Molecular Human Reproduction Vol.13, No.1 pp. 45–53, 2007
Advance Access publication November 4, 2006
doi:10.1093/molehr/gal098

Genomic structure and tissue-specific expression of human and mouse genes encoding homologues of the major bovine seminal plasma proteins

J.Lefebvre[1], J.Fan[1], S.Chevalier[2], R.Sullivan[3], E.Carmona[1] and P.Manjunath[1,4]

[1]Guy-Bernier Research Centre, Maisonneuve-Rosemont Hospital and Department of Medicine, University of Montreal, Montreal, [2]Department of Surgery, McGill University Health Center Research Institute, Montreal and [3]Centre de Recherche, Centre Hospitalier de l'Université Laval, Ste-Foy, Québec, Canada

[4]To whom correspondence should be addressed at: Centre de Recherche Guy-Bernier, Hôpital Maisonneuve-Rosemont, 5415 boulevard de l'Assomption, Montréal, Québec, Canada H1T 2M4. E-mail: puttaswamy.manjunath@umontreal.ca

Sperm capacitation is a maturation event that takes place in the female reproductive tract and is essential for fertilization. A family of phospholipid-binding proteins present in bovine seminal plasma (BSP proteins) binds the sperm membrane at ejaculation and promotes bovine sperm capacitation. Homologues of these proteins have also been isolated from boar, ram, goat, bison and stallion seminal fluid, suggesting that BSP proteins and their homologues are conserved among mammals. However, there have been no reports on BSP-homologous proteins in mice and humans to date. A search of the mouse and human genomes, using the nucleic acid sequences of BSP proteins, revealed the presence of three BSP-like sequences in the mouse genome, named *m*ouse *BSP H*omologue *1* (*mBSPH1*), *mBSPH2* and *mBSPH3*, and one sequence in the human genome (*hBSPH1*). Mouse epididymal expressed sequence tags corresponding to partial sequences of *mBSPH1* and *mBSPH2* were identified. The entire complementary DNA (cDNA) sequences of *mBSPH1* and *mBSPH2* from mouse epididymis and *hBSPH1* from human epididymis were obtained by 5′-/3′-rapid amplification of cDNA ends (RACE) and encode predicted proteins containing two tandemly repeated fibronectin type II domains, which is the signature of the BSP family of proteins. Using RT–PCR, it was revealed that *mBSPH1*, *mBSPH2* and *hBSPH1* mRNA are expressed only in the epididymis. Expression of *mBSPH3* was not detected in any tissue and probably represents a pseudogene. This work shows, for the first time, that BSP homologues are expressed in mouse and human and may be involved in sperm capacitation in these species.

Key words: BSP protein homologue/cDNA cloning/mRNA expression/epididymis/PDC-109

Introduction

Mammalian sperm acquire the capacity to fertilize an oocyte through a complex series of molecular modifications. After testicular spermatogenesis, immature sperm enter into the epididymis to undergo the first maturation events necessary for the acquisition of their fertilizing ability (Cooper, 1995). Numerous proteins are secreted by the epididymal epithelium and added to the sperm membrane at this stage, and several membrane proteins and receptors are exposed. The complex assortment of molecules found on the sperm head after maturation is crucial for sperm to undergo capacitation, reach the site of fertilization, recognize the oocyte, bind to oocyte surface receptors and finally fuse with the oocyte (Cooper, 1995; Jones, 1998).

Despite the maturation steps having occurred during epididymal transit, mammalian sperm emerging from the male reproductive tract are still incapable of fertilizing an oocyte. Complete fertilizing potential is acquired during a second maturation process that takes place in the female reproductive tract and is named capacitation (Austin, 1951; Chang, 1951). Despite years of investigations, sperm capacitation is still poorly understood. This multistep process involves several biochemical and ultrastructural changes in the sperm membrane (Yanagimachi, 1994). Sperm undergoing capacitation exhibit loss of adsorbed proteins originating from seminal plasma (SP), modification of membrane lipid composition, increased permeability to Ca^{2+}, increased intracellular pH, redistribution of surface proteins, changes in intramembranous particle distribution, increased sperm motility (hyperactivation), increased adenylate cyclase and cyclic-adenosine monophosphate (cAMP) and an increased tyrosine phosphorylation of a group of signalling proteins (for reviews, see de Lamirande *et al*., 1997; Visconti and Kopf, 1998). Only those sperm having completed capacitation in the proper time and place will be capable of undergoing the acrosome reaction and fertilizing an oocyte. In the bovine species, a family of SP proteins [bovine seminal plasma (BSP) proteins] bind to the sperm membrane upon ejaculation and are essential for sperm capacitation (Manjunath and Therien, 2002).

BSP proteins (BSP-A1, BSP-A2, BSP-A3 and BSP-30 kDa) are secreted by the seminal vesicle epithelial cells as part of the semen and constitute ~60% of total BSP proteins. BSP-A1 and -A2 differ only in their degree of glycosylation and are considered as one chemical entity, named BSP-A1/-A2 or PDC-109 (Esch *et al*., 1983). The biochemical properties and structure of these proteins have been thoroughly characterized (Manjunath *et al*., 1988). Each BSP protein contains two homologous type II domains (Fn2 domains) (Fan *et al*., 2006) that are responsible for the binding of these proteins to choline phospholipids, glycosaminoglycans (GAGs), collagen and gelatin (Manjunath *et al*., 1987, 1988, 2006). BSP proteins can also bind to

© The Author 2006. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

high- and low-density lipoproteins (Manjunath *et al*., 1989, 2002; Therien *et al*., 2001). The cloning and sequencing of the complementary DNA (cDNA) corresponding to the BSP proteins has been reported (Kemme and Scheit, 1988; Salois *et al*., 1999). Recently, three new BSP-related genes (*BSPH4*, *BSPH5* and *BSPH6*) were identified in the bull, and their expression in the seminal vesicles (*BSPH4*), epididymis and testis (*BSPH5* and *BSPH6*) was ascertained (Fan *et al*., 2006).

Homologues of BSP proteins have been characterized from the seminal fluid of boar, stallion, goat, bison and ram (reviewed in Manjunath *et al*., in press), indicating that they are conserved among mammals and probably share similar roles in sperm capacitation. However, there have been no reports to date revealing the presence of BSP family proteins in mice or humans. This study was aimed at characterizing BSP-homologous genes in these species.

Materials and methods

Materials

Human epididymis, testis and seminal vesicles were obtained through our local organ transplantation programme. Donors were 28–48 years of age with no medical pathologies affecting reproductive function. Tissues were collected while artificial circulation was maintained to preserve organs assigned for transplantation. Seminal vesicle tissue was obtained from patients undergoing prostatectomy by laparoscopy under general anaesthesia. Tissues were immediately sent to the laboratory, dissected, snap-frozen in liquid nitrogen and stored at –80°C until use. All procedures were approved by the ethical committee of Laval University. Human prostate was obtained with informed consent from benign prostatic hyperplasia patients undergoing laser resection of the prostate at the McGill University Health Centre.

