Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Genomics 87 (2006) 382 – 391
GENOMICS
www.elsevier.com/locate/ygeno

Comparative genomics of the sperm mitochondria-associated cysteine-rich protein gene

Sabrina K. Hawthorne, Golnaz Goodarzi, Jana Bagarova, Katherine E. Gallant, Rakhee R. Busanelli, Wendy J. Olend, Kenneth C. Kleene *

Department of Biology, University of Massachusetts at Boston, 100 Morrissey Boulevard, Boston, MA 02125, USA

Received 1 July 2005; accepted 20 September 2005
Available online 1 December 2005

Abstract

The sperm mitochondrial cysteine-rich protein (SMCP) is a rapidly evolving cysteine- and proline-rich protein that is localized in the mitochondrial capsule and enhances sperm motility. The sequences of the SMCP protein, gene, and mRNA in a variety of mammals have been compared to understand their evolution and regulation. SMCP can now be reliably identified by its tripartite structure including a short amino-terminal segment; a central segment containing short tandem repeats rich in cysteine, proline, glutamine, and lysine; and a C-terminal segment containing no repeats, few cysteines, and a C-terminal lysine. The *SMCP* gene is located in the epidermal differentiation complex (EDC), a large gene cluster that functions in forming epithelial barriers. Similarities in chromosomal location, molecular function, intron–exon structure, and protein organization argue that *SMCP* originated from an EDC gene and acquired spermatogenic cell-specific transcriptional and translational regulation and a novel cellular function in sperm motility. The *SMCP* 5′ UTR and 3′ UTR contain conserved elements and uORFs that may function in cytoplasmic regulation of gene expression, and the levels of *SMCP* mRNA in human are much lower than in other mammals, a feature of male-biased expression. The evolution of SMCP has been accompanied by changes in the sequence, number, and length of repeat units, including three alleles in dogs. The major proteins associated with the mitochondrial capsule, SMCP and phospholipid hydroperoxide glutathione peroxidase, provide outstanding examples of changes in cellular function driven by selective pressures on sperm motility, an important determinant of male reproductive success.

Keywords: Sperm mitochondria-associated cysteine-rich protein; Sperm mitochondrial capsule; Translational regulation; Epidermal differentiation complex; Tandem repeats; uORF; Unstructured protein; Male-biased expression

The sperm flagellum generates motility, a factor that is under strong evolutionary pressures in the competition to fertilize the egg among the millions of sperm in the ejaculates of individual and multiple males [1]. The flagella of mammalian sperm are modified by three sperm-specific accessory structures, the mitochondrial sheath, the outer dense fibers, and the fibrous sheath, which provide mechanical factors that enhance motility and a scaffold for localization of proteins that produce ATP and regulate motility [2].

The outer membranes of sperm mitochondria are modified by a structure, the mitochondrial capsule, that is crosslinked by disulfide bridges and resistant to SDS [3]. The mitochondrial capsule confers a stiffness on sperm mitochondria that is very different from that of somatic mitochondria: sperm mitochondria do not swell in hypotonic media and retain the characteristic crescent shape of mitochondria in the mitochondrial sheath after being pelleted in the ultracentrifuge [3]. Biochemical and immunocytochemical studies demonstrate that the mitochondrial capsule is associated with several proteins, including two ~20-kDa proteins that were confused for many years, the sperm mitochondria-associated cysteine-rich protein (SMCP), a small, hydrophilic, cysteine- and proline-rich protein, and a sperm mitochondria-specific isoform of the selenoprotein, phospholipid hydroperoxide glutathione peroxidase, PHGPx [3–7]. SMCP replaces the misleading name “mitochondrial capsule selenoprotein.” A careful examination of the conflicting studies of the composition of the mitochondrial capsule suggests that the primary constituent of the capsule is PHGPx [4,6], while SMCP is present in a

* Corresponding author. Fax: +1 617 287 6650.
E-mail address: kenneth.kleene@umb.edu (K.C. Kleene).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.09.010

"granular layer" on the surface of the capsule [3], but rigorous evidence is lacking for this hypothesis.

Targeted deletion of the mouse *Smcp* gene has no discernable effect on the ultrastructure of the mitochondrial sheath, but sperm motility on the inbred background is impaired, resulting in sperm that fail to migrate in the female reproductive tract and penetrate the egg membranes during fertilization [8]. Expression of the *Smcp* mRNA at high levels is restricted to haploid spermatids, and translation is developmentally regulated: the *Smcp* mRNA is stored in translationally repressed free mRNPs in early haploid cells, and it is actively translated in late haploid cells [5,9]. SMCP is a rapidly evolving protein that exhibits little similarity in divergent species [5,10,11], and the features that distinguish it from other cysteine- and proline-rich proteins have never been identified.

