Gene 481 (2011) 65–75

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Human sphingomyelin synthase 1 gene (*SMS1*): Organization, multiple mRNA splice variants and expression in adult tissues

Alexandra V. Rozhkova [a,*], Veronika G. Dmitrieva [a], Olga N. Zhapparova [b], Olga Yu. Sudarkina [a], Elena S. Nadezhdina [b], Svetlana A. Limborska [a], Lyudmila V. Dergunova [a]

[a] Department of Human Molecular Genetics, Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov sq., 2, 123182 Moscow, Russia
[b] Institute of Protein Research, Russian Academy of Sciences, Vavilova Str., 34, 117334 Moscow, Russia

ARTICLE INFO

Article history:
Accepted 20 April 2011
Available online 29 April 2011

Received by Lynn Jorde

Keywords:
Alternative transcripts
In silico analysis
Gene structure
5′- and 3′-untranslated regions
Real-time PCR quantification

ABSTRACT

We have previously characterized the structure of the human *MOB* gene (*TMEM23*), which encodes a hypothetical transmembrane protein (Vladychenskaya et al., 2002, 2004). The primary structure of the peptide that we predicted coincided completely with the amino acid sequence of the later identified sphingomyelin synthase 1 protein (SMS1), which catalyses the transfer of a phosphorylcholine moiety from phosphatidylcholine to ceramide, producing sphingomyelin and diacylglycerol (Huitema et al., 2004; Yamaoka et al., 2004). The gene we found was the *SMS1* gene. The combination of *in silico* and RT-PCR data helped us identify and characterize numerous new transcripts of the human *SMS1* gene. We identified mRNA isoforms that vary in the 5′-untranslated region (UTR) and encode the full-length protein, and transcripts resulting from alternative combinations of the exons in the coding region of the gene and the 3′-UTR. Comparison of the discovered transcripts' structures with the sequence of human chromosome 10 showed that the human *SMS1* gene comprises at least 24 exons. RT-PCR and real-time PCR data showed that the expression patterns of the alternative *SMS1* transcripts are tissue specific. Our results indicate that the regulation of *SMS1* expression is complex and occurs at the transcriptional, post-transcriptional and translational levels.

1. Introduction

Sphingomyelin synthase (SMS) is essential in eukaryotic cells (Tafesse et al., 2007; Yamaoka et al., 2004). The functional activity of the enzyme catalysing the synthesis of sphingomyelin was described previously, but the protein remained uncharacterized (van Helvoort et al., 1994; Marggraf and Kanfer, 1984; Voelker and Kennedy, 1982). As a transferase, this enzyme transfers a phosphorylcholine residue from phosphatidylcholine to ceramide and produces sphingomyelin and diacylglycerol (DAG) (Marggraf and Kanfer, 1984; Ullman and Radin, 1974; Voelker and Kennedy, 1982). SMS also catalyses the reverse reaction; i.e., the formation of phosphatidylcholine and ceramide from sphingomyelin and DAG (van Helvoort et al., 1994; Marggraf and Kanfer, 1984). Both reactions generate two types of compounds: structural components of the cell (sphingomyelin and phosphatidylcholine) and so-called "signal molecules" (DAG and ceramide). Thus, the reactions catalysed by this enzyme ensure the synthesis of sphingomyelin, a component of the outer layer of cell membranes, and maintain the balance between cell death and survival by regulating the formation of the pro-apoptotic mediator ceramide and synthesis of the anti-apoptotic mediator DAG (Claus et al., 2009; Geilen et al., 1997; Hannun, 1994; Hannun et al., 2001).

In 2004, two independent research groups identified cDNA clones that included the open reading frames (ORFs) encoding a protein with SMS enzymatic activity (Huitema et al., 2004; Yamaoka et al., 2004). The two discovered sequences encode protein isoforms, named SMS1 and SMS2, with the same enzymatic activity (Huitema et al., 2004). *In silico* experiments identified putative amino acid sequences of both proteins and revealed some of their structural features. It was demonstrated that SMS1 is located in the Golgi complex and that SMS2 accumulates mainly on the surface of the plasma membrane (Huitema et al., 2004; Tani and Kuge, 2009; Yeang et al., 2008). SMS1 accounts for 60% to 80% and SMS2 for the remaining 20% to 40% of the SMS activity (Tafesse et al., 2007).