Animal use, tissue collection and RNA extraction

Mice were maintained and handled according to the guidelines of the Animal Care Committee at the Guy-Bernier Research Centre. Mouse tissues were collected immediately after killing the mice and were snap-frozen in liquid nitrogen. Tissues were stored at –80°C until RNA extraction. Total RNA from human testis and caput epididymal tissues was isolated as explained by Legare *et al*. (1999). Human prostate RNA was prepared using the RNeasy Minikit (Qiagen, Mississauga, ON, Canada) according to the manufacturer's specifications. Human seminal vesicle and mouse tissue RNA were extracted using the Trizol™ Reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions.

Cloning of mBSPH1, mBSPH2 *and* hBSPH1 *cDNA*

To determine the internal sequences (Fn2 domains) of *mBSPH1* and *hBSPH1*, PCR products resulting from the amplification of mouse or human epididymal cDNA were excised from the agarose gel, purified using the Qiaex II gel purification kit (Qiagen) and sequenced using the Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems). PCR primers (Table I) were designed according to the partial Fn2 domain cDNA sequences obtained from a BLAST search of the mouse and human genomes using the BSP nucleic acid sequences. The missing 5′- and 3′-ends were obtained using the 5′- and 3′ Rapid Amplification of cDNA Ends (RACE) Systems (Invitrogen) according to manufacturer's instructions.

Briefly, for 5′-RACE of *mBSPH1*, first-strand cDNA was synthesized using total mouse epididymal RNA and *mBSPH1* gene-specific primer-1 (GSP1) (Table I). PCR amplification was performed using *mBSPH1* GSP2 (Table I) and the abridged anchor primer (AAP) (supplied in kit). The PCR conditions were as follows: one cycle at 94°C for 2 min; 35 cycles at 94°C for 45 s, 55°C for 45 s and 72°C for 1.5 min and one cycle at 72°C for 7 min. A nested amplification was performed using a nested *mBSPH1* PCR primer (Table I) and the abridged universal amplification primer (AUAP) (supplied in kit) under the same conditions. For 3′-RACE, first-strand cDNA was synthesized using total mouse epididymal RNA and the adapter primer (supplied in kit). PCR amplification was performed using *mBSPH1* GSP3 and AUAP primers under the same conditions.

For 5′-RACE of *hBSPH1*, first-strand cDNA was synthesized using total human epididymal RNA and *hBSPH1* GSP1 (Table I). cDNA was PCR-amplified using *hBSPH1* GSP2 and AAP primers, under identical PCR conditions as for *mBSPH1*. A semi-nested amplification was performed using the universal amplification primer (supplied in kit), under the same conditions except with only 30 cycles. For 3′-RACE, first-strand cDNA was synthesized using total human epididymal RNA and the adapter primer. A first PCR amplification was performed using *hBSPH1* GSP3 (Table I) and AUAP primers, under the same conditions as for *mBSPH1* 3′-RACE. A semi-nested PCR was performed using a nested *hBSPH1* primer and AUAP.

The open reading frame (ORF) and 3′-untranslated region of *mBSPH2* were obtained by sequencing a commercially available expressed sequence tag (EST) (ATCC #9075483). The 5′-end was obtained by 5′-RACE using the system described above. Briefly, first-strand synthesis was performed using total mouse epididymal RNA and *mBSPH2* GSP1 (Table I). cDNA was PCR-amplified using *mBSPH2* GSP2 (Table I) and AAP primers, under the same conditions described above. A semi-nested amplification was performed using *mBSPH2* GSP2 and the universal amplification primer, under the same PCR conditions except with only 30 cycles.

The final 5′- and 3′-RACE products were subcloned into pCR2.1 (Invitrogen) and sequenced. The ORFs of *mBSPH1*, *mBSPH2* and *hBSPH1* mRNA were identified using the ORF Finder tool from NCBI (http://www.ncbi.nlm.nih.gov/projects/gorf/). The cDNA sequences of *mBSPH1*, *mBSPH2* and *hBSPH1* were deposited into GenBank under accession numbers DQ227498, DQ227499 and DQ227497, respectively. The full cDNA sequences of each gene were compared with the genomic sequences using the BLAST2 algorithm, and each pairwise match was then mapped onto the chromosome. Intron/exon boundaries were determined by the exact match of the

Table I. Oligonucleotide primers used in this study

Name	Sequence (5′–3′)	Experimental use
mBSPH1-F/mBSPH1-GSP3	AGATGGTGCATGTGTCTTTCC	RT–PCR/3′-RACE
mBSPH1-R/mBSPH1-GSP2	CTCATCACTCTATACAATATTTCC	RT–PCR/5′-RACE
mBSPH1-GSP1	CTCTCTACAACCAGTTATCTTC	5′-RACE
mBSPH1-nested	CACCTCTCCATCCTCTGTGCAATC	5′-RACE
mBSPH2-F	CCAAAGTGTATTTTCCCTTTCC	RT–PCR
mBSPH2-R/mBSPH2-GSP1	CATTTGGAGGAACATCTGATGTATC	RT–PCR/5′-RACE
mBSPH2-GSP2	CTAAAAATTGTTAGGAGAACATTGC	5′-RACE
hBSPH1-F/hBSPH1-GSP3	AGATGGGGAGTGTGTCTTTCC	RT–PCR/3′-RACE
hBSPH1-R/hBSPH1-GSP1	CAGTATTTCCAAATTCGGTCC	RT–PCR/5′-RACE
hBSPH1-GSP2	CATCATCAGTACACTCCCAGTAG	5′-RACE
hBSPH1-nested	GACTGCATCAAGTCCAAGGCAAGAC	3′-RACE
β-Actin-F (mouse)	CCCCCTGAACCCTAAGGCCA	RT–PCR
β-Actin-R (mouse)	TCCCTCTCAGCTGTGGTGGT	RT–PCR
γ-Actin-F (human)	CCTGAAGTACCCCATTGAGC	RT–PCR
γ-Actin-R (human)	GTTGGCGTACAGGTCTTTGC	RT–PCR

cDNA with the genomic sequences and the consensus boundary (GT/AG) (Shapiro and Senapathy, 1987).

Analysis of predicted protein sequences

The presence of potential signal peptides within the mBSPH1, mBSPH2 and hBSPH1 protein sequences as well as the expected cleavage sites were examined using the SignalP 3.0 software (Bendtsen *et al*., 2004). A search for potential O-glycosylation sites was performed using NetOGlyc 3.1 (Julenius *et al*., 2005). To compare sequence similarity, we aligned the amino acid sequences with those of BSP-A1/-A2 and BSP-A3 (GenBank numbers: P02784 and P04557) using the ClustalW method (Thompson *et al*., 1994). Similarity was calculated according to the PAM250 matrix. The sequences of mBSPH1, mBSPH2 and hBSPH1 were submitted to the SWISS-MODEL server (Automated Comparative Protein Modeling Server, version 3.0) (Schwede *et al*., 2003) for comparative protein structure modelling. All homology models were generated based on the template of BSP-A1/-A2 (PDC-109; PDB accession number: 1H8P) using Swiss-Pdb Viewer 3.7 (Guex and Peitsch, 1997; Schwede *et al*., 2003) and MOLMOL 2K.2 (Koradi *et al*., 1996).

