Chapter 4: Discovery of Shadow of Prion Protein and Shadoo

Chapter 4: Discovery of Shadow of Prion Protein and Shadoo

4.1 Introduction

Genes could be manually annotated by compiling direct evidence, *ab initio* gene prediction and homology-based evidence (Chapter 2.6).

I applied this approach to discover *PRNP* homologues. Starting from the initial lead of the sequence of a zebrafish cDNA (Prof. Tatjana Simonic, University of Milan) and using the information that already existed in the public databases, I discovered the new human gene that is present also in other vertebrates (from mammals to fish). Sequence similarities of the protein product to PrP, as well as its predominant expression in brain, allowed the hypothesis that it may be functionally related to PrP and that it could contribute to understanding of prion disease pathogenesis (Prusiner, 1998).

Drs. Jill Gready, Jenny Graves and I called this new gene "Shadow of prion protein" *SPRN*, and the protein "Shadoo" (Japanese for shadow). My inspiration for the gene name was the famous Plato's "Allegory of the cave" (4th century BC, from Plato, The Republic). In the Plato's dual world there was the world of ideas, and there was the world of their "shadows". It is the world of "shadows" that we usually see, like we see shadows on the cave wall. But every "shadow" is a mere representative of its own reality, of its own idea. So I thought, Shadow may be a shadow and a representative of its well-known idea, prion protein. Yet, in the following chapters I show that in fact the PrP may be a shadow of Shadow.

The Sho sequence is well conserved between fish (zebrafish, *Fugu*) and mammals (human, mouse, rat), as well as the gene structure and genomic location. By combining the direct evidence (cDNAs, ESTs, gene expression) with homology evidence and gene predictions, I compiled enough support for the new human gene *SPRN* conserved between fish and mammals (Table 4.1).

Table 4.1: 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: 130362897 - 130364466	NIH_BMAP_EH0p, RIKEN full-length enriched, adult male hippocampus
	TIGR	-	-	TC613821	*BB284188, *AI842512	Chr. 7: 130362273 - 130362792	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	-
	NCBI	-	-	BC040198	NT_017795	Chr. 10 q26.3	-
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,	-	UI-R-C4, Rat gene index, Normalized

	TIGR	-	-	TC293056	*BF285504 *BF285497, *AA943106, *BF404416, *AI007831, *BG376848, *AI101223, *BF409274, *BF409392, *BF409521	-	rat brain 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*), Genscans; (*f*), Fgenesh++; (*t*), Twinscan; (*gs*) , Genomescan. [b], ESTs are labelled by asterisk.

4.2 Discovery of *SPRN* Gene

In order to discover new *PRNP* homologues, I searched manually, and in a targeted fashion (Chapter 2.6.6), data in public databases.

4.2.1 Discovery of Mouse *Sprn*

I first discovered the mouse Sho protein (XP_150033) in the NCBI nr protein database using zebrafish *sprn* cDNA sequence (Prof. Tatjana Simonic's group, University of Milan). I used the zebrafish Sho sequence (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. I also found the corresponding *Sprn* gene, transcript, genomic clone and chromosomal coordinates. In the Ensembl mouse genome database (v9.3a.1), I detected mouse Sho protein in the genscan (Burge and Karlin, 1997) predicted peptides database using the server's BLASTP tool (Altschul et al., 1990, and W. Gish, 1994-1997, unpublished) and the XP_150033 amino acid sequence as search query. Information on the gene, EST, genomic clone, chromosomal position and gene predictions 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. I found also the cDNA clone encoding mouse Sho in the FANTOM database using the mouse *Sprn* ORF nucleotide sequence extracted from XM_150033 as the search query and the server's BLASTN (Altschul et al., 1997) program. I found this cDNA in the DDBJ database in the same manner, as well as the contigs, genomic location and expression data in the TIGR database.

4.2.2 Discovery of Human *SPRN*

In the Ensembl human genome database (v9.3a.1), I was able to detect *SPRN* genscan and Fgenesh++ predictions in the genomic position analogous to that of mouse *Sprn*. I found the putative *SPRN* genomic location in the NCBI human genome database after keyword search and detection of the annotated adjacent gene *ECHS1*. The human *SPRN*

transcript was found in the NCBI nr database using the mouse ORF sequence from XM_150033 and the server's BLASTN program.

