Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 314 (2003) 89–102
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Shadoo, a new protein highly conserved from fish to mammals and with similarity to prion protein

Marko Premzl[a,1], Lorenzo Sangiorgio[b,1], Bice Strumbo[b], Jennifer A. Marshall Graves[c], Tatjana Simonic[b], Jill E. Gready[a,*]

[a] *Computational Proteomics and Therapy Design Group, Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra ACT 2601, Australia*
[b] *Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Sezione di Biochimica e Fisiologia Veterinaria, Università di Milano, via Celoria 10, 20133 Milan, Italy*
[c] *Comparative Genomics Group, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra ACT 2601, Australia*

Received 30 December 2002; received in revised form 7 May 2003; accepted 16 May 2003

Received by S. Salzberg

Abstract

We report evidence from cDNA isolation and expression analysis as well as analyses of genome, expressed sequence tag (EST), cDNA and expression databases for a new gene named *SPRN* (shadow of prion protein). *SPRN* comprises two exons, with the open reading frame (ORF) contained within exon 2, and codes for a protein of 130–150 amino acids named Shadoo (Japanese shadow), predicted to be extracellular and GPI-anchored. The *SPRN* gene was found in fish (zebrafish, *Fugu*) and mammals (mouse, rat, human). Conservation of order and transcription orientation of two proximal genes between fishes and mammals strongly indicates gene orthology. Sequence comparison shows: a highly conserved N-terminal signal sequence; Arg-rich basic region containing up to six tetrarepeats of consensus XXRG (where X is G, A or S); a hydrophobic region of 20 residues with strong homology to prion protein (PrP); a less conserved C-terminal domain containing a conserved glycosylation motif; and a C-terminal peptide predicted to be a signal sequence for glycophosphotidylinositol (GPI)-anchor attachment. Fish Shadoos (Sho) show well conserved sequences (identity 54%) over 106 amino acids (zebrafish length), and conservation among the mammalian sequences is very high (identity 81–96%). The fish and mammalian sequences are also well conserved, particularly for zebrafish, to beyond the end of the hydrophobic sequence (identity 41–53%, 78 amino acids, zebrafish length). The overall structure appears closely related to PrPs, although the C-terminal domains of Shos are quite different from those of PrPs, for which conformational changes in mammals are implicated in disease. The structural similarity is particularly interesting given recent reports of three new genes with similarities to PrPs found in *Fugu* (PrP-like, PrP-461/stPrP-1 and stPrP-2) and other fish, but for which direct evolution to higher vertebrate PrPs is unlikely and for which no other mammalian homologues have been found. Database information indicates expression of *SPRN* in embryo, brain and retina of mouse and rat, hippocampus of human, and in embryo and retina of zebrafish, and we directly confirmed a strikingly specific expression of the mammalian (human, mouse, rat) transcripts in whole brain. This result together with some common structural features led to the suggestive hypothesis of a possible functional link between mammalian PrP and Sho proteins.

Keywords: Comparative genomics; Conserved contiguity; Repeats; Gene prediction; Zebrafish; DNA

1. Introduction

Comparative genomics is a powerful method for identification of new genes, allowing initial characterization of their distribution among the kingdoms of life, and tracking of the sequence and regulatory changes underpinning their conserved or diverged functions (Hedges and Kumar, 2002). Application of the method is now enormously facilitated by the ability to rapidly search for and integrate genome,

Abbreviations: AdT, anchored oligo-dT; Dpl, Doppel protein; EST, expressed sequence tag; GPI-anchor, glycophosphotidylinositol-anchor; l/c, lower case; ORF, open reading frame; nr, non-redundant; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR; PrP, prion protein; *PRND*, gene for Dpl; *PRNP*, gene for PrP; PrP-like, fish orthologue of PrP; Sho, Shadoo protein; *SPRN*, gene for Sho; TCs, tentative consensus sequences; u/c, upper case.

* Corresponding author. Tel.: +61-2-6125-8304; fax: +61-2-6125-0415.
E-mail address: Jill.Gready@anu.edu.au (J.E. Gready).

[1] These two authors contributed equally.

0378-1119/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0378-1119(03)00707-8

expressed sequence tag (EST), cDNA and expression information from databases; based on analysis of the data, deductions can be made on gene structure, location and regulation, and on translated protein sequences, including cellular location and possible function.

We report an application of this approach which, starting from the initial lead of the sequence of a zebrafish cDNA we isolated, enabled us to discover other homologous proteins in fish and mammals, and to start to characterize the gene and protein using the substantial information which already exists in the databases. Striking sequence similarities of this protein to prion protein (PrP), as well as its expression almost entirely specific to brain, allow the hypothesis that it may be functionally related to PrP and important for understanding prion disease (Prusiner, 1998). We propose to call this new gene *SPRN* (shadow of PrP) and the protein Shadoo (Japanese shadow).

PrP is the remarkable protein associated with lethal neurodegenerative diseases grouped as transmissible spongiform encephalopathies (Collinge, 2001). Although normally expressed in many tissues other than brain, recent work indicates expression at high levels only in discrete subpopulations of cells, particularly nerve and immune cells of the neuroimmune, neuroendocrine and peripheral nervous systems (Ford et al., 2002). However, its function is still unclear. Although the first cDNA encoding mammalian PrP was cloned and characterized in hamster, and immediately confirmed in mouse and human, almost 20 years ago (Oesch et al., 1985), it was more difficult to find homologues in birds (Harris et al., 1991), reptiles (Simonic et al., 2000) and amphibians (Strumbo et al., 2001). Identification of homologues using conventional experimental methods has been difficult because despite the conservation of all, or almost all, the structural motifs found in mammalian PrPs (Fig. 1), PrPs from different vertebrate classes exhibit remarkable differences in their primary sequences.

As shown in Fig. 1, the PrP sequence is very unusual in having several regions with distinct amino acid composition, and potential 3-D structure and contributions to function, within a quite short protein of ~250 residues (mammals) encoded by a single exon (Lee et al., 1998). Both conserved and variable features between amphibians and mammals are apparent. Most notably, PrPs show extremely high conservation in the middle hydrophobic sequence (although this is shorter in frog PrP), in the presence of one disulfide bond and two *N*-glycosylation

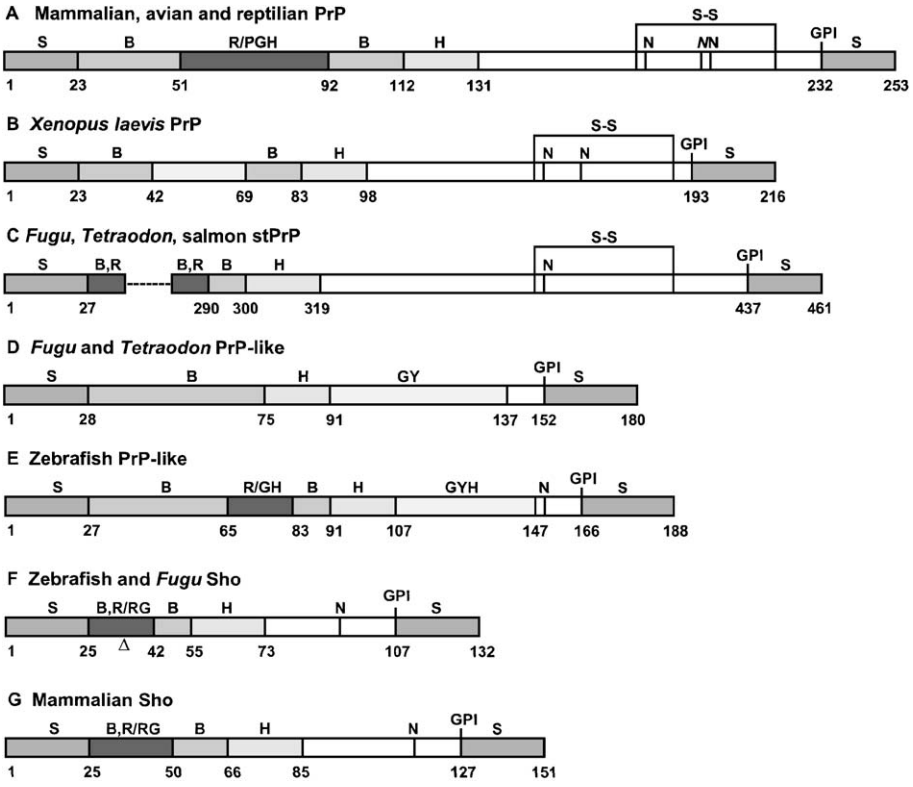

Fig. 1. Overall structures of PrP, stPrP, PrP-like and Sho proteins showing: S, signal sequence; B, basic region; H, hydrophobic region; R/PGH, PGH-rich repeats; R/GH, GH-rich repeats; B,R/RG, RG-rich basic repeats; B,R, basic repeats; N, *N*-glycosylation site; S—S, disulfide bridge; GPI, glycophosphatidylinositol anchor; GY and GYH, GY- and GYH-rich regions. Regions and attachment positions are approximately to scale. Numbers indicate the first residue of each section, and last one of each protein. (A) Mammalian, avian and reptilian PrPs; numbers refer to human; additional N site for avian in italics. (B) *X. laevis* PrP. (C) *Fugu*, *Tetraodon* and salmon stPrP; numbers for PrP-461 (Rivera-Milla et al., 2003). (D) *Fugu* and *Tetraodon* PrP-like; numbers refer to *Fugu*. (E) Zebrafish PrP-like. (F) Zebrafish and *Fugu* Sho; the arrow indicates insertion region in *Fugu*; numbers refer to zebrafish. (G) Mammalian Sho; numbers refer to human.