RT–PCR

RT–PCRs were performed with the Superscript-III First-Strand Synthesis System (Invitrogen) according to manufacturer's protocol. Briefly, 2 μg of total RNA from each tissue was treated with DNase (New England BioLabs, Beverly, MA, USA) and reverse transcribed. Two microlitres of the first-strand reaction was used as a template for PCR amplification with gene-specific primers. The *mBSPH1*, *mBSPH2* and *hBSPH1* gene fragments were amplified using primers designed as described above according to the partial Fn2 domain cDNA sequences (Table I). A mouse *β-actin* gene fragment was amplified as a control for *mBSPH1* and *mBSPH2*, whereas a human *γ-actin* gene fragment was amplified as a control for *hBSPH1* using the primers described in Table I. PCR conditions were as follows: one cycle at 94°C for 3 min; 35 cycles at 94°C for 45 s, 55°C for 45 s and 72°C for 1 min and one cycle at 72°C for 7 min. The RT–PCR products were analysed in 1.5% agarose gels containing ethidium bromide.

Results and discussion

Presence of BSP-homologous sequences in the mouse and human genomes

BSP proteins play a crucial role in bovine sperm capacitation, and homologues of these proteins have been isolated and characterized from the seminal fluid of numerous species. Quite recently, we performed an extensive bioinformatics analysis of all proteins containing Fn2 domains in fully or partially sequenced genomes of several mammalian species and found that those contained within BSP-related proteins are unique compared with other Fn2-containing proteins, thus allowing the identification of many yet unidentified BSP-related sequences (Fan *et al*., 2006).

The mouse genome was searched for BSP-related sequences, allowing the identification of three BSP-related genes on chromosome 7, designated *m*ouse *BSP H*omologue *1* (*mBSPH1*), *mBSPH2* and *mBSPH3*. In a similar fashion, the recently updated human genome shotgun assembly (Istrail *et al*., 2004) was searched, yielding one hit with a score of 47% identity in predicted amino acid sequence. The human BSP-homologous gene was named *hBSPH1* and is orthologous to *mBSPH1* (Fan *et al*., 2006). Based on the mapping data from the shotgun assembly, the *hBSPH1* gene is located in loci along chromosome 19.

To verify whether the newly identified sequences are actively transcribed genes, we searched the EST database using the mouse BSP-homologous gene sequences and identified several ESTs of epididymal origin (GenBank accession numbers: BY721041 and BB73190 for *mBSPH1*; and BY721134, AV381075, BB073010, BU961078 and AV379540 for *mBSPH2*). However, there were no ESTs sharing sequence similarity with *mBSPH3*. The epididymal expression of *mBSPH1* and *mBSPH2* was confirmed by RT–PCR although no expression was detected for *mBSPH3* (discussed below), which seemingly represents a pseudogene.

Full cDNA sequences of mBSPH1, mBSPH2 *and* hBSPH1

The full-length mRNAs encoded by *mBSPH1* and *hBSPH1* were obtained by RT–PCR of epididymal RNA followed by 5′- and 3′-RACE and are shown in Figure 1A and C. The *mBSPH1* cDNA sequence (DQ227498) is 716 bp in length with an ORF of 402 bp (including the stop codon), encoding a predicted protein of 133 amino acids. In the case of *hBSPH1* (Figure 1C), the entire cDNA (DQ227497) spans 654 bp, with a complete ORF of 399 bp (including the stop codon), coding for a predicted protein of 132 amino acids. The cDNA sequence of *mBSPH2* (Figure 1B) was obtained by sequencing a commercially available EST, which allowed the determination of the entire ORF as well as the 3′-untranslated region, whereas the sequence of the 5′-untranslated region was obtained by 5′-RACE. The complete cDNA sequence of *mBSPH2* (DQ227499) is 530 bp in length with an ORF of 396 bp, encoding a protein of 131 amino acids.

Intron/exon organization of the mouse and human BSP-homologous genes

The mouse *mBSPH1* gene spans 24 kb of chromosome 7 and consists of five exons and four introns. A similar organization was observed for *mBSPH2*, also on chromosome 7, which contains five exons and four intronic sequences spanning ~21 kb. In the case of *hBSPH1*, the 26-kb genomic DNA sequence found on chromosome 19 encompasses six exons and five introns. The intron/exon splice sites as well as the sizes of each intron and exon for *mBSPH1*, *mBSPH2* and *hBSPH1* are indicated in Table II. All intron–exon boundaries were consistent with the GT/AG rule for eukaryotic splice junctions (Shapiro and Senapathy, 1987).

The non-coding regions of the *mBSPH1*, *mBSPH2* and *hBSPH1* genes were analysed by the blastx and blastn programs, revealing that introns in *mBSPH1* (introns 2 and 3) and in *mBSPH2* (intron 2) encoded a reverse transcriptase that may have originated from L1-retrotransposable elements (Shehee *et al*., 1987; Martin, 1995; Goodier *et al*., 2001). In addition, homologous fragments of the *Plasmodium yoelii yoelii* hypothetical protein PY07367 coding sequence (XP_728215) (Carlton *et al*., 2002) were also evident within intron 1 of both *mBSPH1* and *mBSPH2*, a situation that may have resulted from the mobilization of retrotransposons or from horizontal gene transfer from the rodent malaria parasite DNA. The introns in the human *hBSPH1* gene were rich in *Alu* repeat elements (Jurka and Milosavljevic, 1991; Claverie and Makalowski, 1994). These mobile and repeat elements, as the main components of the mouse and human genomes, may play a role in the evolution of BSP-homologous genes.

Chromosomal mapping of the mouse and human BSP-homologous genes

Mapping of the mouse BSP-homologous genes revealed that the *mBSPH1* and *mBSPH2* coding strands are arranged in a 'head-to-head' orientation on chromosome 7, suggesting that the two genes may share a common promoter and/or regulatory elements (Doerwald *et al*., 2004; Trinklein *et al*., 2004). Moreover, the human and mouse BSP-homologous genes map to syntenic segments of their respective genomes, which signifies that orthologous genes are present in the same order and indicates a common evolutionary origin. The BSP-homologous genes are found in large syntenic segments of the chromosomes, which share the same gene order although there are differences in the spacing between each gene. Because there is but a single BSP-homologous