4.2.3 Discovery of Rat *Sprn*

I detected the rat *Sprn* contigs in the TIGR database comparative genome browser. The rat *Sprn* tentative consensus sequences TCs (virtual transcripts created by merging the EST data, which may be full or partial cDNA length) were aligned with their corresponding TCs from mouse (TC613820 and TC613821). An additional EST BF391059 and expression data were also present in the TIGR database. The protein sequence was derived by conceptual translation of the TIGR contigs. Genomic clones 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, I found the protein, gene, transcript, and chromosomal location information in the updated Ensembl rat database (v12.2.1).

4.2.4 Discovery of *Fugu SPRN*

I used the zebrafish Sho amino acid sequence as query in the BLASTP analysis of Ensembl's *Fugu rubripes* genscan peptides database. The prediction of *Fugu* Sho, *SPRN* and its transcript corresponded to the genomic fragment Chr_scaffold_28. I detected the same genomic scaffold by using SINFRUT00000151627 as query sequence and the BLASTN tool on the HGMP web server. I translated the *Fugu* Sho amino acid sequence form the scaffold_28, which contained the complete ORF.

4.2.5 Discovery of *Tetraodon SPRN*

I used the *Fugu SPRN* ORF as query sequence to search the Genoscope *Tetraodon nigroviridis* whole genome shotgun database with the Genoscope BLASTN server. Although the genomic clone FS_CONTIG_4144_1 harboured *Tetraodon SPRN* ORF, conceptual translation of the nucleotide sequence was impossible due to the poor quality

of the sequence data. I therefore used this sequence to design primers and amplify, clone and sequence *Tetraodon SPRN* ORF (Chapter 5.5.7).

4.2.6 Discovery of Zebrafish *sprn* Gene and ESTs

Using the zebrafish cDNA sequence as search query with BLASTN (Altschul et al., 1997) on the local server at the Massachusetts General Hospital Renal Unit I first found the genomic sequence assembly z06s038879 harbouring the zebrafish *sprn* in the Sanger Institute zebrafish genome Assembly 6. I also detected the sequence read zfishC-a2446d07.g1c in the Ensembl trace respository using the web search tool Sequence Search and Alignment by Hashing Algorithm (SSAHA; Ning et al., 2001). I identified the ESTs covering parts of the zebrafish *sprn* cDNA by searching the NCBI est_others database. A BLAST search of the 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.

4.2.7 Data from Other Species

I found no *SPRN* in other species by analysis of available data in the Flybase, M Base, NCBI (including *Xenopus laevis* and other vertebrates), Sanger Institute (*X. tropicalis*) and WormBase. However, the vertebrate databases are still very incomplete; it is very likely that *SPRN* will be also found in other vertebrates.

This database evidence enabled me to annotate the new vertebrate *SPRN* gene.

4.3 Translation of Shos

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 (Chapter 3.1.3). 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 TCs. The *Fugu* Sho amino acid sequence was derived by conceptual

translation of the genomic scaffold_28 from *Fugu* Ensembl v12.2.1 (ORF is encoded by the nucleotide sequence 392342-392800 bp).

4.4 Analysis of Shos

I aligned and analysed the set of predicted Shos, and predicted their posttranslational modifications.

4.4.1 Sequence Alignment

I first aligned the protein sequences. The mouse (XP_150033), human (genscan prediction AL 161645.14.1.161644.7867.9560), rat (derived by conceptual translation) and *Fugu* (deduced from the genomic data) Shos were aligned with the zebrafish Sho (Figure 4.1). This alignment was done with the Taylor (1990) algorithm within the program Cameleon (Chapter 3.1.3).

4.4.2 Prediction of Signal Peptides

Next, I predicted the signals peptides in the sequences. Proximal signal peptide cleavage sites were predicted using the SignalP program (Chapter 3.1.3). These predictions are consistent for all Sho proteins (Figure 4.1), indicating cleavage before the first (conserved) Lys residue and strongly supporting its extracellular location both in fish and mammals. BigPI (Chapter 3.1.3) predicts GPI anchor attachment for mammalian Sho proteins (human, mouse and rat at G126, G122 and G122, respectively). Although confidence level for 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 glycines described above (Figure 4.1), suggests they also might be GPI-anchored.