Table 1
Summary of database information for *SPRN* gene and Sho protein in fish and mammals

Species	DB	(A) Protein ID[a]	(B) Gene ID[a]	(C) Transcript ID[a]	(D) Clone ID[b]	(E) Genomic location	(F) Library
Mouse	Ensembl (v9.3a.1)	ENSMUSESTP 00000026614, 7.130000001-131000000. 362987.364892 (*g*), C7002492 (*f*), chr7.131.012.a (*t*), Mm7_WIFeb01_181_17_6.1 (*gs*)	ENSMUSESTG 00000026976	ENSMUSESTT 00000026614, 7.130000001-131000000.362987.364892 (*g*), chr7.131.012.a (*t*), Mm7_WIFeb01_181_17_5.1 (*gs*)	7.130000001-131000000	Chr.7 F5: 130360737-130364466	–
	NCBI	XP_150033	LOC212518	XM_150033	NW_000335	Chr.7 F4	–
	FANTOM	PC31542	–	TF31542	C630041J07	–	–
	DDBJ	AK049995-1	–	AK049995	C630041J07	–	–
	TIGR	–	–	TC613820	*BB653489, *BM944750	Chr. 7: 130.362.897-130.364.466	NIH_BMAP_EH0p, RIKEN full-length enriched, adult male hippocampus
	TIGR	–	–	TC613821	*BB284188, *AI842512	Chr. 7: 130.362.273-130.362.792	NIH_BMAP_MHI_N, RIKEN full-length enriched, adult male hippocampus
Human	Ensembl (v9.3a.1)	AL 161645.14.1. 161644.7867.9560 (*g*), C10001717 (*f*)	–	AL161645.14.1.161644.7867.9560 (*g*)	AL161645	Chr. 10 q26.3: 134150994-134151449 Chr. 10 q26.3	–
	NCBI	–	–	BC040198	NT_017795	–	–
Rat	Ensembl (v12.2.1)	ENSRNOP 00000025609, C1003566 (*f*)	ENSRNOG 00000018927	ENSRNOT 00000025609	RNOR01010413	Chr.1 q36: 199692648-199694883	–
	NCBI	–	–	–	NW_043400	Chr. 1	–
	TIGR	–	–	–	*BF391059	–	–
	TIGR	–	–	TC297203	*AW530092, *BF564096, *BF285504	–	UI-R-C4, rat gene index, normalized rat brain
	TIGR	–	–	TC293056	*BF285497, *AA943106, *BF404416, *AI007831, *BG376848, *AI101223, *BF409274, *BF409392, *BF409521	–	UI-R-CA1, normalized rat brain
Fugu	Ensembl (v12.2.1)	SINFRUP 00000151627 (*g*)	SINFRUG 00000142820 (*g*)	SINFRUT 00000151627 (*g*)	scaffold_28	Chr_scaffold_28: 392510-392800	–
	HGMP	–	–	–	scaffold_28	–	–
Tetraodon	Genoscope	–	–	–	FS_CONTIG_4144_1	–	–
Zebrafish	Sanger (Assembly6)	–	–	–	z06s038879	–	–
	Ensembl (Trace)	–	–	–	zfishC-a2446d07.g1c	–	–
	Ensembl (Zv2)	–	–	–	ctg9556.1	–	–
	NCBI	–	–	–	*AL913623, *AL913622, *AL913625, *BG738569, *AL913624	–	PJR-Z1+Z2, zebrafish adult retina cDNA
	EMBL	–	–	(AJ490525)	–	–	–

[a] Gene-prediction derived protein sequences and corresponding predicted transcripts are labelled according to the program: (*g*), Genscan; (*f*), Fgenesh++; (*t*), Twinscan; (*gs*), Genomescan.
[b] ESTs are labelled by *.

sites in the C-terminal domain, in the presence of an N-terminal basic region, and in N- and C-terminal signal sequences for extracellular export and glycophosphotidylinositol (GPI)-anchor attachment. However, the N-terminal repeat region is highly variable in repeat length and sequence among eutherian and marsupial mammal and avian/turtle PrPs (Windl et al., 1995; Wopfner et al., 1999; Simonic et al., 2000), and is entirely absent in frog PrP (Strumbo et al., 2001). Solution NMR structures for mammalian PrPs show only the C-terminal domain (from residue ~126) to be folded, with the whole N-terminal region (23–125) being flexibly disordered (Zahn et al., 2000).

As suggested by the amphibian PrP structure, this protein appears to have gained new structural features during evolution (Fig. 1), and, therefore, potentially new functions. The recent reports of, at least, three new genes in fish with similarities to PrP are, therefore, of great interest. Suzuki et al. (2002) reported isolation and characterization of a cDNA coding for a protein of 180 amino acids in *Fugu rubripes*. This sequence was called PrP-like, based mainly on conservation of the hydrophobic sequence (albeit four residues shorter), and other features in common with PrP, including its basic nature, N-terminal signal sequence, possible GPI-anchor attachment, and single-exon coding region. However, overall sequence homology between PrPs and *Fugu* PrP-like (and a similar database sequence for *Tetraodon*) is low, with the latter lacking repeats, disulfide and glycosylation sites, and having a different C-terminal domain. Nonetheless, orthology was suggested between *Fugu* PrP-like and human PrP by conservation of adjacent genes (Suzuki et al., 2002). The fish gene region did not, however, contain the Doppel gene *PRND*, which is reported so far only in mammals and is adjacent to *PRNP*; *PRND* and *PRNP* are regarded as paralogues arising from gene duplication (Moore et al., 1999). Suzuki et al. (2002) also reported a (database) PrP-like sequence for *Danio rerio* (zebrafish) which although homologous by sequence alignment to *Fugu* PrP-like (not shown) shows significant divergence (see Fig. 1); as the genomic context of this sequence has yet to be elucidated it is unclear whether the *Fugu/Tetraodon* and *Danio* PrP-like genes are true homologues.

More recently, two other genes with similar structural features to PrPs of higher vertebrates (amphibians to mammals) have been reported in fish (Rivera-Milla et al., 2003; Oidtmann et al., 2003). The cDNA of a *F. rubripes* gene (Genbank AF531159) reported by Rivera-Milla et al. (2003) encoded a protein of 461 amino acids (PrP-461) with a greatly expanded repeat region, but apparently similar C-terminal domain to those in PrPs, including disulfide bridge and one of the conserved *N*-glycosylation sites. A homologous *F. rubripes* cDNA (AY141106) of slightly different length (450 residues; stPrP-1) as well as that for Atlantic salmon *Salmo salar* (AY141107) with an even more expanded repeat region (605 residues) were obtained by Oidtmann et al. (2003), while the *Tetraodon* homologue was reported by both groups from genomic data. This stPrP-1 gene in *Fugu* is at a different genomic location from the PrP-like gene, while the locations of the *Tetraodon* and salmon genes are undetermined. However, Oidtmann et al. (2003) reported the cDNA for a second *Fugu* gene, stPrP-2 (AY188583), closely related to stPrP-1 but with a hydrophobic region disrupted by charged residues, which was located adjacent to PrP-like. These findings provide a puzzle as while in terms of structural similarity stPrP-1 is closest to PrPs, in terms of genomic localization stPrP-2 and PrP-like seem more closely linked to PrPs (Oidtmann et al., 2003).

During our own searches for fish PrP candidates in EST databases, we identified an interesting zebrafish sequence fragment, and now report the isolation and cloning of the cDNA. This sequence encodes a novel protein, different from the PrP-like, stPrP-1 and stPrP-2 sequences, but again with intriguing similarities to PrPs, as summarized in Fig. 1. Database searches indicate that this protein sequence is well conserved between fish (zebrafish, *Fugu*) and mammals (human, mouse, rat). Where available, data for chromosomal location, adjacent genes, gene structure and overall protein structure and sequence all show high homology (see Table 1). Extending expression data already in the databases, we also demonstrate expression of the corresponding mammalian mRNAs, which is almost entirely in brain.