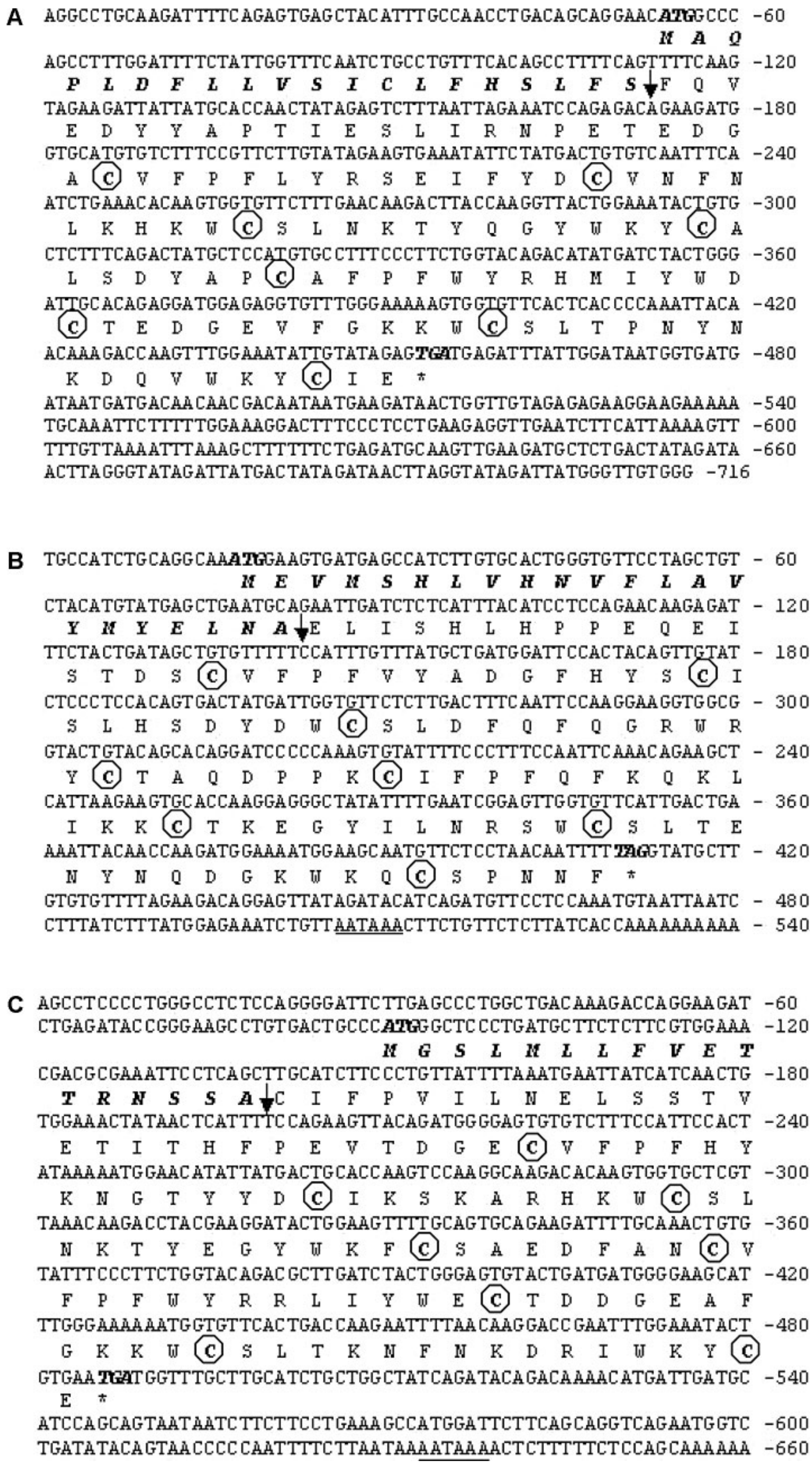

Figure 1. Complementary DNA (cDNA) and deduced amino acid sequences of the mouse and human bovine seminal plasma (BSP protein) homologues. The cDNA sequences of (**A**) *mBSPH1*, (**B**) *mBSPH2* and (**C**) *hBSPH1* were obtained as described in the *Methods* section. The predicted signal peptide is indicated in bold italic letters and the position of the predicted cleavage site is marked by an upside-down arrow. The eight characteristic cysteine residues constituting the two Fn2 domains are circled.

Table II. Nucleotide sequences of the exon/intron boundaries of the *mBSPH1*, *mBSPH2* and *hBSPH1* genes

Exon	Size (bp)	Exon–intron junction		
		5′-splice sequence	Intron size (bp)	3′-splice sequence
mBSPH1				
				actggatgccAGGCCTGCAA
1	124	CAAGTAGAAG**gt**aagggatg	7223[a]	tgttttc**ag**ATTATTATGC
2	21	CCAACTATAG**gt**gagtgtgt	12 026	aatttttc**ag**AGTCTTTAAT
3	30	GAGACAGAAG**gt**aagttgat	1824	gttcttt**ag**ATGGTGCATG
4	132	GCTCTTTCAG**gt**aagtgctt	689	ccttttac**ag**ACTATGCTCC
5	408	GGGTTGTGGGatttaagatg		
mBSPH2				
				ctcctgtgtgTGCCATCTGC
1	86	CTGAATGCAG**gt**gaggtgaa	10 155[a]	tgtctttc**ag**AATTGATCTC
2	33	CCAGAACAAG**gt**aagcaacc	4479	tgttttcc**ag**AGATTTCTAC
3	141	ACAGCACAGG**gt**aagaatcc	217	ttctctgc**ag**ATCCCCCAAA
4	148	CTCCTAACAA**gt**aagacaat	1359	ttcacctc**ag**TTTTTAGGTA
5	122	TTCTCTTATCaacattgatg		
hBSPH1				
				ttttccaagaAGCCTCCCTG
1	161	GTTATTTTAA**gt**atttcttt	11 067	tgtttttc**ag**ATGAATTATC
2	21	TCAACTGTGG**gt**gagttgat	1321	tatttttc**ag**AAACTATAAC
3	30	GAAGTTACAG**gt**aagtcgat	2065	ttttctct**ag**ATGGGGAGTG
4	132	AGTGCAGAAG**gt**gagtgtcc	518	tcctctgc**ag**ATTTTGCAAA
5	145	GTGAATGATG**gt**gagattta	8500	tttgggtt**ag**GTTTGCTTGC
6	164	TTTCTCCAGCatttctagca		

The nucleotide sequence of each exon/intron boundary and the sizes of the exon and intron are shown. Exon sequences are in upper case letters; intron sequences are in lower case letters and the boundary-conserved nucleotides are underlined.
[a]The intron contains ambiguous nucleotides or an unsequenced region.

gene in humans, it is possible that the mouse *mBSPH2* and *mBSPH3* sequences arose from gene duplication. Our previous studies indicated that there is also a single BSP-homologous gene in the genomes of other primates such as chimpanzee and monkey (Fan *et al*., 2006). The syntenic mapping of the human and mouse BSP-homologous genes suggests that the mouse may be a good model to study the functions of human genes in reproductive biology.

Putative signal peptides in the mBSPH1, mBSPH2 and hBSPH1 predicted proteins

The cDNA sequences determined for *mBSPH1*, *mBSPH2* and *hBSPH1* were translated and the predicted protein sequences are also indicated in Figure 1. In bovine, the BSP protein sequences contain signal peptides targeting the proteins for secretion. Similarly, BSP-homologues identified in other species also are secreted proteins. Using the SignalP 3.0 server, it was revealed that the mBSPH1, mBSPH2 and hBSPH1 protein sequences contain 20-, 22- and 17-amino acid predicted signal peptides, respectively, at the N-terminus of the immature proteins (expected cleavage sites indicated in Figure 1). The predicted amino acid structures strongly suggest the presence of two Fn2 domains, which are responsible for the binding of BSP proteins to the sperm membrane (Manjunath *et al*., 1994) as well as to other extracellular ligands such as lipoproteins (Manjunath *et al*., 1989, 2002) and GAGs (Therien *et al*., 2005). Consequently, mBSPH1, mBSPH2 and hBSPH1 should be secreted proteins.