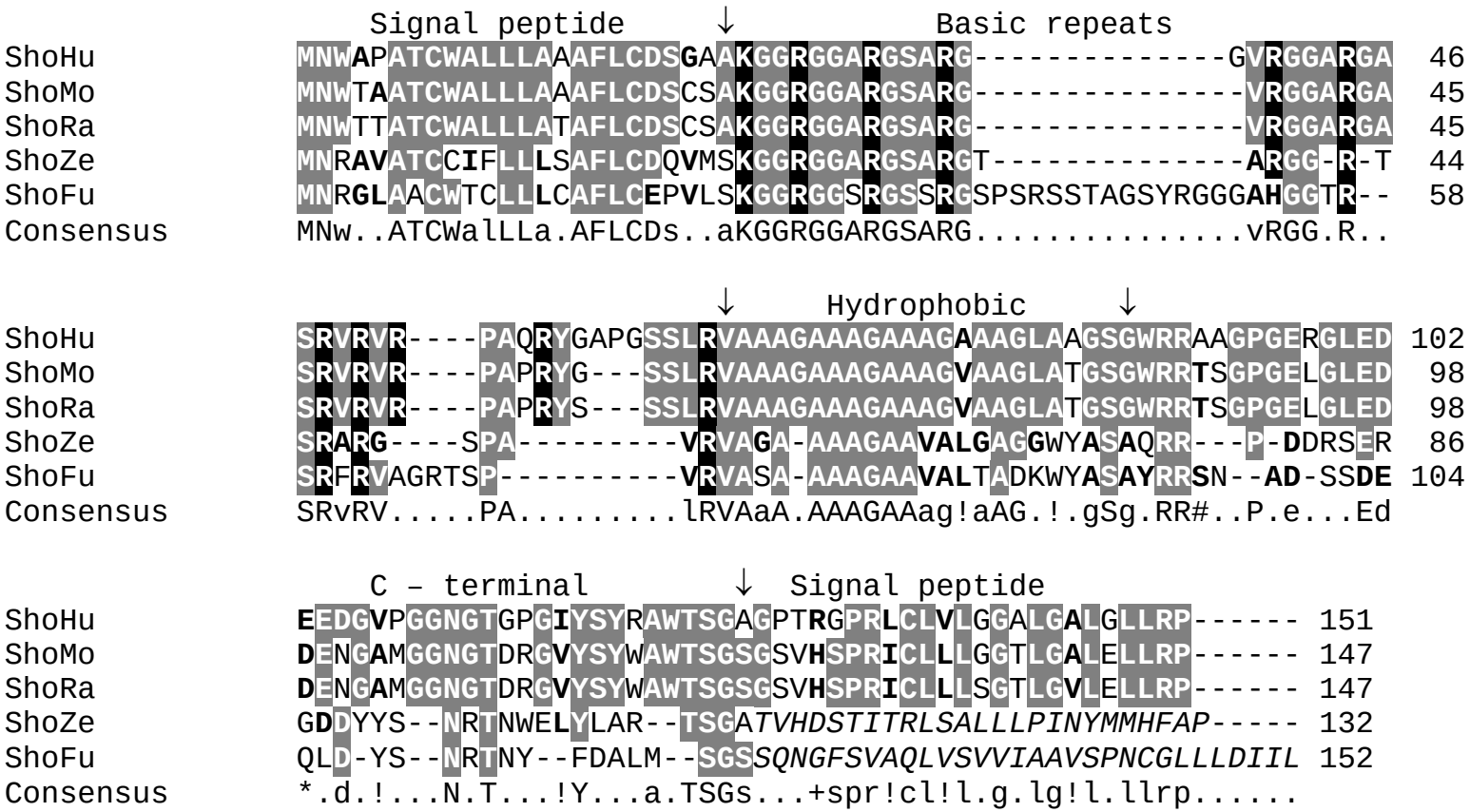

Figure 4.1: Alignment of Sho proteins for fish and mammals. Arrows show cleavage sites and beginning and end of hydrophobic segment. Reversed black, conserved basic; reversed grey, conserved in at least 3 species; bold letters, conservatively conserved. In consensus line: capital letters, conserved in 4 or more species; lower-case letters, conservatively conserved in 4 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*.

4.5 *SPRN* Gene Structure

The transcripts and genomic sequences in hand enabled me to infer the *SPRN* gene structure (Figure 4.2). I aligned the zebrafish *sprn* cDNA (908 bp) with the genomic sequence (ctg9556.1) downloaded from the Ensembl Zv2 pre-assembly database. The zebrafish *sprn* gene has 2 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: GAG|gta (cDNA sequence 120-122 bp in u/c, ctg9556.1 sequence 1778-1899 bp in l/c) and cag|GGT (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.