2. Materials and methods

2.1. RT-PCR and PCR assays

Oligonucleotides were purchased from MWG-Biotech (Germany). Total RNA was extracted from zebrafish head and hepatopancreas, from mouse and human adult brain and from several rat tissues (Chomczynsky and Sacchi, 1987), genomic DNA from zebrafish hepatopancreas following the standard procedure (Sambrook et al., 1989), and plasmid DNA using QIAprep Spin Miniprep Kit (QIAGEN, Germany). Reverse transcriptions were primed by random primers, anchored oligo-dT primer (AdT) or fish specific primers, depending on experimental purposes. One to five micrograms of total RNA was reverse transcribed using Superscript II Reverse Transcription system (Life Technologies, UK), following the manufacturer's instructions. Polymerase chain reactions (PCRs) were performed on 20–200 ng of genomic DNA, on 1–10 pg of plasmid DNA or on reverse transcribed RNA, using Platinum *Taq* DNA Polymerase system (Life Technologies), following the manufacturer's protocol. PCR conditions were 30 s at 94 °C, 30 s at 47–56 °C, and 30 s–3 min at 72 °C, depending on the expected size of fragments, for 31–33 cycles. The 3′ end of zebrafish cDNA was isolated using 3′ RACE as described previously (Simonic et al., 1997). PCR products were cloned into *Sma* I linearized pUC9, modified as a T-vector

(Marchuk et al., 1991) and *E. coli* XL1-Blue competent cells were transformed with the recombinant plasmids. The 5′ end of zebrafish cDNA was obtained by semi-nested PCR on a cDNA library, using a forward primer annealing on vector and specific reverse primers. The adult zebrafish retina cDNA library was kindly provided by Dr. G.A. Niemi from Howard Hughes Medical Institute (University of Washington, Seattle). To verify expression of mammalian mRNAs, reverse transcriptase (RT)-PCR assays were performed using random and AdT primed reverse transcripts as template. Mouse, rat and human cDNAs were amplified as follows: 30 s at 94 °C, 30 s at 51–55 °C, and 30–40 s at 72 °C. The reaction products were separated on agarose gel 2% and stained with ethidium bromide. Gels were scanned using a Typhoon 9200 imager (Amersham Pharmacia Biotech, UK) with Typhoon Scanner Control software.

2.2. Sequencing and computer-assisted analyses

Sequence analysis of DNA fragments was performed on both strands, either on plasmids or directly on purified PCR products by the *Taq* Big Dye di-deoxy terminator method, using an automated 3100 DNA sequencer (Applera, USA). DNA sequences were analyzed using MT Navigator PPC software (release 1.0.2b3). The cDNA sequence (named Shadoo) has been deposited in the EMBL database under accession number AJ490525 (see Table 1).

2.3. Databases, searches and search tools

Several public databases were searched in attempts to detect homologues of zebrafish Sho/*SPRN*: DDBJ (http://www.ddbj.nig.ac.jp/); Ensembl (http://www.ensembl.org/); FANTOM (http://www.gsc.riken.go.jp/e/FANTOM/); FlyBase (http://www.flybase.org/); Genoscope (http://www.genoscope.cns.fr/); HGMP (http://fugu.hgmp.mrc.ac.uk/); M Base (http://mbase.bioweb.ne.jp/~dclust/medaka_top.html); NCBI (http://www.ncbi.nlm.nih.gov/); Sanger Institute (http://www.sanger.ac.uk/ and local BLAST server for Sanger Institute zebrafish database on http://danio.mgh.harvard.edu/blast/blast_grp.html); TIGR (http://www.tigr.org/) and WormBase (http://www.wormbase.org/). Data found, with identifier codes, are collected in Table 1. The following subsections summarize the order of the searches, and search sequences used, starting from our primary data for the zebrafish Sho cDNA; column A to F reference the data in Table 1.

2.3.1. Identification of mouse Sho

The mouse Sho protein XP_150033 (column A), and accordingly its gene (column B), transcript (column C), genomic clone (column D) and chromosomal coordinates (column E) were detected in the NCBI non-redundant (nr) protein database using fish Sho information (amino acid residues 25 to 54) as the search query and the server's BLASTP (Altschul et al., 1997) search program with short-nearly-exact matches option. In the Ensembl mouse genome database (v9.3a.1), mouse Sho protein (column A) was then detected in the Genscan (Burge and Karlin, 1997) predicted peptides database using the server's BLASTP tool (Altschul et al., 1990, and Gish, 1994–1997, unpublished) and the XP_150033 amino acid sequence as search query. Information on the gene (column B), EST (column C), genomic clone (column D), chromosomal position (column E) and gene predictions (column A) by the Twinscan (Korf et al., 2001), Fgenesh++ (Solovyev, 2001) and Genomescan (Burge and Karlin, 1997) programs was also found with the interactive Ensembl web service. In the new release of the database (v12.3.1), only the gene predictions are present. The cDNA clone (column D) encoding mouse Sho (columns A and C) was detected in the FANTOM database using the mouse Sho ORF nucleotide sequence extracted from XM_150033 as the search query and the server's BLASTN (Altschul et al., 1997) program. This cDNA clone (columns A, C and D) was found in the DDBJ database in the same manner, as well as the contigs (columns C and D), genomic location (column E) and expression data (column F) in the TIGR database.

2.3.2. Identification of human Sho

In the Ensembl human genome database (v9.3a.1), we were able to detect *SPRN* Genscan and Fgenesh++ predictions (column A) in the genomic position (columns D and E) analogous to that of mouse *SPRN*. However, in the new release of the database (v12.31.1) only the Fgenesh++ prediction is present, and the chromosomal coordinates have changed to Chr.10 q26.3: 134364175-134368086. The putative *SPRN* genomic location in the NCBI human genome database (column D) was in turn located after keyword search and detection of the annotated adjacent gene encoding ECHS1 (see also Fig. 5). The human Sho transcript (column C) was found in the NCBI nr database using the mouse ORF sequence from XM_150033 and the server's BLASTN program.

2.3.3. Identification of rat Sho

The rat Sho contigs (column C) were detected in the TIGR database comparative genome browser in which genomic contexts of the rat TCs are aligned with their corresponding genomic contexts from the mouse (TC613820 and TC613821), which were already known and located. TCs are tentative consensus sequences, virtual transcripts created by merging the EST data, and may be full or partial cDNA length. An additional EST BF391059 (column D) and expression data (column F) were also present in the TIGR database. The protein sequence presented in Section 3 and Fig. 3 was derived by conceptual translation of the TIGR contigs. Genomic clones (column D) harbouring the rat *SPRN* were located in the NCBI and Ensembl databases by using the BF391059 nucleotide sequence as query sequence with the servers' BLASTN programs. In addition, the protein

(column A), gene (column B), transcript (column C), and chromosomal location (column E) information were found in the updated Ensembl rat database (v12.2.1).

2.3.4. Identification of Fugu Sho

The zebrafish Sho amino acid sequence was used as query in BLASTP analysis of Ensembl's (v12.2.1) *F. rubripes* Genscan peptides database. The prediction of *Fugu* Sho (column A), *SPRN* (column B) and transcript (column C) correspond to the genomic fragment Chr_scaffold_28 (column D). The same genomic scaffold was detected by using SINFRUT00000151627 as query sequence and the BLASTN (Altschul et al., 1997) tool on the HGMP web server. The *Fugu* Sho amino acid sequence presented in Section 3 and Fig. 3 was derived by conceptual translation of scaffold_28, which contains the complete ORF.

2.3.5. Identification of zebrafish genomic clones and ESTs

The genomic sequence assembly z06s038879 (column D) harbouring the zebrafish *SPRN* was found in the Sanger Institute zebrafish genome data (Assembly 6) by using our zebrafish cDNA sequence as search query with BLASTN (Altschul et al., 1997) on the local server at the Massachusetts General Hospital Renal Unit. The sequence read zfishC-a2446d07.g1c was found in the Ensembl trace repository using the zebrafish nucleotide cDNA sequence as query sequence and the available web service SSAHA (Ning et al., 2001) search tool. ESTs (column D) covering parts of the zebrafish Sho cDNA were detected using its nucleotide sequence as query with BLASTN search of the NCBI est_others database. A BLAST search of the recent Ensembl whole genome shotgun assembly Zv2 using, again, the same cDNA sequence as search query identified a genomic contig ctg9556.1 on Chromosome fragment assembly_203 that contained the whole *SPRN* gene.

2.3.6. Identification of Sho in Tetraodon

The *Fugu* Sho ORF was used as query sequence to search the Genoscope *Tetraodon* whole genome shotgun database with the Genoscope BLASTN server. Although the genomic clone FS_CONTIG_4144_1 harbours the *Tetraodon SPRN* ORF in its whole length, conceptual translation of the nucleotide sequence was impossible due to the poor quality of the sequence data.

2.3.7. Data from other species

Searches for Sho/*SPRN* homologues in other species by analyzing available data in the Flybase, M Base, NCBI (including *Xenopus laevis* and other vertebrates), Sanger Institute (*X. tropicalis*) and WormBase produced no results. However, the *Xenopus* DBs are still very incomplete; it seems very likely that Sho/*SPRN* will be found to be present in animals placed evolutionarily between fishes and mammals.

2.4. Analysis software

The Lasergene software package (DNASTAR, Madison WI, USA) was used for basic handling and analyses of the nucleotide sequence and protein data. Exon–intron structures of *SPRN* genes were determined using the Local Pair Alignment tool (Huang and Miller, 1991) on the ANGIS interactive web interface Biomanager (http://www.angis.org.au). The zebrafish genomic sequence was annotated using the NIX interactive web tool (Williams et al., 1998, http://www.hgmp.mrc.ac.uk/NIX/). The programs used for prediction of protein signal-peptide cleavage sites (SignalP, Nielsen et al., 1997, http://www.cbs.dtu.dk/services/SignalP/) and GPI-anchor addition sites (bigPI-predictor, Eisenhaber et al., 1999, http://mendel.imp.univie.ac.at/sat/gpi/gpi_server.html) are available as free web services. Multiple sequence alignment using the algorithm of Taylor (1990), and manual editing, was done using Cameleon v3.14 (Oxford Molecular 1995, now owned by Accelrys). Alignment figures were prepared for publication using CHROMA (Goodstadt and Ponting, 2001) available free on the web (http://www.lg.ndirect.co.uk/chroma). Other figures were drawn using PowerPoint.