The amino acid sequence of the SMS1 protein coincides completely with the primary structure of the hypothetical transmembrane peptide encoded by the *MOB* gene, which we identified and which localizes in the 10q11.2 region of human chromosome 10 (Dergunova et al., 1998, 2003; Vladychenskaya et al., 2002, 2004). The 320 kb *MOB* gene (*TMEM23*) contains 11 exons. The 3734-nucleotide (nt) transcript

Abbreviations: ARE, adenylate uridylate-rich element; bp, base pair(s); C, cytidine; cDNA, DNA complementary to RNA; DAG, diacylglycerol; DMEM, Dulbecco's modified Eagle's medium; DNase, deoxyribonuclease; EGFP, enhanced green fluorescent protein; EST, expressed sequence tag; G, guanosine; GAPDH, glyceraldehyde-3-phosphate-dehydrogenase; kb, kilobase(s) or 1000 bp; kDa, kilodalton(s); nt, nucleotide(s); ORF, open reading frame; p, plasmid; RLM-RACE, RNA ligase mediated rapid amplification of cDNA ends; RT-PCR, reverse transcriptase-polymerase chain reaction; SAM, sterile alpha motif; TM, transmembrane domain; uORF, upstream ORF; UTR, untranslated region(s).

* Corresponding author. Fax: +7 499 196 0221.
E-mail address: avrojk@yandex.ru (A.V. Rozhkova).

0378-1119/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2011.04.010

encodes a protein comprising 413 amino acid residues. The protein includes several domains: the N-terminal sterile α motif (SAM) domain and five presumed transmembrane domains. An alternative transcript of this gene (*MOB1*), which lacks exon 7, has been discovered (Vladychenskaya et al., 2004). Here, we used *in silico* and RT-PCR data to identify a number of new transcripts of the human *SMS1* gene. Comparison of the structure of all the discovered transcripts with the nucleotide sequence of human chromosome 10 showed that the human *SMS1* gene comprises at least 24 exons. The data obtained by RT-PCR and real-time PCR revealed that the expression pattern of the alternative *SMS1* transcripts is tissue specific. These results indicate that the regulation of *SMS1* gene expression is complicated and might occur at the transcriptional, post-transcriptional and translational levels.

2. Materials and methods

2.1. Bioinformatics

The search for alternative transcripts of the *SMS1* gene was conducted using a powerful strategy of gene discovery, named "*in silico cloning*" and first described by Wilcox et al., (1991). The sequence of interest ("virtual probe") was hybridized with all accumulated expressed sequences (expressed sequence tags (ESTs) and/or cDNA) using the BLAST search on the server of The National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi). This allowed us to identify previously unknown alternative transcripts of the *SMS1* gene. Comparing these transcripts with the nucleotide sequence of the human chromosome 10 genome contig (accession no. NT_030059) allowed us to localize the alternative exons within the sequence of chromosome 10. To compare the nucleotide sequences, the GenBank Human G+T (human genomic plus transcript), NR (nucleotide collection) and EST databases were used. The promoter regions were studied with the Gene2Promoter software available at the Genomatrix web server (http://www.genomatix.de). Putative ORFs were revealed and the protein sequences in single-letter code of amino acids were obtained using the ORFinder program from the server of The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/projects/gorf/). Putative transmembrane domains within the amino acid sequence of MOB (SMS1) were revealed using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

2.2. RNA isolation and treatment with DNase I

Total RNA was isolated from human brain and other tissues using guanidine thiocyanate (Chomczynski and Sacchi, 1987). The investigation was performed after approval from the local ethics committee was obtained. Total RNA from cultured HeLa cells electrotransfected with pEGFP-C2–SMS1_2a and pEGFP-N2–SMS1_2a plasmids was isolated using an RNeasy kit (Qiagen). RNA integrity was assessed by analysing the ratio between rRNA bands after agarose gel electrophoresis under denaturing conditions. RNA samples were stored in ethanol at −70 °C. To remove the residual genomic DNA from the total RNA samples, we treated each sample (100 μg) with DNase I (MBI Fermentas, Lithuania). RNA was then extracted with a 1:1 phenol–chloroform mixture and precipitated with sodium acetate (3.0 M, pH 5.2).