I aligned the nucleotide sequence of the mouse FANTOM clone C630041J07 (1326 bp) with the corresponding genomic nucleotide sequence downloaded from the Ensembl database (Chr. 7: 130361035-130363236 bp; Ensembl mouse v9.3a.1). This demonstrated that the mouse *Sprn* gene also has 2 exons and is very small (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: CCA|gta (cDNA sequence 146-148 bp in u/c, and intron sequence in l/c) and cag|ATT (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. Worth noting here is that 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 and may be incomplete.

The length of *SPRN* in human is 3907 bp and it, also, has two exons. I aligned the cDNA (BC040198, 3150 bp) encoding human Sho with its corresponding genomic fragment extracted from the Ensembl database (Chr.10: 134364175 – 134368086 bp; Ensembl human v9.3a.1). Exon 1 is 101 bp, but I noted that 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: GTG | gcg

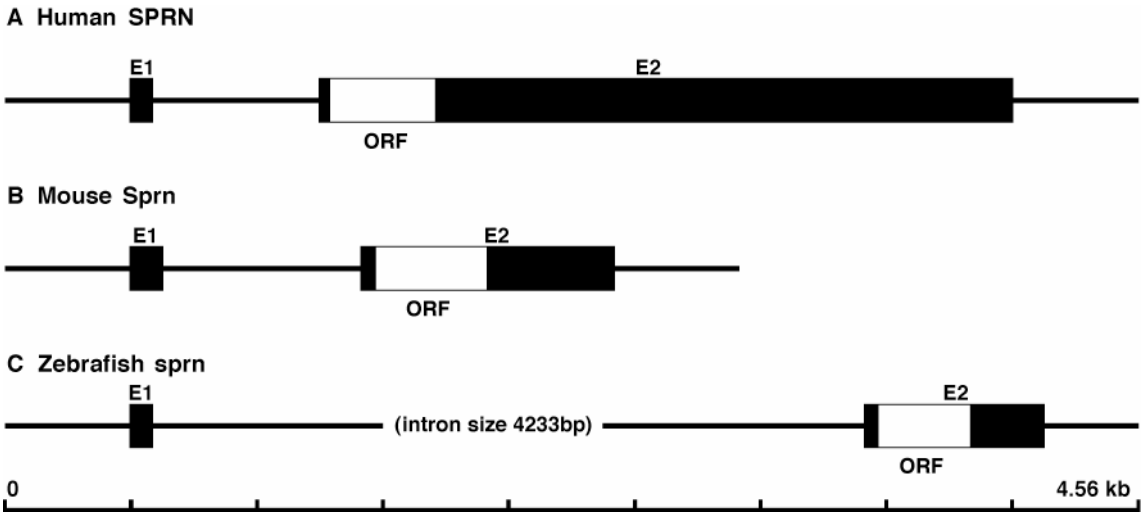

Figure 4.2: Intron-exon structure for human, mouse and zebrafish *SPRN* genes. Figure is to scale as shown by the ruler, except the very long intron in zebrafish has been abbreviated. E1, exon 1; E2, exon 2; ORF, open reading frame.

(cDNA sequence 99-101 bp in u/c, and intron sequence in l/c) and cag | GTT (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.

The structure of *SPRN* is conserved from fish to mammals. However, the zebrafish *sprn* gene has much longer intron than mammalian *SPRN*s.

4.6 Genomic Context of *SPRN*

Genome browsers indicate local genomic context for genes. The genomic contexts (Figure 4.3) of the mouse, human and *Fugu SPRN* genes are clear from Ensembl's interactive genome browsers, while I determined that of the zebrafish *sprn* by using the NIX program (Chapter 3.1.2). A gene encoding a GTP-binding protein of unknown function is the proximal adjacent gene present both in mammals and fish, and its tail-to-tail orientation relative to *SPRN* is also conserved from fish 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 (2 genes in zebrafish) with its conserved gene order (*AO-GTP-SPRN*) and orientation is an example of conserved contiguity (Gilligan et al., 2002) between fish and mammals strongly indicating 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.

The genomic context of *SPRN* gene is conserved from fish to mammals. Whereas the three genes span roughly 50 kb in mammals, they span about 10 kb in *Fugu* and about 16 kb in zebrafish.