3. Results

3.1. Isolation of Zebrafish Sho protein cDNA and features of the encoded protein

Repeated screenings of EST databases aimed at finding fish sequences coding for proteins with structural motifs peculiar to known PrPs were done. Whole protein and/or selected regions of the amino acid sequences of known PrPs from different vertebrate species were used as protein queries for translated BLAST searches. Results were analyzed by translating the whole cDNA fragments corresponding to promising amino acid sequences exhibiting some similarities to the query. From a number of unsatisfactory outcomes, a single interesting zebrafish clone (GenBank/EMBL/DDBJ accession number BG738569; Washington University Zebrafish EST Project) was identified. This DNA fragment spanned 407 bp and comprised a region coding for the N-terminal portion of an unknown protein. A second upstream ATG (−61) could be the start of a shorter ORF, but other data discussed later cast doubt on its significance. Four new cDNA fragments recently deposited in the EST database (accession numbers AL913622, AL913623, AL913624, AL913625) (see Table 1) do not extend the sequence we report here.

In order to verify the EST data, a control DNA fragment was amplified on a standard reverse transcript using primers designed from EST clone BG738569 sequence information. Numbering of cDNA fragments and of the deduced protein refers to Fig. 2. A first PCR performed with primers ZF1 (GAATGAACAGGGCAGTCGC) and ZR1 (CCGTTCTGACCTGTCGTC), followed by a seminested step

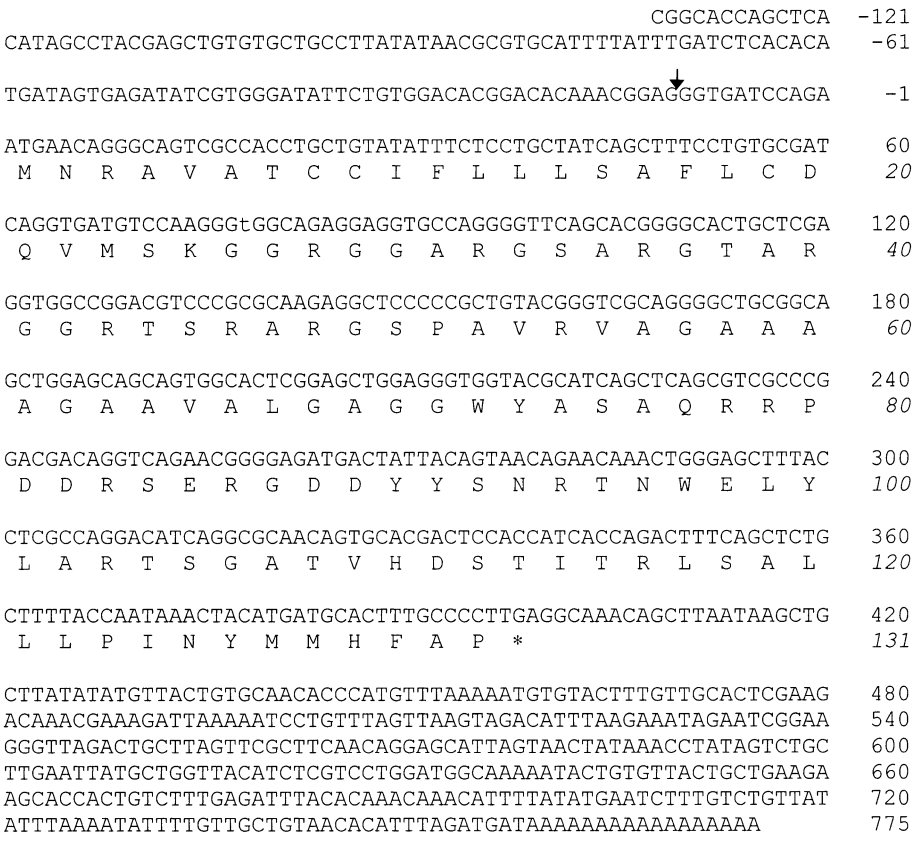

Fig. 2. Nucleotide and deduced amino acid sequences of the cDNA coding for zebrafish Shadoo. The first base of the translation start codon is designated +1 and the stop codon is marked by an asterisk. The missing base (T78) in EST BG73856 is shown as lower case. The arrow indicates the intron location in the corresponding gene.

with the inner reverse primer ZR2 (CGAGTGCCACTGCTGCTC) gave a 204 bp product (positions −2 to +202) identical to the known sequence except for a T insertion at position 78, which was repeatedly verified by independent experiments and is now also confirmed by EST AL913625. Due to its localization within the coding region, this change shifted the translation reading frame.

Subsequently, 3′ RACE allowed isolation of the 3′ region of this cDNA using an anchored template for two PCR assays, the second seminested, with an anchor primer coupled in turn with the forward primers ZF1 and ZF2 (CGAGTGCCACTGCTGCTC). The resulting DNA fragment spanned 591 bp, overlapped the 3′ end of the EST BG738569 sequence and contained the unknown 3′ portion of cDNA (positions 185–775). Finally, a zebrafish retina cDNA library has been used as template to extend the 5′ end by PCR using a vector anchor primer coupled with zebrafish specific primers. Three subsequent reactions, the latter two seminested, using in turn primers ZR1, ZR2 and ZR3 (GAATGAACAGGGCAGTCGC), were carried out to obtain products that were sequenced after cloning into pUC9. All these products, of 89, 94 and 110 bp in size, belonged to this same cDNA and confirmed the available sequence, the longest (positions −133 to −25) extending the 5′ end of the EST BG738569 clone by 27 bp.

The Zebrafish Sho cDNA spans 908 nt and comprises a 5′ UTR of 133 nt (positions −133 to −1), a coding sequence of 399 nt (positions 1–399) including the stop codon TGA and a 3′ UTR of 376 nt (positions 400–775). The 17 adenylates present at the end of this cDNA do not seem to belong to a polyA tail, because they have been also found in a zebrafish genomic sequence assembly (z06s038879) matching most parts of this cDNA (see Table 1). The coding region and the 3′ UTR are uninterrupted, while an intron of 4233 bp is recognizable by aligning cDNA and genomic contig ctg9665.1 before the translation start, between positions −12 and −11 of the 5′ UTR.

The encoded product comprises 132 amino acids, has a calculated molecular mass of 13.9 kDa and a theoretical pI of 10.51. Ala (17.4%), Gly (12.1%) and Arg (12.1%) are the most represented residues. The overall structure of this zebrafish protein will be discussed in the next section, together with that of its close homologue in *Fugu* and very similar mammalian homologues, but we note that the zebrafish Sho is the shortest of these.

3.2. Discovery of homologues of zebrafish Sho

Using the zebrafish cDNA and deduced ORF protein sequences as probes, the publicly available sequence data-

bases were methodically searched; full details are given in Section 2.3 and the results are shown in Table 1. Much database evidence supports the existence of mouse Sho protein: transcripts and ESTs from the NCBI, Ensembl, FANTOM and TIGR databases as well as gene prediction by the Genscan, Fgenesh++, Twinscan and Genomescan programs. Moreover, its expression has been shown experimentally in several cDNA libraries. The cDNA and its expression data from the NCBI nr database, and gene predictions using Ensembl's Genscan and Fgenesh++ programs, provide database support for human Sho. Gene prediction from the Ensembl DB and ESTs and expression data from the TIGR database are available for rat Sho. *Fugu* Sho was predicted by the Genscan program on Ensembl's server. The *SPRN* gene from *Tetraodon* is found in the genomic sequence from the Genoscope database. Expression data are available for zebrafish Sho in the NCBI est_others database, and the gene was predicted by Genefinder and Genemark (Borodovsky and McIninch, 1993) with NIX analysis of the genomic clone sequence.

3.3. Deduced protein sequences

Where no protein sequence data were available directly, the protein sequences of the Sho proteins were deduced from either genomic or EST nucleotide sequence data. Specifically, the rat Sho amino acid sequence was deduced from the TC297203 (translation for residues 1–27) and TC293056 (translation for residues 28–147) TIGR contigs. The *Fugu* Sho amino acid sequence was derived by conceptual translation of the genomic scaffold_28 (ORF is encoded by the nucleotide sequence 392342–392800 bp).

3.4. Sequence alignment

The amino acid sequence of our zebrafish Sho was aligned (Fig. 3) with the homologous protein sequences from mouse (XP_150033), human (Genscan prediction AL 161645.14.1.161644.7867.9560), rat (derived by conceptual translation, see Section 3.3) and *Fugu* (deduced from the genomic data, see Section 3.3). Alignment was done with the Taylor (1990) algorithm within the program Cameleon.

