrate gene.

Southern blotting demonstrated the presence of a single hybridization band for the *EcoR*, *Hind* an *BstX* restriction enzymes and two hybridization bands for the *Bgl* enzyme (Figure 3). All four restriction enzymes used did not digest *AmphiS27* cDNA strings. The fact that *Bgl* digestion yielded two bands may be due to either the presence of polymorphism or digestion within the intron region of the gene. The presence of introns in S27 genes has been found in humans (Gene ID 6232), rats (Gene ID 94266) and yeasts (Gene ID 853700) and it is highly likely that the amphioxus S27 gene also has one or more introns. The preponderance of a single hybridization band in the Southern blot analyses suggests that only one copy of the *AmphiS27* gene is present in the genome of amphioxus *B. belcheri tsingtauense*, agreeing with the observation that amphioxus retains a genome uncomplicated by extensive gene duplication (Holland and Garcia-Fernandez, 1996). It is also interesting to

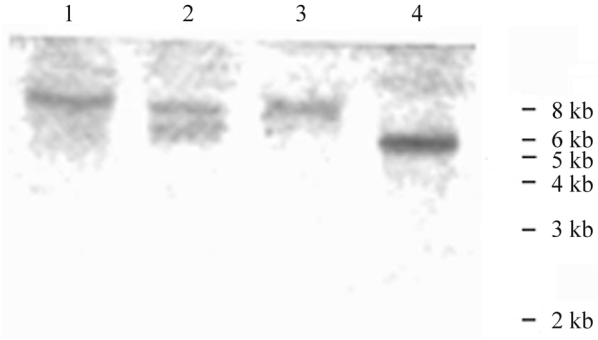

Figure 3 - Southern blotting analysis of genomic DNAs extracted from adult amphioxus *B. belcheri tsingtauense*. The digoxigenin (DIG)-labeled DNA probes correspond to the cDNA clone 094. Lane 1, *Bst*X; Lane 2, *Bgl*; Lane 3, *Hind*; Lane 4, *EcoR*.

note that there are 4 to 6 copies of the *S27* gene in rats (Chan *et al*, 1993) and 2 copies of this gene in lobster and mussel (Snyder, 1999). The presence of *S27* pseudogenes cannot be ruled out in lobster and mussel, therefore our results may not contradict the proposal that during the early evolution of the vertebrate lineage two rounds of extensive gene duplication took place, one close to the origin of the vertebrates and the other close to the origin of the jawed

vertebrates (Holland *et al*., 1994; Sharman and Holland, 1996; Sidow, 1996).

Acknowledgments

We thank our colleagues for their help in sample dissection. The work was supported by a grant from the Ministry of Science and Technology (MOST) of China (n. 2004AA626050) to S.C. Zhang.

References

Altschul SF, Madden TL, Schaffer AA, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl Acids Res 25:3389-3402.

Burland TG (2000) DNASTAR's lasergene sequence analysis software. Methods Mol Biol 132:71-91.

Chan YL, Suzuki K, Olvera J and Wool IG (1993) Zinc finger-like motifs in rat ribosomal proteins S27 and S29. Nucl Acids Res 21:649-55.

Gao J, Kim SR, Chung YY, Lee JM and An G (1994) Developmental and environmental regulation of two ribosomal protein genes in tobacco. Plant Mol Biol 25:761-770.

Hahn D and Kück U (1995) cDNA nucleotide sequences and expression of the genes encoding the cytoplasmic ribosomal proteins S18 and S27 from the green alga *Chlamydomonas reinhardtii*. Plant Sci 111:73-79.

Holland PWH, Garcia-Fernadez J, Williams NA and Holland ND (1994) Gene duplication and the origin of vertebrate development. Dev Suppl 125-133.

Holland RW, Garcia-Fernandez J (1996) Hox genes, developmental evolution and the origin of vertebrates. Ontogenez 27:273-279.

Kay MA & Jacobs-Lorena M (1987) Developmental genetics of ribosome synthesis in *Drosophila*. Trends Genet 3:347-351.

Liu Z, Zhang S, Yuan J, Sawant MS, Wei J and Xu A (2002) Molecular cloning and phylogenetic analysis of AmphiUbf80, a new member of ubiquitin family from the amphioxus *Branchiostoma belcheri tsingtauense*. Curr Sci 83:50-53.

Meyer A and Schartl M (1999) Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11:699-704.

Rehm BH (2001) Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. Appl Microbiol Biotechnol 57:579-592.

Revenkova E, Masson J, Koncz C, Afsar K, Jakovleva L and Paszkowski J (1999) Involvement of *Arabidopsis thaliana* ribosomal protein S27 in mRNA degradation triggered by genotoxic stress. EMBO J 18:490-4999.

Sankoff D (2001) Gene and genome duplication. Curr Opin Genet Dev 11:681-684.

Sharman AC and Holland PWH (1996) Conservation duplication and divergence of developmental genes during chorade evolution. Netherlands J Zool 46:47-67.

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrate. Curr Opin Genet Dev 6:715-722.

Snyder MJ (1999) Ribosomal protein S27E, P2 and L37A from marine invertebrates. Mar Biotechnol 1:184-190.

Steele LF and Jacobson A (1986) Ribosomal proteins are encoded by single copy genes in *Dictyostelium discoideum*. Gene 41:165-172.

Strokes MD and Holland ND (1998) The lancelet: Also known as "amphioxus", this curious creature has returned to the limelight as a player in the phylogenetic history of the vertebrates. Am Sci 86:552-560.

Thomas EA, Alvarez CE and Sutcliffe JG (2000) Evolutionarily distinct classes of S27 ribosomal protein with differential mRNA expression in rat hypothalamus. J Neurochem 74:2259-2267.

Wool IG, Chan YL and Glück A (1995) structure and evolution of mammalian ribosomal proteins. Biochem Cell Biol 73:933-947

Wool IG, Chan YL and Glück A (1996) Mammalian ribosomes: The structure and the evolution of the protein. In: Hershey JWB, Mathews MF and Sonenberg N (eds) Translation Control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 685-732.

Wool IG, Endo Y, Chan YL and Glück A (1990) Studies of the structure, function and evolution of mammalian ribosomes. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D and Warner JR (eds) Ribosome Structure, Function and Evolution. Society for Microbiology, Washington DC, pp 203-214.

Zhang SC, Yuan JD and Li HY (2001) Amphioxus-model animal for insights into the origin and evolution of the vertebrates. Chin Bull Life Sci 13:214-218 (in Chinese with English abstract).

Associate Editor: Horácio Schneider