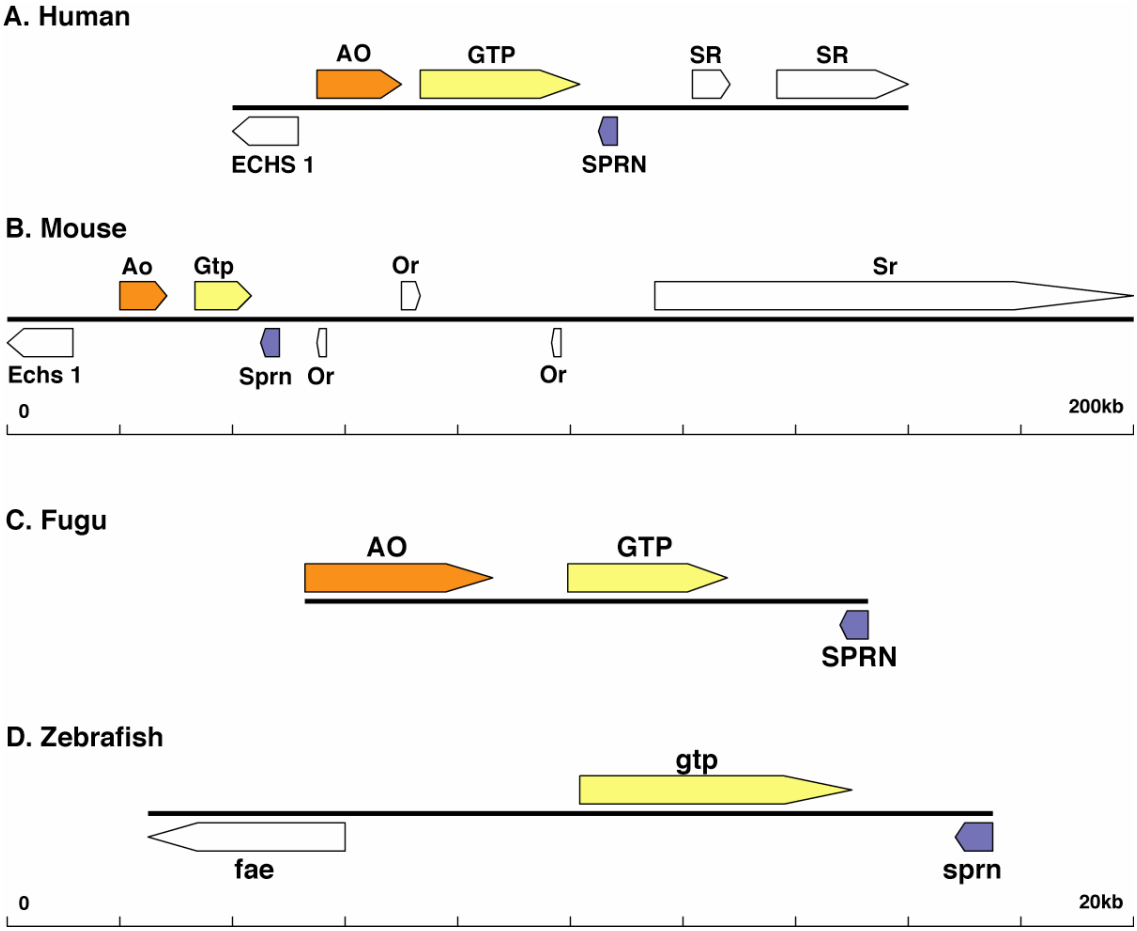

Figure 4.3: Summary of conserved contiguity for fish and mammalian *SPRN*. Figure is to scale as shown by rulers. *ECHS1*, enoyl-CoA hydratase/isomerase gene; *AO*, amine oxidase gene; *GTP*, GTP-binding protein gene; *SPRN*, Shadow of prion protein gene; *Or*, olfactory receptor gene; *SR*, speract/scavenger receptor gene; *FAE*, long-chain fatty-acyl elongase gene.

4.7 Expression of *SPRN*

Transcripts and ESTs deposited in databases indicate the range of tissues in which the *SPRN* gene is expressed.

My database searches found entries indicating expression of human, mouse, rat and zebrafish *SPRN* in the embryo, brain and retina, as summarized in Table 4.1. Evidence of expression is available at the TIGR database for mouse and rat *Sprn.* The mouse *Sprn* 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 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 brain (EST AA943106: 0.02% of the total library normalized rat brain, Lee et al., 1998, unpublished). The ESTs related to the rat contig TC293056 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 al., 2002, unpublished) and EST BG738569 is expressed in the adult retina library (Clark et al., 1998, unpublished).