3.5. Prediction of signal sequences

Proximal signal peptide cleavage sites were predicted using the SignalP program (Nielsen et al., 1997). These predictions are consistent for all Sho proteins (see Fig. 3), indicating cleavage before the first (conserved) Lys residue and strongly supporting its extracellular location both in fishes and mammals. BigPI (Eisenhaber et al., 1999) predicts GPI-anchor attachment for mammalian Sho proteins (human, mouse and rat at G126, G122 and G122, respectively) at a similar confidence level to predictions for bird, turtle and frog PrPs (Simonic et al., 2000; Strumbo et al., 2001); nevertheless, experimental evidence has been reported only for chicken PrP (Harris et al., 1991). Anchor attachment is more strongly predicted for mammalian PrPs and Doppel, and is confirmed experimentally (Silverman et al., 2000). Although anchor prediction for fish Sho proteins (*Fugu* at G121 and zebrafish at G106) is weaker, their strong overall similarity with the mammalian Sho sequences, as well as local conservation around the abovementioned glycines, (see Fig. 3) suggests they might be GPI-anchored also.

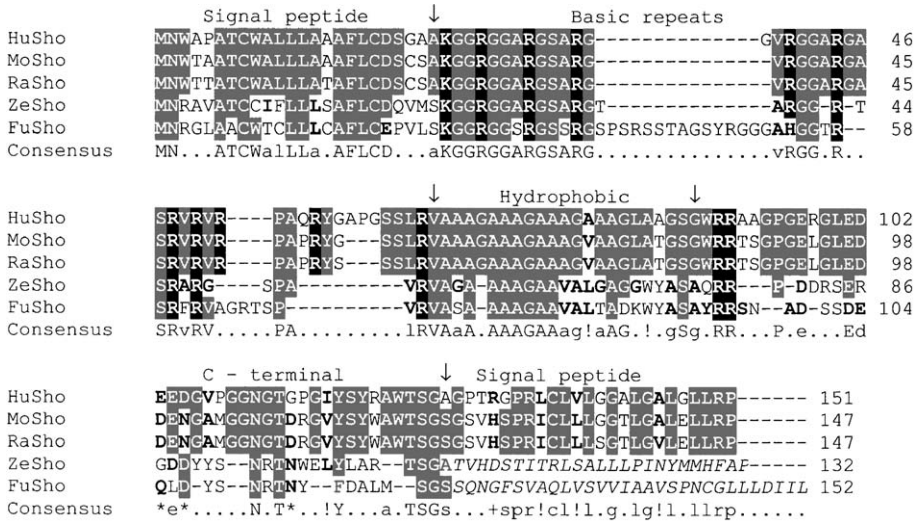

Fig. 3. Alignment of Sho proteins for fish and mammals. Arrows show cleavage sites and beginning and end of hydrophobic segment. Reversed black, conserved basic; reversed grey, conserved in at least three species; bold letters, conservatively conserved. In consensus line: capital letters, conserved in four or more species; lower case letters, conservatively conserved in four or more species; !, conserved hydrophobic; *, conserved polar; +, conserved basic. C-terminal signal sequence for ZeSho and FuSho in italics has not been aligned. Hu, human; Mo, mouse; Ra, rat; Ze, zebrafish; Fu, *Fugu*.

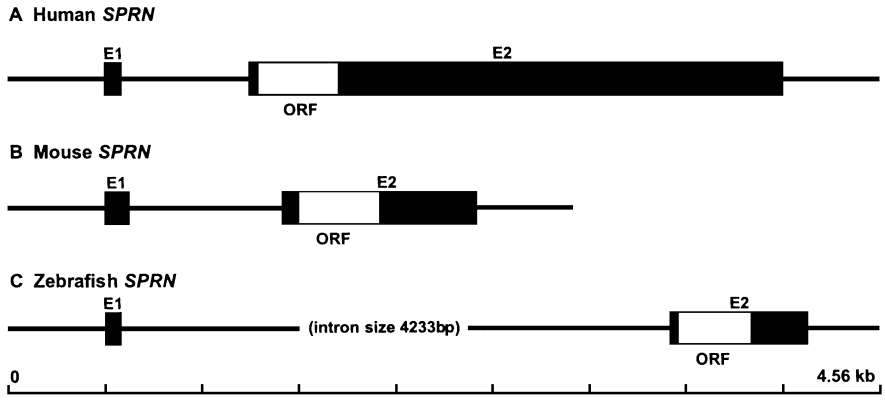

Fig. 4. Intron–exon structure for human, mouse and zebrafish *SPRN* genes. Figure is to scale as shown by the ruler, except the very long intron in zebrafish has been abbreviated. E1, exon 1; E2, exon 2; ORF, open reading frame.

3.6. Gene structure

SPRN gene structure was determined for species (zebrafish, human and mouse) in which both cDNA and genomic sequence information is available; these are shown in Fig. 4. Our zebrafish Sho cDNA (908 bp) was aligned to the genomic sequence downloaded from the Ensembl Zv2 pre-assembly database (ctg9556.1). The zebrafish *SPRN* gene has two exons: the proximal part of the cDNA sequence from 1 to 122 bp comprises exon 1 and the rest of the cDNA (nucleotides 123–908 bp; 786 bp) comprises exon 2. The intron size is 4233 bp. Nucleotide sequences are consistent with known consensus sequences at exon–intron boundaries: GAGgta (cDNA sequence 120–122 bp in upper case (u/c), ctg9556.1 sequence 1778–1899 bp in lower case (l/c)) and cagGGT (cDNA sequence 123–125 bp in u/c, and ctg9556.1 sequence 6133–6924 bp in l/c). The entire ORF (399 bp, cDNA sequence 134–532 bp) is encoded by exon 2.

The nucleotide sequence of the mouse FANTOM clone C630041J07 (1326 bp) was aligned to the corresponding genomic nucleotide sequence downloaded from the Ensembl database (Chr. 7: 130361035–130363236). This demonstrated that the mouse *SPRN* gene also has two exons and is very compact (2.2 kb). Exons 1 and 2 are 148 and 1178 bp, respectively, and the intron is 876 bp. The exon–intron boundaries are concordant with consensus sequences: CCAgta (cDNA sequence 146–148 bp in u/c, and intron sequence in l/c) and cagATT (cDNA sequence 149–151 bp in u/c, and intron sequence in l/c). The whole ORF (444 bp, cDNA sequence 164–607 bp) is included in exon 2. Remarkably, neither of the database cDNAs available for mouse Sho, the FANTOM clone C630041J07 and XM_150033 (845 bp) from NCBI, appear to contain a polyA tail.

The length of *SPRN* in human is 3907 bp and it, also, has two exons. The NCBI cDNA (BC040198, 3150 bp) encod-

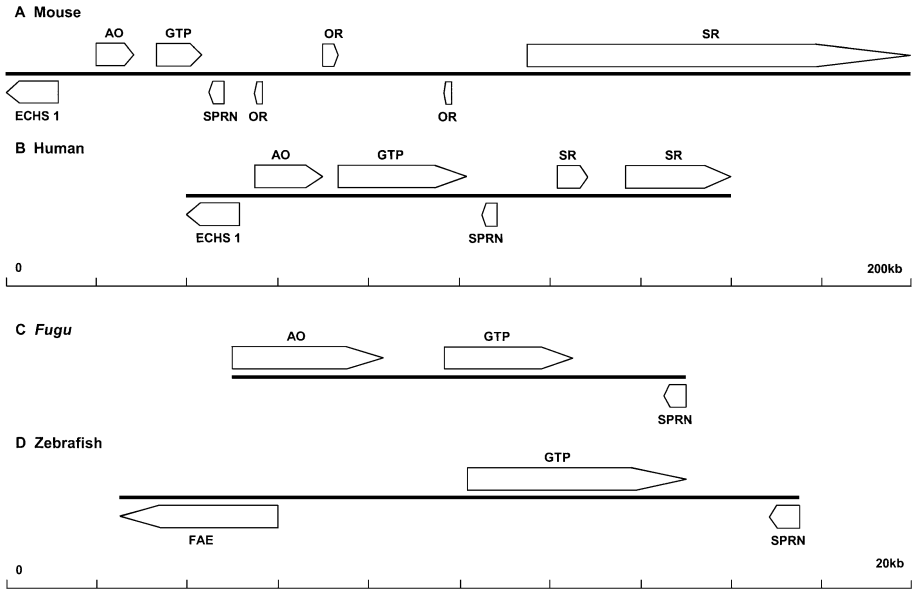

Fig. 5. Summary of conserved contiguity for fish and mammalian *SPRN* genes. Figure is to scale as shown by rulers. ECHS1, enoyl-CoA hydratase/isomerase; AO, amine oxidase; GTP, GTP-binding protein; SPRN, Shadoo; OR, olfactory receptor; SR, speract/scavenger receptor; FAE, long-chain fatty-acyl elongase.

ing human Sho was aligned to its corresponding genomic fragment extracted from the Ensembl database (Chr.10: 134364175–134368086 bp). Exon 1 is 101 bp, but we note sequence 1–5 bp was not aligned. Exon 2 is 3032 bp, and the intron length is again short as in mouse with 779 bp. The exon–intron boundaries are consistent with consensus sequences: GTGgcg (cDNA sequence 99–101 bp in u/c, and intron sequence in l/c) and cagGTT (cDNA sequence 102–104 bp in u/c, and intron sequence in l/c). The entire ORF (456 bp, cDNA sequence 118–573 bp) is encoded by exon 2.