Molecular characteristics of the mBSPH1, mBSPH2 and hBSPH1 predicted proteins

After cleavage of the putative signal peptide of mBSPH1, the mature protein would contain 113 amino acids (including the eight cysteines characteristic of the BSP-family Fn2 domains), with a predicted molecular weight of 13.8 kDa and an isoelectric point (pI) of 5.12 (Figure 1A). Mature mBSPH2 would be a 109-residue protein, with a calculated molecular weight of 12.8 kDa and a pI of 6.46, and would also contain the eight characteristic cysteine residues (Figure 1B). In the case of the human protein, the predicted length of the mature protein is 115 amino acids, accounting for a theoretical molecular weight of 13.8 and a pI of 8.10. The porcine BSP-homologue, pB1, is also a basic protein with a pI of 8.6 (Jonakova *et al*., 1998). Mature hBSPH1 contains nine cysteine residues (Figure 1C), in contrast to bovine, porcine and murine BSP proteins, which contain eight. The N-terminal cysteine is expected to be free as the eight others are found within the Fn2 domains and would therefore be participating in intramolecular disulphide bridges. The free cysteine may participate in intermolecular disulphide bridges with other hBSPH1 molecules or with distinct proteins. Alternatively, despite the predictions, the N-terminal cysteine may not be part of the mature protein.

Analysis of potential O-glycosylation of mBSPH1, mBSPH2 and hBSPH1

BSP-A1, BSP-A2 and BSP-30 kDa are glycoproteins, whereas BSP-A3 does not contain any carbohydrate (Manjunath and Sairam, 1987; Manjunath *et al*., 1988; Calvete *et al*., 1996). Porcine pB1 and equine SP-1 are also glycosylated proteins (Calvete *et al*., 1995, 1997). In all cases, the carbohydrate linkage was shown to be O-glycosidic. Therefore, we analysed the predicted protein sequences of the three new BSP homologues for potential O-glycosylation sites. Of the 10 serine/threonine residues in mature mBSPH1, the 14 Ser/Thr residues found in mBSPH2 and the 14 Ser/Thr residues in hBSPH1, none displayed significant scores with respect to O-glycosylation. The BSP homologues from mice and human are therefore not expected to be glycosylated.

mBSPH1, mBSPH2 and hBSPH1 display features characteristic of the BSP family

To compare the amino acid sequences, we aligned the sequences of mBSPH1, mBSPH2 and hBSPH1 with those of bovine BSP-A1/-A2 and BSP-A3 (Figure 2A). The newly identified BSP-homologues share many conserved motifs with the bovine BSP proteins, especially surrounding the cysteine residues. The -C-X-F-P-F- motif (where X is usually valine or another non-polar amino acid), found at the first cysteine residue of each Fn2 domain, is characteristic of BSP-family proteins (Esch *et al*., 1983; Seidah *et al*., 1987; Calvete *et al*., 1996, 1997; Plucienniczak *et al*., 1999; Villemure *et al*., 2003; Bergeron *et al*., 2005). As seen in Figure 2A, the -W-C-S-L- motif, which borders the third cysteine of each Fn2 domain, is preserved in the mouse and human proteins. In addition, the -W-(K/R)-Y-C- motif, which surrounds the fourth cysteine of each Fn2 domain, is found in the new homologues, except for a slight change in the second Fn2 domain of mBSPH2 and in the first Fn2 domain of hBSPH1. The conserved tryptophan and tyrosine residues are thought to be important for forming the hydrophobic cluster that lines the phosphorylcholine binding pocket, as was shown for BSP-A1/-A2 (PDC-109) (Wah *et al*., 2002). Many other single amino acids are also conserved throughout the sequences (Figure 2A).

A schematic representation of the domain organization of the mouse and human BSP-homologous proteins is shown in Figure 2B. The mature proteins are predicted to have a variable N-terminal extension followed by Fn2 domain A, a short linker peptide, Fn2 domain B and a variable C-terminal extension. This domain repartition is strikingly similar to that of the bovine BSP proteins, except for the absence of a C-terminal extension in the bovine proteins. In addition, with the exception of the variable N-terminal domain, all domains of the mouse and human homologues are composed of the same number of amino acids as those found in the bovine proteins (Fn2A, 38 residues linker, 7 residues and Fn2B, 42 residues). The boar (pB1; NP_998997) and stallion (SP1, SP2 and CAE46515, CAE46517) BSP homologues also share a highly similar domain organization.