To confirm this indicative database information Tatjana Simonic's group (University of Milan, Italy) demonstrated expression of the human, mouse and rat *SPRN* mRNAs in whole brain by RT-PCR. The *Sprn* mRNA expression was also assayed in different rat tissues, and fainter bands were detectable also in lung and stomach.

Both database and experimental data therefore indicate predominant *SPRN* expression in brain.

4.8 Discussion

I first describe the protein features of Sho and compare with those of PrP and its homologues. The features of vertebrate *SPRN*s are then discussed. Finally, the Sho, PrP and Dpl proteins are compared.

4.8.1 Protein Structure

Sho is the first human/mammalian protein known that, apart from PrP, contains remarkable middle hydrophobic sequence.

4.8.1.1 Major Features of Sho

Sho proteins have the following major features (Figure 4.1):
(a) An N-terminal peptide sequence (aa 1-24) comprising the 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 in *Fugu* there is an insertion of 14 residues in the middle of the region. This region ends with Arg-54 (zebrafish).
(c) A hydrophobic stretch in the middle of the protein (aa 55-74, zebrafish), which contains the same composition of aliphatic residues (G,A,V) as PrP and PrP homologues (Chapter 2.1).

(d) A C-terminal region with a putative N-glycosylation site (Asn-93, zebrafish) and a predicted GPI anchor site (Gly-106, zebrafish). Composition of this region could be compatible with folded structure. However, secondary structure predictions (not shown) revealed no consistent predictions.
(e) A C-terminal sequence predicted as a signal sequence for GPI-anchor attachment.

4.8.1.2 Conservation of Sho Sequences

Conservation among the mammalian (human, mouse, rat) sequences is high (identity 81-96%). Shos show good conservation from fish to mammals (Figure 4.1), particularly for zebrafish, to slightly beyond the end of the hydrophobic sequence (identity 41-53%, zebrafish 1-78 including 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. The sequences of the *Fugu* and zebrafish Shos show good conservation over the whole length (identity 54%, zebrafish 1-106, excluding C-terminal signal sequence).

4.8.1.3 Comparison of Overall Sho, PrP, stPrP and PrP-like Sequence Features

There are similarities between the Sho, PrP and PrP homologues in overall sequence features.

The overall protein structures of Sho, PrP, stPrP and PrP-like sequences (Figure 2.1 and Figure 4.1) indicate that all appear to be extracellular proteins attached to the outer leaflet of the cell membrane by GPI anchors. All also contain an internal hydrophobic segment (Figure 4.4). The alignment of this region shows strong conservation across all PrPs and Shos; 12 of the 20 residues (112-131, HuPrP) are identical, and another 6 are conserved hydrophobic. The sequence of Sho proteins is the same length as that of PrPs (except for *Xenopus* PrP) (Figure 4.4), whereas this region of stPrPs is 1-2 residues shorter and that of PrP-like proteins 4 residues shorter, this gap being positioned as for *Xenopus* PrP.

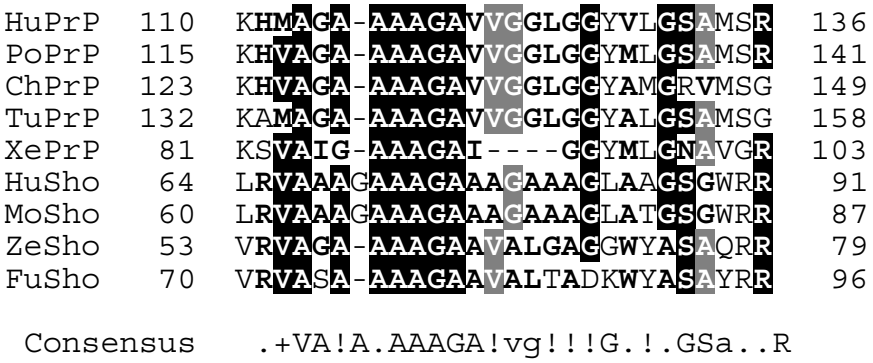

Figure 4.4: Alignment of hydrophobic regions of PrPs and Shos. In consensus line: capital letters, conserved in 7 or more species; lower-case letters, conserved in 6 or more species; !, conservatively conserved hydrophobic; +. conserved basic. Hu, human; Po, (marsupial) possum; Ch, chicken; Tu, turtle; Xe, *Xenopus*; Mo, mouse; Ze, zebrafish; Fu, *Fugu*.