3.7. Genomic context of SPRN

The genomic contexts (Fig. 5) of the mouse, human and *Fugu SPRN* genes are clear from Ensembl's interactive genome browsers, while that of the zebrafish *SPRN* was determined by using the NIX program (Williams et al., 1998). A gene encoding a GTP-binding protein of unknown function is the proximal adjacent gene present both in mammals and fishes, and its tail-to-tail orientation relative to *SPRN* is also conserved from fishes to mammals. The next most proximal gene, which encodes an amine oxidase (AO), is conserved between *Fugu* and mammals, as is its tail-to-tail orientation with *SPRN*. This three-gene block (two genes in zebrafish; see Fig. 5) with its conserved gene order (AO–GTP–*SPRN*) and orientation is an example of conserved contiguity (Gilligan et al., 2002) between fishes and mammals, that strongly indicates gene orthology. However, the genes distal to *SPRN* are not conserved between mouse and human indicating a chromosome rearrangement in either the mouse or human genome. In the human genome, a gene for speract/scavenger receptor lies adjacent to *SPRN*, whereas in the mouse genome three genes encoding olfactory receptors are located between *SPRN* and the scavenger-receptor gene.

3.8. Expression of SPRN

The database searches found entries indicating expression of human, mouse, rat and zebrafish *SPRN* in the embryo, brain and retina, as summarized in Table 1. Evidence of expression is available at the TIGR database for mouse and rat *SPRN*. The mouse ESTs merged into the contig TC613820 are expressed in embryo brain (EST BM944750: 0.03% of the total library NIH_BMAP_EH0p; NIH-MGC, 1999, unpublished) and in the adult hippocampus (EST BB653489: 0.01% of the total library RIKEN full-length enriched, adult male hippocampus; Arakawa et al., 2001, unpublished). The ESTs that comprise the contig TC613821 are expressed in the hippocampus of young mice (EST AI842512: 0.09% of the total library NIH_BMAP_MHI_N; Bonaldo et al., 1996) and adult retina (EST BB284188: 0.01% of the total library RIKEN full-length enriched, adult retina; Konno et al., 2000, unpublished). The ESTs that make the rat TC297203 are expressed in the embryo (ESTs AW530092 and BF564096: 0.06% of the total library UI-R-C4, Bonaldo et al., 1996), adult rat mixed tissue (ESTs BF285504 and BF285497: 0.01% of the total library Rat gene index; Malek et al., 2000, unpublished), and in the brain (EST AA943106: 0.02% of the total library normalized rat brain, Lee et al., 1998, unpublished). The rat contig TC293056 related ESTs are expressed in adult brain (ESTs BF404416, BG376848, BF409274, BF409392 and BF409521: 0.05% of the total library UI-R-CA1; Bonaldo et al., 1996: ESTs AI007831 and AI101223: 0.03% of the total library normalized rat brain; Lee et al., 1998, unpublished). Human hippocampus was the source for the NIH_MGC_95 library (Jones et al., unpublished) from which human cDNA BC040198 originates. The zebrafish ESTs AL913622, AL913623, AL913624 and AL913625 found in the NCBI est_others database are expressed in the whole embryo PJR-Z1+Z2 library (Lee et

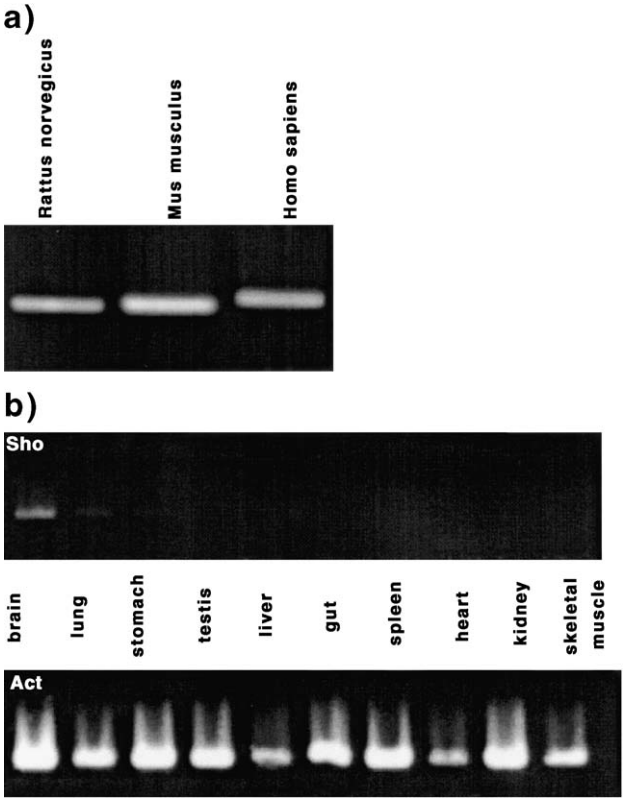

Fig. 6. RT-PCR analysis of mammalian Shadoo mRNA expression. (a) Whole brain: seminested PCRs were performed using two forward primers, coupled with a single reverse primer. In particular GCCTTCCTCTGTGACAGC followed by TCTGGCTGGAGGAGGACC for rodents and GCCTTCCTCTGCGACAGC followed by TCGGGCTGGAGAAGGGCC for human, coupled with CTAGGATGGAGCTTAGCC (rat), CTGGGATGGAGTTTAGCC (mouse) and CAGATGTGGTCCCCGAGC (human). The sizes of the fragments are 220 bp for rodents and 226 bp for human. (b) Rat tissues: the upper panel is for Shadoo, where the one step PCR product of 308 bp was obtained using the primer couple GCCTTCCTCTGTGACAGC/CCCAGTAGCTGTAGACTCC; the lower panel is for control actin, amplified using the primers AGCCATGTACGTAGCCATCC and AATGTCACGCACGATTTCCC.

al., 2002, unpublished) and EST BG738569 is expressed in the adult retina library (Clark et al., 1998, unpublished).

To confirm this indicative database information, expression of the mammalian mRNAs was demonstrated in whole brain by RT-PCR, followed by sequencing of the products of the expected size. To this purpose, two PCR assays, the second seminested, were carried out on random primed reverse transcripts, using primer pairs designed on the basis of sequences deposited in databases (Fig. 6a). Sho mRNA expression was also assayed in different rat tissues by one step PCR using a new downstream primer, leading to an obvious change in the product size (308 bp). Such substitution was necessary to obtain clean and reliable results, as initially a misleading band, very close to the expected size of 220 bp, was detectable in some non-brain tissues and identified by sequencing as Factor V. Results reported in Fig. 6b show that, besides brain, fainter bands are detectable only in two (lung and stomach) out of nine tested tissues. These data demonstrate the very limited tissue expression pattern of the Sho transcript, and serve to highlight its presence in brain.

4. Discussion

4.1. Protein structure

4.1.1. Sequence regions

Inspection of Fig. 3 shows that Sho proteins have the following major features:

(a) An N-terminal peptide sequence (aa 1–24) consistent with an endoplasmic reticulum targeting signal for extracellular export.
(b) A basic RG-rich region starting from Lys-25 with up to six tetrarepeats of consensus XXRG, where X is G, A or S. For mammals, the pattern is GGRG GARG SARG (G/-)VRG GARG ASRV; this pattern is well conserved in zebrafish but is distorted in *Fugu* by an insertion of 14 residues in the middle. This region finishes with Arg-54 (zebrafish).
(c) A hydrophobic stretch in the middle of the protein (aa 55–74, zebrafish), which contains the same unusual composition of small aliphatic residues (GAV) as for PrP and PrP-like proteins (see Section 4.1.3).
(d) A C-terminal region with a putative *N*-glycosylation site (Asn-93, zebrafish) and a possible GPI-anchor site (Gly-106, zebrafish). This region is enriched with acidic and hydroxyl-containing residues, but otherwise has a more standard mix of amino acids which might be compatible with folded structure. However, secondary structure predictions (not presented) indicate mostly coil with no consistent or strong predictions of secondary structure. The region is also shorter in fish Shos.
(e) A C-terminal sequence predicted as a signal sequence for GPI-anchor attachment.

4.1.2. Conservation of Sho sequences

In contrast to the PrP-like proteins (Suzuki et al., 2002), the sequences of the *Fugu* and zebrafish Shadoo proteins show good conservation over the whole length (identity 54%, zebrafish 1–106, excluding C-terminal signal sequence). Conservation among the mammalian (human, mouse, rat) sequences is very high (identity 81–96%). The fish and mammalian sequences are also well conserved, particularly for zebrafish, to slightly beyond the end of the hydrophobic sequence (identity 41–53%, zebrafish 1–78); remarkably, this includes the N-terminal signal sequence. The C-terminal region is less well conserved between fish and mammals but all show a predicted *N*-glycosylation site at equivalent positions.

4.1.3. Comparison of Sho, PrP, stPrP and PrP-like sequences

Considering now the overall protein structures of Sho, PrP, stPrP and PrP-like sequences shown in Fig. 1, together with Fig. 3, we can make the following observations. All appear to be exported extracellular proteins attached to the membrane by GPI anchors. All show an internal hydrophobic segment "anchored" on either side by basic residues (see Fig. 7, and Suzuki et al., 2002; Oidtmann et al., 2003). The highly conserved hydrophobic sequence of Sho proteins is the same length as that of PrPs (except *Xenopus* PrP), while this region of stPrPs is one to two residues shorter and that of PrP-like proteins four residues shorter, this gap being positioned as for *Xenopus* PrP (Fig. 7). The alignment of the hydrophobic segment (Fig. 7) shows strong conservation across all PrPs and Shos; 12 of the 20 residues (112–131, HuPrP) are identical or almost identical, and another 6 are conserved hydrophobic.