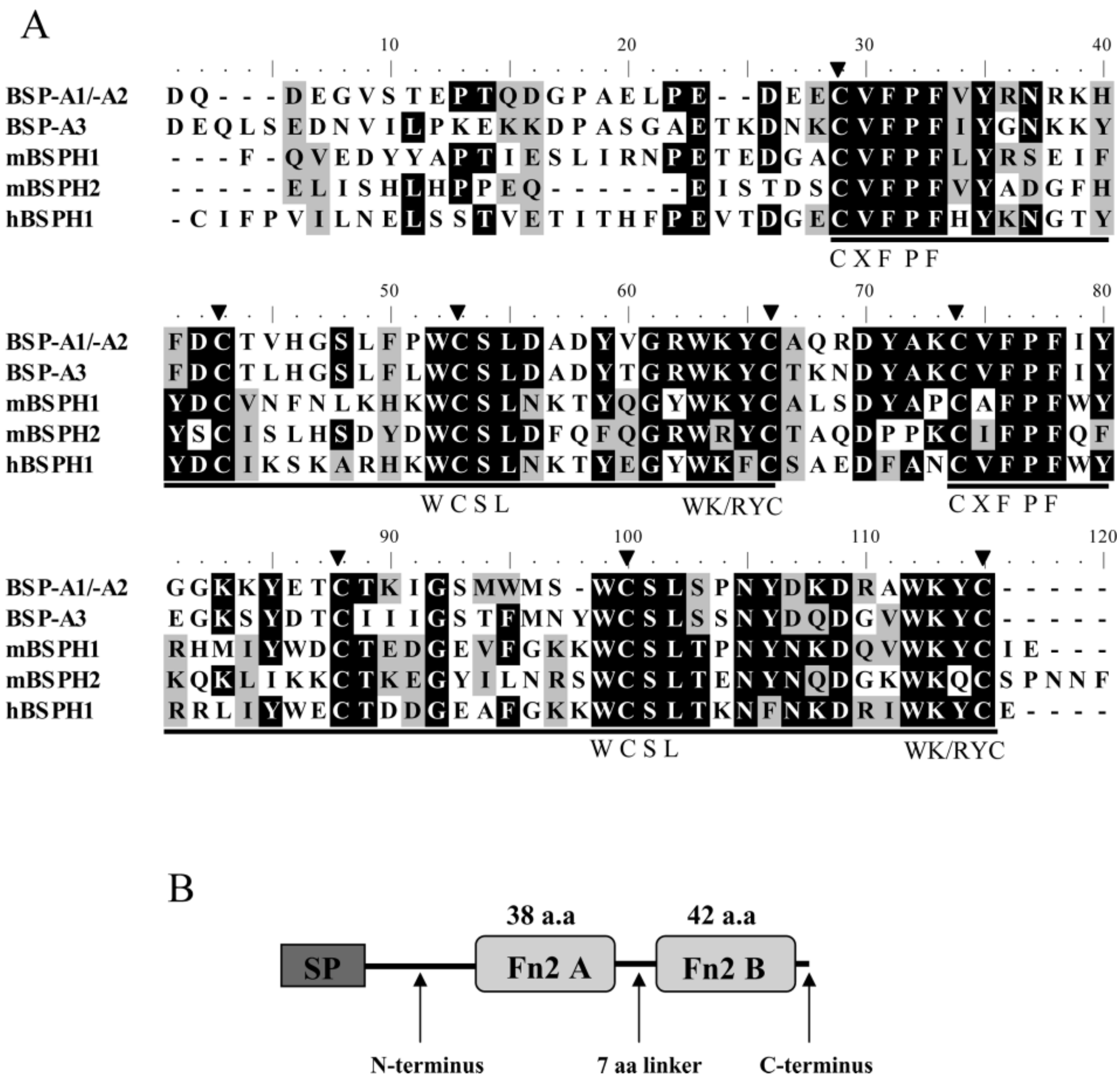

Figure 2. Analysis of the mBSPH1, mBSPH2 and hBSPH1 predicted proteins. (**A**) Relationships between bovine seminal plasma (BSP) proteins and the mouse and human BSP homologues. The amino acid sequences of each new BSP homologue were deduced from their cDNA sequences and aligned with the sequences of BSP-A1/-A2 (P02784) and BSP-A3 (P04557). Upside-down arrowheads indicate the four cysteine residues in each Fn2 domain. Black solid lines underline the sequences corresponding to the Fn2 domains. Identical amino acids are highlighted in black, whereas similar amino acids are highlighted in grey. Characteristic motifs conserved throughout the BSP family are indicated below the alignment. (**B**) Schematic representation of the domain structure of the mBSPH1, mBSPH2 and hBSPH1 predicted proteins. SP, signal peptide; Fn2A, first Fn2 domain; Fn2B, second Fn2 domain and a.a, amino acid.

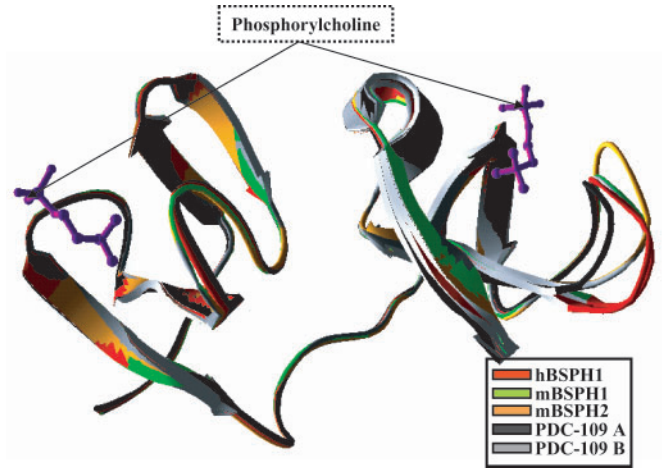

Figure 3. Ribbon representation of homology models of the mouse and human bovine seminal plasma (BSP protein) homologues. A potential ligand, phosphorylcholine, is represented in ball and stick structure. The figure was prepared with Swiss-PbdViewer 3.7b2 (Guex and Peitsch, 1997) and MOLMOL 2K.2 (Koradi *et al*., 1996). The BSP-A1/-A2 (PDC-109; PDB accession number, 1H8P) template is from Wah *et al*. (2002), in which chains PDC-109 A and PDC-109 B correspond to two monomers because this protein forms dimers when associated with phosphorylcholine.

To visualize the three-dimensional organization of the predicted mouse and human BSP-homologues, we created homology models of mBSPH1, mBSPH2 and hBSPH1 (Figure 3) based on the available crystal structure of BSP-A1/-A2 (PDC-109; PDB accession number, 1H8P) (Wah *et al*., 2002). Phosphorylcholine, a known ligand of BSP proteins, is shown in the figure within the presumed choline-binding pocket. The important sequence similarity together with the identical domain and three-dimensional organization indicates that the newly identified proteins from mouse and human belong to the BSP family.

The mouse and human BSP homologues share strong sequence similarity with the bovine proteins

The per cent identity and similarity among the new homologues as well as between these proteins and bovine BSP-A1/-A2 and BSP-A3 were calculated and are presented in Table III. mBSPH1 shares 35–40% identity and 53–55% similarity with the bovine proteins and with mBSPH2. It is most similar to hBSPH1 (56% identity and 78% similarity), which was expected as *mBSPH1* and *hBSPH1* are orthologous genes (Fan *et al*., 2006). Similarly, hBSPH1 shares 33–41% identity and 54–56% similarity with the bovine proteins and mBSPH2 but is most similar to its orthologue, mBSPH1. mBSPH2 shares 33–37% identity and 50–56% similarity with the bovine, mouse and human proteins. When these calculations are performed on the sequences of the Fn2 domains (excluding the variable N- and C-terminal sequences), the values obtained are on average 10 points higher than those calculated for the entire proteins (data not shown). The Fn2 domains of hBSPH1 are 47–48% identical in amino acid sequence to BSP-A1/-A2 and BSP-A3, which is the value that was obtained when the initial search of the human genome was performed. The high level of conservation within the Fn2 domains suggests a conservation of the characteristic binding properties of BSP proteins, which are conferred by the Fn2 domains, and may also indicate shared biological functions.

Table III. Comparison of the degree of identity and similarity between the new mouse and human BSP-homologues and the bovine seminal plasma (BSP) proteins

	BSP-A1/-A2	BSP-A3	mBSPH1	mBSPH2	hBSPH1
BSP-A1/-A2	–	75	53	54	54
BSP-A3	61	–	53	50	55
mBSPH1	43	38	–	55	78[a]
mBSPH2	37	36	34	–	56
hBSPH1	41	38	56	33	–

Predicted protein sequences were aligned according to ClustalW (Thompson *et al*., 1994) and identities and similarities were calculated. The similarity matrix used was PAM250. Values for per cent identity are on the left side of the table below the diagonal, and those for per cent similarity are indicated on the right side above the diagonal.
[a]Highest similarity.

Reproductive tissue-specific expression of mouse and human BSP-homologous mRNA

To assess the expression pattern of the BSP-homologous genes, we performed RT–PCR with RNA isolated from mouse and human tissues. Using the genomic sequences obtained in the BLAST search described above, we designed primers for PCR amplification of the new mouse and human BSP-homologous genes, within the predicted exons (exons 3 and 4 for *mBSPH1* and *mBSPH2* and exons 4 and 5 for *hBSPH1*) encoding the Fn2 domains (Table I). As shown in Figure 4A, expression of *mBSPH1* and *mBSPH2* mRNA was detected solely in the mouse epididymis. As expected, our results indicate that *mBSPH3* is not an actively transcribed gene, because expression was not detected in any tissue using several different primer pairs and numerous experimental conditions (data not shown).