The main regions of structural divergence between Shos and PrPs, and also PrP-likes (Suzuki et al., 2002), are the C-terminal region and the N-terminal repeat region. In contrast to the fish PrP-like proteins, the Sho C-terminal region sequence (spanning zebrafish Sho residues 75-106) is not low complexity. The Sho C-terminal sequences are also very different from those of PrPs and 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 (Mo et al., 2001), and from the stPrPs (Rivera-Milla et al., 2003; Oidtmann et al., 2003). The respective regions are shorter in Sho; human Sho 87-126 (40 residues) compared with human PrP 132-231 (100 residues). Whereas PrPs have two (or three) glycosylation 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) nonconserved sites (Oidtmann et al., 2003), Shos have only one potential glycosylation site.

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 or 9 (eutherian mammal), 9 or 10 (marsupial) or 6 or 7 (avian and turtle) residues flanked by the basic region (Chapter 6.3), but Shos have R-rich tetrarepeats starting at the beginning of the region. Zebrafish PrP-like, on the other hand, contains trirepeats of consensus HXG (where X is mostly T) inserted towards the end of the N-terminal region from residues 65-82.

Thus, the general structure of these proteins could be divided in four regions (Figure 2.1): the basic region 1, the repeat or low-complexity region 2, the hydrophobic region 3, and the C-terminal region 4 (Chapter 6).

4.8.2 Gene Structure and Expression

Vertebrate *SPRN*s all have the same gene organization, consisting of two exons and one intron. The complete ORF resides within the 3' terminal exon, which also is the case for the *PRNP* gene family (Chapters 5,6).

The database entries indicated expression of *SPRN* in the embryo, brain and retina of mouse and rat, in hippocampus of human, and in embryo and retina of zebrafish. Dr. Tatjana Simonic's laboratory used RT-PCR to confirm expression of the mammalian (human, mouse, rat) *SPRN* in whole brain, stomach and lung.

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 (Figure 4.3). Conservation of gene order and orientation for a three-gene block of an amine oxidase, GTP-binding protein and *SPRN* between pufferfish (two-gene block in zebrafish) and mammals, strongly indicates gene orthology (Gilligan et al., 2002).

4.8.3 Relationship Between Mammalian Sho, PrP and Dpl

Comparison of the Shos, PrPs and Dpls could assist in revealing their evolution and function.

Dpls have been reported only in mammals (Moore et al., 1999; Li et al., 2000; Mastrangelo and Westaway, 2001). Dpls show sequence homology with PrPs in the C-terminal domain, which, indeed, has been shown by NMR to have the same 3-D structure (Chapter 2.2.2), and they both are attached to the membrane by a GPI anchor. These two proteins are regarded as paralogues (Mastrangelo and Westaway, 2001), but there is no redundancy between the *PRNP* and *PRND* genes (Chapter 2.2.2). Dpls lack some sequence characteristics of PrP (basic region, repeats, hydrophobic region), and have different tissue expression. PrP co-exists with Dpl in sperm in the N-terminally

truncated form (Peoc'h et al., 2002) and in the C-terminally truncated form (Shaked et al., 1999; Shaked et al., 2002).

In contrast to Dpl, predominant brain expression, conserved hydrophobic region and similarity of sequence suggest that the mammalian Sho proteins may be good candidates for redundancy with mammalian PrPs (Chapter 5).

4.9 Conclusion

Experimental data and database analysis revealed the presence of a new vertebrate gene *SPRN* related to *PRNP*. The conserved contiguity, gene structure and protein sequence features between fish and mammals predict that *SPRN* is also present in other vertebrates.

The *Prnp* gene-ablated mice have no obvious phenotype due to the redundancy between *Prnp* and other gene(s) (Chapter 2.5.1). The Shadoo protein is at present the only human/mammalian candidate protein for this redundancy. My comparative genomic analysis (Chapter 5) further indicated this possibility.

Protein X, the auxilliary factor involved in pathogenic transformation of PrP is unknown (Chapter 1.2.4). PrP homologues, and thus Sho, could also be regarded as candidates for the protein X (Chapter 7).