The main regions of structural divergence between Shos and PrPs, and also PrP-likes (Suzuki et al., 2002), are the post-hydrophobic C-terminal region and the N-terminal repeat region. In contrast to the fish (only) PrP-like proteins, the C-terminal region sequence (spanning zebrafish Sho residues 75–106) is not low complexity, and, hence, as noted above, might adopt folded or stable secondary struc-

Fig. 7. Alignment of hydrophobic regions of PrPs and Shos. In consensus line: capital letters, conserved in seven or more species; lower-case letters, conserved in six or more species; !, conservatively conserved hydrophobic; +, conserved basic. Hu, human; Po, (marsupial) possum; Ch, chicken; Tu, turtle; Xe, *Xenopus*; Mo, mouse; Ze, zebrafish; Fu, *Fugu*.

ture. However, the C-terminal sequences are very different from those of PrPs and its mammalian homologue, Doppel protein (Dpl) (Moore et al., 1999), for which a tertiary structure comprising three helices, two short β-strands, and a large proportion of coiled structure has been determined (Zahn et al., 2000; Mo et al., 2001), and for the apparently homologous stPrPs (Rivera-Milla et al., 2003; Oidtmann et al., 2003). The respective regions are also much shorter; human Sho 87–126 (40 residues) compared with human PrP 132–231 (100 residues) while the fish sequences are even shorter—zebrafish Sho and *Fugu* Sho, of 32 and 30 residues, respectively. While PrPs have two (or three) *N*-glycan sites in the C-terminal region, and stPrPs have one (*Fugu* stPrP-1) or both (*Tetraodon*) conserved glycosylation sites or two (salmon) or three (*Fugu* stPrP-2) non-conserved sites (Oidtmann et al., 2003), Shos have only one potential site. Deactivation experiments for each PrP *N*-glycosylation site showed one was associated more with cell-trafficking while the second appeared to stabilize the structure (Lehmann and Harris, 1997).

Comparing now the N-terminal region (from the end of the signal sequence to the hydrophobic segment) for Shos, PrPs, stPrPs and PrP-likes, we note that all are basic and low complexity, with the region being highly extended in stPrPs. PrPs (except *Xenopus*), Shos and zebrafish PrP-like all show distinct repeating sequences within the region, but of different sequence and at different positions. PrP sequences show PGH-rich repeats of 8 (mammalian), 10 (marsupial) or 6 (avian and turtle) residues inserted in the middle of the basic region, whereas Shos have Arg-rich tetrarepeats starting at the beginning of the region. Zebrafish PrP-like, on the other hand, contains six trirepeats of consensus HXG (where X is mostly T) inserted towards the end of the N-terminal region from residues 65 to 82. The full picture, which emerges from consideration of Fig. 1, suggests that this basic N-terminal region should be considered as one region which has a propensity to insert sequence; formal repeats in mammalian, marsupial, avian and turtle PrPs and zebrafish PrP-like, and non-repeat sequence in *Xenopus* PrP, fish stPrPs and *Fugu* Sho.

On this model, it is useful to consider the structure of the mature forms of this type of "prion" protein as consisting of four regions: the basic region 1 which shows a tendency for insertion of repeat or other sequence (region 2), the hydrophobic region 3, and the C-terminal region 4. Setting aside the question of the evolutionary relationships of these proteins, which are not clear (Suzuki et al., 2002; Rivera-Milla et al., 2003; Oidtmann et al., 2003; Premzl et al., in preparation), as also discussed in the Introduction, then a dynamic structural scaffold is evident upon which variations might have produced proteins with quite different functions in different lineages. In particular, there appears to have been a gradual gain in structural features and new potential functions in PrPs. The C-terminal region of Sho is increased in length between fishes and mammals.

4.2. Gene structure and expression

From combined cDNA and genomic data we were able to determine that human, mouse and zebrafish *SPRN* genes all contain two exons, with the ORF contained entirely within exon 2. While the intron in human and mouse *SPRN* is very small, the intron of zebrafish *SPRN* is much longer. Although a second upstream start codon was detected in the zebrafish cDNA, which might have coded for a shorter overlapping frame-shifted ORF, this is of doubtful significance as there is no evidence for a second ORF in other *SPRN* sequences.

The database searches found entries indicating expression of *SPRN* in embryo, brain and retina of mouse and rat, in hippocampus of human, and in embryo and retina of zebrafish. We have confirmed expression of the mammalian (human, mouse, rat) transcripts in whole brain directly using RT-PCR, and shown that brain expression is highly specific by examination of nine other tissues in the rat. Transcript expression for *Fugu* PrP-like was reported in skin and retina of embryos, and retina and weakly in brain of adults (Suzuki et al., 2002); this pattern seems similar to that of zebrafish *SPRN*. For salmon stPrP and *Fugu* stPrP-1/PrP-461, transcript was detected in all organs examined but predominantly in brain, whereas for *Fugu* stPrP-2 it was not detected in brain and other tissues were not examined (Rivera-Milla et al., 2003; Oidtmann et al., 2003).

Analysis of the available genomic sequences for human, mouse, *Fugu* and zebrafish allowed us to map proximal and distal genes, and ascertain their relative transcription directions (Fig. 5). Conservation of gene order and orientation for a proximal three-gene block of an amine oxidase, GTP-binding protein and *SPRN* (only second two in zebrafish) between fishes and mammals, strongly indicates gene orthology. The distal genes in mouse and human show evidence of rearrangements, with olfactory receptor genes inserted between *SPRN* and a scavenger receptor gene in mouse. Whether there is any functional significance for the linkage of these genes with *SPRN* is unknown.

4.3. Mammalian Shadoo/PrP and Doppel/PrP links

There is an urgent need to define functions of mammalian PrPs because of the increasing importance of PrP diseases. Comparison of the relationships of mammalian PrPs to Dpls, and now PrPs to Shos, could help to reveal the normal function and relationships of these proteins, and the generation of abnormal forms of PrP associated with transmissible spongiform encephalopathies.

So far Dpls have been reported only in mammals (Moore et al., 1999; Li et al., 2000; Mastrangelo and Westaway, 2001). Dpls show clear sequence homology with PrPs in the C-terminal domain, which, indeed, has been shown by NMR to have the same 3-D structure (Mo et al., 2001). They both have N- and C-terminal signal sequences, and are both expressed extracellularly attached to the membrane by a GPI anchor (Silverman et al., 2000). Their adjacent

gene locations (Moore et al., 1999) suggest a duplicated gene and they are regarded as paralogues (Mastrangelo and Westaway, 2001). However, Dpls lack the other PrP sequence characteristics (basic region, repeats, hydrophobic region), and increasing evidence showing quite different cellular location (Shaked et al., 2002) and tissue expression (Peoc'h et al., 2002) points to distinctly different functions. Mammalian PrPs are expressed largely in certain cell types of the neuroendocrine, neuroimmune, and peripheral nervous systems (Ford et al., 2002). Dpls, however, are not normally expressed at high levels in brain and are in fact toxic when overexpressed in *PRNP* gene-ablated mice as the result of intergenic splicing events (Mastrangelo and Westaway, 2001). They have now been shown to be sperm proteins required for fertilization (Behrens et al., 2002). PrP co-exists with Dpl in sperm only in the N-terminally truncated form (Peoc'h et al., 2002); a C-terminally truncated form of PrP in sperm has also been reported (Shaked et al., 1999).

In contrast, brain expression and similarity of all sequence features other than the C-terminal domain makes mammalian Sho proteins good candidates for further investigation for overlapping functions, and possibly direct physical interactions, with mammalian PrPs in brain. The recently characterized ability of mammalian PrP to bind heparan sulfate proteoglycans (i.e. negatively-charged sugars) through the two parts of their N-terminal basic region (Warner et al., 2002) might also be shared by the Arg-rich N-terminal region of Sho, although binding of this region to polynucleotides, as also shown for mammalian PrP for the N-terminal fragment (Grossman et al., 2003), is another possibility.

4.4. Conclusion—a new vertebrate gene related to PRNP

Experimental data and database analysis, therefore, reveals the presence of a novel highly conserved gene *SPRN* related to *PRNP*, which encodes prion protein. The lower level of database information for the human *SPRN* gene is somewhat surprising, particularly given the extensive genomic and expression data available for the mouse *SPRN* gene. While no information was discovered in databases for other vertebrates, the conserved contiguity, gene structure and protein sequence between fish and mammals predict that *SPRN*/Sho is also present in other vertebrate classes.

The normal function of prion protein is still not apparent, despite much work over almost two decades. One of the puzzling features is the finding that *PRNP* gene-ablated mice have no obvious phenotype, implying either that the function of PrP "is not apparent in a laboratory setting or that other molecules have overlapping functions" (Mastrangelo and Westaway, 2001). We suggest here that the Shadoo protein is PrP's long-sought shadow.