Because the bioinformatics analysis indicated that *hBSPH1* is orthologous to *mBSPH1*, we predicted the expression of this gene in the human epididymis. Indeed, *hBSPH1* mRNA is expressed in the human epididymis and also weakly in testis but not in prostate or seminal vesicles (Figure 4B). No expression was detected in brain, intestine or lung; therefore, results suggest the restriction of expression to the male reproductive tract. However, because of difficulties in obtaining human material, other tissues were not yet examined. For all RT–PCR experiments, samples for which the RT reaction was performed without

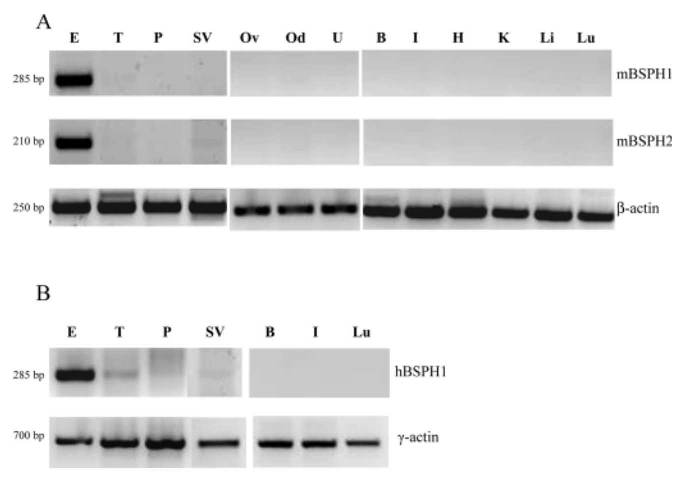

Figure 4. mRNA expression analysis of the mouse and human bovine seminal plasma (BSP protein) homologues. Total RNA from mouse or human tissues was prepared as described in the *Methods* section and subjected to RT–PCR. (**A**) Expression of *mBSPH1* and *mBSPH2*. (**B**) Expression of *hBSPH1*. The expression of the mouse *β-actin* or the human *γ-actin* gene was used as an internal control. B, brain; E, epididymis; H, heart; I, intestine; K, kidney; Li, liver; Lu, lung; Od, oviduct; Ov, ovary; P, prostate; SV, seminal vesicles; T, testis; U, uterus.

reverse transcriptase served as a negative control (data not shown). The PCR products, corresponding to the partial transcripts of the *mBSPH1*, *mBSPH2* and *hBSPH1* genes, were confirmed by DNA sequencing, as described in the *Methods* section.

A recent molecular evolutionary analysis revealed that all BSP-related sequences could be grouped into three subfamilies: BSPH4, which is expressed in seminal vesicles, and BSPH5 and BSPH6, which are expressed in the epididymis and testis (Fan *et al*., 2006). The BSPH4 subfamily includes BSP-A1/-A2, BSP-A3 and BSP-30 kDa, which are all expressed in the seminal vesicles. Human *hBSPH1* and mouse *mBSPH1* are included within the BSPH5 subfamily, and mouse *mBSPH2* is in the BSPH6 subfamily. Thus, the expression of these genes in the epididymis is consistent with the phylogenetic predictions.

Potential biological functions for the mouse and human BSP homologues

In the bovine species, BSP proteins are intimately involved in the process of sperm capacitation (reviewed in Manjunath and Therien, 2002), and BSP homologues from other mammals are also believed to play similar roles. This study has identified BSP-homologous genes in mice and human, the mRNA of which are expressed in the epididymis, differing from the seminal vesicle expression seen in other species. This may be due to species-specific differences in sperm maturation, because, in contrast to bull semen, mouse and human semen are known to coagulate after ejaculation, in which case sperm coating by SP proteins would be highly inefficient. Our preliminary experiments indicate almost undetectable levels of BSP-homologous antigens in human SP, consistent with the idea that the contact between sperm and hBSPH1 would be taking place inside the epididymis, where a small number of sperm reside for a long period. Thus, organisms that express BSPs in the epididymis need not synthesize them in excess to coat the sperm surface, unlike in other species where large amounts of BSP homologues are produced by the seminal vesicles and added to sperm at ejaculation.

On the contrary, the mouse and human BSP homologues may fulfil somewhat different biological functions than those exerted by BSP proteins from other species. Numerous studies have shown that proteins secreted by the epididymis associate with the sperm membrane during epididymal transit and confer the ability to interact or fuse with the oocyte (Sullivan, 1999; Cohen *et al*., 2001; Weerachatyanukul *et al*., 2003). BSP-homologous proteins may also be added to sperm during epididymal transit and remain there until sperm enter the female reproductive tract and are ready to undergo capacitation, in which case BSP homologues in mice and human could fulfil a role similar to that of the bovine BSP proteins.

In summary, we have shown for the first time that mice and humans also express homologues of the major BSP proteins. The full cDNA sequences and genomic structure of the mouse and human BSP homologues were determined and mapped to syntenic segments of mouse chromosome 7 and human chromosome 19. Expression of these genes seems to be restricted to tissues of the male reproductive tract, namely the epididymis, strongly suggesting a role in sperm maturation. On the basis of the known role of BSP proteins in sperm capacitation, a similar role is predicted for the mouse and human BSP homologues. Further studies are underway to characterize these proteins and confirm their biological implications.

Acknowledgements

This work was funded by the Canadian Institutes of Health Research. J.L. is a recipient of a Doctoral Training Award from the Fonds de la Recherche en Santé du Québec. We thank Dr Yves Paquette and Josée Tessier for their advice in troubleshooting, as well as Dr Lucie Hamel for the isolation of human prostate RNA.

References

Austin CR (1951) Observations on the penetration of the sperm in the mammalian egg. Aust J Sci Res (B) 4,581–596.

Bendtsen JD, Nielsen H, von Heijne G and Brunak S (2004) Improved prediction of signal peptides: signalP 3.0. J Mol Biol 340,783–795.

Bergeron A, Villemure M, Lazure C and Manjunath P (2005) Isolation and characterization of the major proteins of ram seminal plasma. Mol Reprod Dev 71,461–470.

Calvete JJ, Mann K, Schafer W, Sanz L, Reinert M, Nessau S, Raida M and Topfer-Petersen E (1995) Amino acid sequence of HSP-1, a major protein of stallion seminal plasma: effect of glycosylation on its heparin- and gelatin-binding capabilities. Biochem J 310,615–622.

Calvete JJ, Mann K, Sanz L, Raida M and Topfer-Petersen E (1996) The primary structure of BSP-30K, a major lipid-, gelatin-, and heparin-binding glycoprotein of bovine seminal plasma. FEBS Lett 399,147–152.

Calvete JJ, Raida M, Gentzel M, Urbanke C, Sanz L and Topfer-Petersen E (1997) Isolation and characterization of heparin- and phosphorylcholine-binding proteins of boar and stallion seminal plasma. Primary structure of porcine pB1. FEBS Lett 407,201–206.

Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL *et al*. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature 419,512–519.

Chang MC (1951) Fertilizing capacity of spermatozoa deposited into the fallopian tubes. Nature 168,697–698.

Claverie JM and Makalowski W (1994) Alu alert. Nature 371,752.

Cohen DJ, Ellerman DA, Busso D, Morgenfeld MM, Piazza AD, Hayashi M, Young ET, Kasahara M and Cuasnicu PS (2001) Evidence that human epididymal protein ARP plays a role in gamete fusion through complementary sites on the surface of the human egg. Biol Reprod 65,1000–1005.

Cooper TG (1995) Role of the epididymis in mediating changes in the male gamete during maturation. Adv Exp Med Biol 377,87–101.