The accompanying comparative analysis of the sequences of the *SMCP* gene and protein in various mammals identifies the distinctive features of SMCP and conserved elements of the mRNA that may function in translational regulation. We also establish unexpectedly that the *SMCP* gene is a paralogue of the genes in the epidermal differentiation complex, a large gene cluster that functions in forming epithelial barriers in cornified epithelia, and that alternations in the number of copies of a short repeated unit have played an important part in the evolution of the protein.

Results

SMCP sequences in various mammals

To identify candidate SMCPs, the EST databases were searched with TBLASTN with SMCPs from mouse, human, and bull. The putative *SMCP* ESTs were evaluated by several criteria: (1) the EST was derived from a library containing testis cDNAs, (2) predicted proteins that exhibit the distinctive features identified here, (3) conserved sequences are present in the 5′ UTR and 3′ UTR as identified below. Selected cDNAs were purchased and sequenced completely on both strands. The bull, hamster, dog, and deer mouse *SMCP* mRNAs were RT-PCR amplified using universal primers for conserved sequences in the 5′ UTR and 3′ UTR, and the sequences of the dog and human 5′ UTRs were extended by 5′ RACE.

Structure and chromosomal location of SMCP genes

BLASTN searches of the mouse, rat, and human genomes reveal that the *SMCP* gene is located near the middle of the epidermal differentiation complex (EDC), a cluster of ~50 genes encoding proteins that form epithelial barriers in syntenic locations at chromosome 3F1 in mouse, chromosome 1q21 in human, and chromosome 2q34 in rat [13,14], confirming in situ hybridization studies [10,11]. In each genome, the *SMCP* gene is located between the involucrin gene and a homolog of the human *XP5/LED10* gene [13]. The Discussion assembles several lines of evidence arguing that the *SMCP* gene evolved from an *EDC* gene.

Comparison of the sequences of the *SMCP* cDNAs and genomic DNAs reveal that the *SMCP* gene in mouse, bull, rat, human, and dog contains a single intron, located 20 nt upstream of the *SMCP* ATG translation initiation codon. The intron is 3829 nt in dog, 4134 nt in bull, 4269 nt in mouse, 4147 nt in rat, and 5956 nt in human, and the splice junctions conform to the AG:GT rule [15]. The structure of the mouse, rat, and human *SMCP* genes deduced here agrees with the structures deduced from genomic clones [10–12].

SMCP is expressed at very low levels in epithelial tissues

The realization that the *SMCP* gene is an *EDC* paralogue raises questions whether the *SMCP* promoter is active in epithelial tissues, a question that was not answered by previous Northern blot analyses of nonepithelial somatic tissues [10,11]. RT-PCR using 30 amplification cycles detects a strong ethidium bromide-stained band in testis and weaker bands in uterus, tongue, stomach, skin, spleen, heart, and kidney (data not shown). Fifteen amplification cycles and a Southern blot detect *Smcp* mRNA at high levels in testis and at much lower levels in somatic tissues (Fig. 1). This protocol gives a more accurate indication of the differences in *Smcp* mRNA levels in various tissues than does ethidium bromide staining and demonstrates that *Smcp* is expressed at very low levels in epithelial tissues in which a variety of EDC mRNAs are expressed at high levels [13,14]. RT-PCR of 18S rRNA shows that similar amounts of PCR products are detected in all lanes (Fig. 1), and the absence of ethidium bromide-

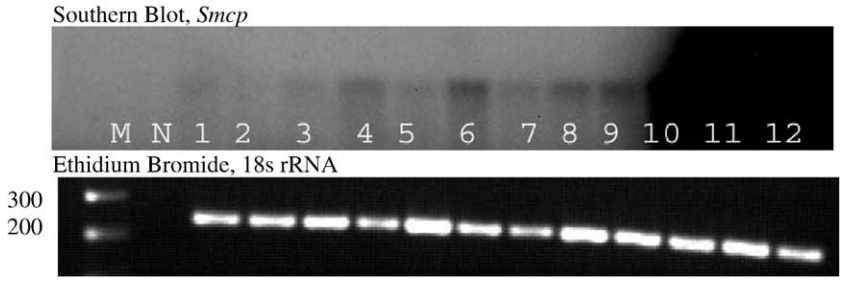

Fig. 1. RT-PCR analysis of the levels of *Smcp* mRNA in various adult mouse tissues. Top: *Smcp* mRNA was RT-PCR amplified for 15 cycles, a point at which ethidium bromide staining detected a PCR product only in testis, Southern blotted, and probed with an *Smcp* probe. Bottom: 18S rRNA was RT-PCR amplified for 15 cycles and the product was detected by ethidium bromide staining. M, 100 bp ladder size marker; N, negative control (no reverse transcriptase); lane 1, uterus; lane 2, stomach; lane 3, esophagus; lane 4, tongue; lane 5, lung; lane 6, spleen; lane 7, skin; lane 8, heart; lane 9, kidney; lane 10, liver; lane 11, brain; lane 12, testis. The sequences of the primers used in the RT-PCR are shown in Table 1.