Acknowledgements

JEG and JAMG are supported by the ANU IAS block grant, and TS is supported by EC grant QLK5-2002-00866 and FIRST. We thank DNASTAR for an extended trial of their Lasergene software. The authors thank Dr. G.A. Niemi for providing the zebrafish retina cDNA library.

References

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

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Behrens, A., Genoud, N., Naumann, H., Rulicke, T., Janett, F., Heppner, F.L., Ledermann, B., Aguzzi, A., 2002. Absence of the prion protein homologue Doppel causes male sterility. EMBO J. 21, 3652–3658.

Bonaldo, M.F., Lennon, G., Soares, M.B., 1996. Genome Res. 6, 791–806.

Borodovsky, M., McIninch, J.D., 1993. GeneMark: parallel gene recognition for both DNA strands. Comput. Chem. 17, 123–133.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94.

Chomczynsky, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal. Biochem. 162, 156–159.

Collinge, J., 2001. Prion diseases of humans and animals: their causes and molecular basis. Annu. Rev. Neurosci. 24, 519–550.

Eisenhaber, B., Bork, P., Eisenhaber, F., 1999. Prediction of potential GPI-modification sites in proprotein sequences. J. Mol. Biol. 292, 741–758.

Ford, M.J., Burton, L.J., Morris, R.J., Hall, S.M., 2002. Selective expression of prion protein in peripheral tissues of the adult mouse. Neuroscience 113, 177–192.

Gilligan, P., Brenner, S., Venkatesh, B., 2002. *Fugu* and human sequence comparison identifies novel human genes and conserved non-coding sequences. Gene 294, 35–44.

Goodstadt, L., Ponting, C.P., 2001. CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics 17, 845–846.

Grossman, A., Zeiler, B., Sapirstein, V., 2003. Prion protein interactions with nucleic acid: possible models for prion disease and prion function. Neurochem. Res. 28, 955–963.

Harris, D.A., Falls, D.L., Johnson, F.A., Fishbach, G.D., 1991. A prion-like protein from chicken brain copurifies with an acetylcholine receptor-inducing activity. Proc. Natl. Acad. Sci. U. S. A. 88, 7664–7668.

Hedges, S.B., Kumar, S., 2002. Vertebrate genomes compared. Science 297, 1283–1285.

Huang, X., Miller, W., 1991. A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. 12, 337–357.

Korf, I., Flicek, P., Duan, D., Brent, M.R., 2001. Integrating genomic homology into gene structure prediction. Bioinformatics 17, S140–S148.

Lee, I.Y., Westaway, D., Smit, A.F.A., Wang, K., Seto, J., Chen, L., Acharya, C., Ankener, M., Baskin, D., Cooper, C., Yao, H., Prusiner, S.B., Hood, L.E., 1998. Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res. 8, 1022–1037.

Lehmann, S., Harris, D.A., 1997. Blockade of glycosylation promotes acquisition of scrapie-like properties by the prion protein in cultured cells. J. Biol. Chem. 272, 21479–21487.

Li, A., Sakaguchi, S., Atarashi, R., Roy, B.C., Nakaoke, R., Arima, K., Okimura, N., Kopacek, J., Shigematsu, K., 2000. Identification of a novel gene encoding a PrP-like protein expressed as chimeric transcripts fused to PrP exon 1/2 in ataxic mouse line with a disrupted PrP gene. Cell. Mol. Neurobiol. 20, 553–567.

Marchuk, D., Drumm, M., Saulino, A., Collins, F.S., 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19, 1154.

Mastrangelo, P., Westaway, D., 2001. The prion gene complex encoding PrP(C) and Doppel: insights from mutational analysis. Gene 275, 1–18.

Mo, H., Moore, R.C., Cohen, F.E., Westaway, D., Prusiner, S.B., Wright, P.E., Dyson, H.J., 2001. Two different neurodegenerative diseases caused by proteins with similar structures. Proc. Natl. Acad. Sci. U. S. A. 98, 2352–2357.

Moore, R.C., Lee, I.Y., Silverman, G.L., Harrison, P.M., Strome, R., Heinrich, C., Karunaratne, A., Pasternak, S.H., Azhar Chishti, M., Liang, Y., Mastrangelo, P., Wang, K., Smit, A.F.A., Katamine, S., Carlson, G.A., Cohen, F.E., Prusiner, S.B., Melton, D.W., Tremblay, P., Hood, L.E., Westaway, D., 1999. Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. J. Mol. Biol. 292, 797–817.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Ning, Z., Cox, A.J., Mullikin, J.C., 2001. SSAHA: a fast search method for large DNA databases. Genome Res. 11, 1725–1729.

Oesch, B., Westaway, D., Waelchli, M., McKinley, M.P., Kent, S.B.H., Aebersold, R.H., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E., Prusiner, S.B., Weissmann, C., 1985. A cellular gene encodes scrapie PrP 27-30 protein. Cell 40, 735–746.

Oidtmann, B., Simon, D., Holtkamp, N., Hoffmann, R., Baier, M., 2003. Identification of cDNAs from Japanese puffer fish (*Fugu rubripes*) and Atlantic salmon (*Salmo salar*) coding for homologues to tetrapod prion proteins. FEBS Lett. 538, 96–100.

Peoc'h, K., Serres, C., Frobert, Y., Martin, C., Lehmann, S., Chasseigneaux, S., Sazdovitch, V., Grassi, J., Jouannet, P., Launay, J.M., Laplanche, J.L., 2002. The human "prion-like" protein Doppel is expressed in both Sertoli cells and spermatozoa. J. Biol. Chem. 277, 43071–43078.

Prusiner, S.B., 1998. Prions. Proc. Natl. Acad. Sci. U. S. A. 95, 13363–13383.

Rivera-Milla, E., Stuermer, C.A.O., Malaga-Trilla, E., 2003. An evolutionary basis for scrapie disease: identification of a fish prion mRNA. Trends Genet. 19, 72–75.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Shaked, Y., Rosenmann, H., Talmor, G., Gabizon, R., 1999. A C-terminal-truncated PrP isoform is present in mature sperm. J. Biol. Chem. 274, 32153–32158.

Shaked, Y., Hijazi, N., Gabizon, R., 2002. Doppel and PrP(C) do not share the same membrane microenvironment. FEBS Lett. 530, 85–88.

Silverman, G.L., Qin, K., Moore, R.C., Yang, Y., Mastrangelo, P., Tremblay, P., Prusiner, S.B., Cohen, F.E., Westaway, D., 2000. Doppel is *N*-glycosylated GPI-anchored protein: expression in testis and ectopic production in the brains of *Prnp*^{00} mice predisposed to Purkinje cell loss. J. Biol. Chem. 275, 26834–26841.

Simonic, T., Duga, S., Negri, A., Tedeschi, G., Malcovati, M., Tenchini, M.L., Ronchi, S., 1997. cDNA cloning and expression of the flavoprotein D-aspartate oxidase from bovine kidney cortex. Biochem. J. 322, 729–735.

Simonic, T., Duga, S., Strumbo, B., Asselta, R., Ceciliani, F., Ronchi, S., 2000. cDNA cloning of turtle prion protein. FEBS Lett. 469, 33–38.

Solovyev, V.V., 2001. Statistical approaches in eukaryotic gene prediction. In: Balding, D., et al. (Eds.), Handbook of Statistical Genetics. Wiley, New York, pp. 83–127.

Strumbo, B., Ronchi, S., Bolis, L.C., Simonic, T., 2001. Molecular cloning of the cDNA coding for *Xenopus laevis* prion protein. FEBS Lett. 508, 170–174.

Suzuki, T., Kurokawa, T., Hashimoto, H., Sugiyama, M., 2002. cDNA sequence and tissue expression of *Fugu rubripes* prion protein-like: a candidate for the teleost orthologue of tetrapod PrPs. Biochem. Biophys. Res. Commun. 294, 912–917.

Taylor, W.R., 1990. Hierarchical method to align large numbers of biological sequences. Methods Enzymol. 183, 456–474.

Warner, R.G., Hundt, C., Weiss, S., Turnbull, J.E., 2002. Identification of the heparan sulfate binding sites in the cellular prion protein. J. Biol. Chem. 277, 18421–18430.

Williams, G.W., Woollard, P.M., Hingamp, P., 1998. NIX: a nucleotide identification system at the HGMP-RC. URL: http://www.hgmp.mrc.ac.uk/NIX/.

Windl, O., Dempster, M., Estibeiro, O., Lathe, R., 1995. A candidate marsupial PrP gene reveals two domains conserved in mammalian PrP proteins. Gene 159, 181–186.

Wopfner, F., Weidenhofer, G., Schneider, R., von Brunn, A., Gilch, S., Schwarz, T.F., Werner, T., Schatzl, H.M., 1999. Analysis of 27 mammalian and 9 avian PrPs reveals high conservation of flexible regions of the prion protein. J. Mol. Biol. 289, 1163–1178.

Zahn, R., Liu, A., Luhrs, T., Riek, R., von Schroetter, C., Lopez Garcia, F., Billeter, M., Calzolai, L., Wider, G., Wuthrich, K., 2000. NMR solution structure of the human prion protein. Proc. Natl. Acad. Sci. U. S. A. 97, 145–150.