Doerwald L, van Rheede T, Dirks RP, Madsen O, Rexwinkel R, van Genesen ST, Martens GJ, de Jong WW and Lubsen NH (2004) Sequence and functional conservation of the intergenic region between the head-to-head genes encoding the small heat shock proteins alphaB-crystallin and HspB2 in the mammalian lineage. J Mol Evol 59,674–686.

Esch FS, Ling NC, Bohlen P, Ying SY and Guillemin R (1983) Primary structure of PDC-109, a major protein constituent of bovine seminal plasma. Biochem Biophys Res Commun 113,861–867.

Fan J, Lefebvre J and Manjunath P (2006) Bovine seminal plasma proteins and their relatives: a new expanding superfamily in mammals. Gene 375,63–74.

Goodier JL, Ostertag EM, Du K and Kazazian HH, Jr. (2001) A novel active L1 retrotransposon subfamily in the mouse. Genome Res 11,1677–1685.

Guex N and Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18,2714–2723.

Istrail S, Sutton GG, Florea L, Halpern AL, Mobarry CM, Lippert R, Walenz B, Shatkay H, Dew I, Miller JR *et al*. (2004) Whole-genome shotgun assembly and comparison of human genome assemblies. Proc Natl Acad Sci USA 101,1916–1921.

Jonakova V, Kraus M, Veselsky L, Cechova D, Bezouska K and Ticha M (1998) Spermadhesins of the AQN and AWN families, DQH sperm surface protein and HNK protein in the heparin-binding fraction of boar seminal plasma. J Reprod Fertil 114,25–34.

Jones R (1998) Plasma membrane structure and remodelling during sperm maturation in the epididymis. J Reprod Fertil Suppl 53,73–84.

Julenius K, Molgaard A, Gupta R and Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 15,153–164.

Jurka J and Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32,105–121.

Kemme M and Scheit KH (1988) Cloning and sequence analysis of a cDNA from seminal vesicle tissue encoding the precursor of the major protein of bull semen. DNA 7,595–599.

Koradi R, Billeter M and Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14,29–32.

de Lamirande E, Leclerc P and Gagnon C (1997) Capacitation as a regulatory event that primes spermatozoa for the acrosome reaction and fertilization. Mol Hum Reprod 3,175–194.

Legare C, Gaudreault C, St-Jacques S and Sullivan R (1999) P34H sperm protein is preferentially expressed by the human corpus epididymidis. Endocrinology 140,3318–3327.

Manjunath P and Sairam MR (1987) Purification and biochemical characterization of three major acidic proteins (BSP-A1, BSP-A2 et BSP-A3) from bovine seminal plasma. Biochem J 241,685–692.

Manjunath P and Therien I (2002) Role of seminal plasma phospholipid-binding proteins in sperm membrane lipid modification that occurs during capacitation. J Reprod Immunol 53,109–119.

Manjunath P, Sairam MR and Uma J (1987) Purification of four gelatin-binding proteins from bovine seminal plasma by affinity chromatography. Biosci Rep 7,231–238.

Manjunath P, Baillargeon L, Marcel YL, Seidah NG, Chretien M and Chapdelaine A (1988) Diversity of novel proteins in gonadal fluids. In KW McKerns and M Chretien (eds) *Molecular Biology of Brain and Endocrine Peptidergic Systems*. Plenum Press, New York, pp. 259–273.

Manjunath P, Marcel YL, Uma J, Seidah NG, Chretien M and Chapdelaine A (1989) Apolipoprotein A-I binds to a family of bovine seminal plasma proteins. J Biol Chem 264,16853–16857.

Manjunath P, Chandonnet L, Leblond E and Desnoyers L (1994) Major proteins of bovine seminal vesicles bind to spermatozoa. Biol Reprod 50,27–37.

Manjunath P, Nauc V, Bergeron A and Menard M (2002) Major proteins of bovine seminal plasma bind to the low-density lipoprotein fraction of hen's egg yolk. Biol Reprod 67,1250–1258.

Manjunath P, Bergeron A, Lefebvre J and Fan J (in press) Seminal plasma proteins: functions and interaction with protective agents during semen preservation. In ER Roldan and M Gomendio (eds) *Society for Reproduction and Fertility Supplement*. Nottingham University Press, Thrumpton, Nottingham, UK.

Martin SL (1995) Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1. Gene 153,261–266.

Plucienniczak G, Jagiello A, Plucienniczak A, Holody D and Strzezek J (1999) Cloning of complementary DNA encoding the pB1 component of the 54-kilodalton glycoprotein of boar seminal plasma. Mol Reprod Dev 52,303–309.

Salois D, Menard M, Paquette Y and Manjunath P (1999) Complementary deoxyribonucleic acid cloning and tissue expression of BSP-A3 and BSP-30-kDa: phosphatidylcholine and heparin-binding proteins of bovine seminal plasma. Biol Reprod 61,288–297.

Schwede T, Kopp J, Guex N and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31,3381–3385.

Seidah NG, Manjunath P, Rochemont J, Sairam MR and Chretien M (1987) Complete amino acid sequence of BSP-A3 from bovine seminal plasma. Homology to PDC-109 and to the collagen-binding domain of fibronectin. Biochem J 243,195–203.

Shapiro MB and Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15,7155–7174.

Shehee WR, Chao SF, Loeb DD, Comer MB, Hutchison CA, III and Edgell MH (1987) Determination of a functional ancestral sequence and definition of the 5′ end of A-type mouse L1 elements. J Mol Biol 196,757–767.

Sullivan R (1999) Interaction between sperm and epididymal secretory proteins. In C Gagnon (ed) *The Male Gamete. From Basic Science to Clinical Applications*. Cache River Press, Vienna, IL, pp. 93–104.

Therien I, Bousquet D and Manjunath P (2001) Effect of seminal phospholipid-binding proteins and follicular fluid on bovine sperm capacitation. Biol Reprod 65,41–51.

Therien I, Bergeron A, Bousquet D and Manjunath P (2005) Isolation and characterization of glycosaminoglycans from bovine follicular fluid and their effect on sperm capacitation. Mol Reprod Dev 71,97–106.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22,4673–4680.

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP and Myers RM (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14,62–66.

Villemure M, Lazure C and Manjunath P (2003) Isolation and characterization of gelatin-binding proteins from goat seminal plasma. Reprod Biol Endocrinol 1,39.

Visconti PE and Kopf GS (1998) Regulation of protein phosphorylation during sperm capacitation. Biol Reprod 59,1–6.

Wah DA, Fernandez-Tornero C, Sanz L, Romero A and Calvete JJ (2002) Sperm coating mechanism from the 1.8 A crystal structure of PDC-109-phosphorylcholine complex. Structure (Camb) 10,505–514.

Weerachatyanukul W, Xu H, Anupriwan A, Carmona E, Wade M, Hermo L, da Silva SM, Rippstein P, Sobhon P, Sretarugsa P *et al*. (2003) Acquisition of arylsulfatase A onto the mouse sperm surface during epididymal transit. Biol Reprod 69,1183–1192.

Yanagimachi R (1994) Mammalian Fertilization. In E Knobil and JD Neill (eds) *The Physiology of Reproduction*, 2nd edn. Raven Press, New York, pp. 189–317.

Submitted on August 8, 2006; resubmitted on October 6, 2006; accepted on October 10, 2006