2.3. RT-PCR and real-time PCR quantification

Five-microgram aliquots of DNase I-treated total RNA samples from human tissues and electrotransfected HeLa cells were used for cDNA synthesis using a RevertAid™ First-strand cDNA Synthesis kit and random primers (MBI Fermentas) according to the supplier's recommendations. To detect *SMS1* transcripts, RNA from human cerebellum, forebrain cortex, hippocampus, testicle, liver, kidney, spleen and lymph node was analysed. First-strand human tissue cDNA was used as the template for PCR with *SMS1* gene-specific primers. Oligonucleotide primers were designed using OLIGO software; their sequences and complementarity to alternative exons are summarized in Table 1. All amplifications were performed in a Tercyc MC2 thermocycler (DNA Technology, Russia). PCR was performed in 25 μl of reaction mixture containing 1 μl of the reverse transcription mixture, 10 μl of a 2.5× ready-to-use reaction mixture, 5 pM of each gene-specific primer and 1.25 units of Taq DNA polymerase (Syntol, Russia). The PCR parameters were as follows: 10 min at 95 °C followed by 30 cycles of 95 °C for 1 min, 65 °C for 1 min and 72 °C for 1 min, followed by a 10 min final elongation at 72 °C. Aliquots of the PCR products were fractionated by size on 8% non-denaturing polyacrylamide gels in Tris-borate-EDTA buffer and silver stained (Budowle et al., 1991). As a control, the *GAPDH* gene was amplified. To analyse all the cDNAs with primers specific to *GAPDH* (forward: 5′-TTAGCACCCCTGGCCAAGG-3′; reverse: 5′-CTTACTCCTTGGAGGCCATG-3′), 25 cycles were performed with the same parameters. The expression activity of the *SMS1* transcripts was estimated by three independent PCR experiments for each tissue type; the amplification products were analysed three times on separate polyacrylamide gels.

The first-strand HeLa cDNA was amplified with *SMS1* gene-specific primers and *EGFP* gene-specific primers. To identify the *EGFP–SMS1_2a* transcript, the direct primer FEGFP (5′-TCCTGCTGGAGTTCGTGACC-3′) and reverse primers R9 and R10 were used (Table 1). To identify the *SMS1_2a–EGFP* transcript, the direct primers F9 and F10 and the reverse primer REGFP (5′-GTCGCCGTCCAGCTCGACCA-3′) were used (Table 1). Amplification and electrophoresis were performed as described earlier.

For real-time PCR, equal amounts of cDNA from human forebrain cortex and testis were added to a set of PCR with MasterMix Buffer B SYBR Green I (Syntol, Russia). To detect the *SMS1_1a* transcript, F1/R8 primers were used. The nucleotide sequences of the primers and complementarity to alternative exons are summarized in Table 1. Reactions were carried out using a StepOnePlus™ Real-Time PCR System (Applied Biosystems, USA). The following real-time PCR protocol was used for all transcripts: denaturation at 95 °C for 10 min; amplification and quantification repeated for 40 cycles at 95 °C for 1 min, 65 °C for 1 min and 72 °C for 1 min; and a single fluorescence measurement. All PCR reactions were run in triplicate, the crossing point

Table 1
The nucleotide sequences of the primers corresponding to the *SMS1* transcripts. The primer names coincide with the names of exons carrying the annealing sites for the primers. F and R indicate the direction of primers: forward and reverse, respectively. The figures in lowered printing denote the number of the primer.

Primer	Nucleotide sequence, 5′–3′
Direct primers for *SMS1* transcripts	
F1	GGCTGACTGCTCTCCCCTC
$F1a_1$	ACTCCAATCACGGCTCCAC
$F1a_2$	TGCCTGGTGTCGGATTGG
F2a	CAAAGCAGGAAGATGGTGAAC
F6	TACTGGAACAATGGGAGAAGC
$F6a_1$	CGCTCGCCTTTCCGCCC
$F6a_2$	GCGCAAACTTGGTGCCCCAGGA
F7u	TAGTCCCTGAAACGAATACAT
F7	GCCAAACAAGTCTCTGCTCAT
F7a1[a]	CCTGTCCTTCTACTGGGGTCC
F7b	CTGCATTTGTGTCAGACTTCAC
F9	TGGCTTGTCTATCACTGGCTC
F10	CACTGGATTTGCTGGCTTCTC
Reverse primers for *SMS1* transcripts	
R7	GGTGGGGATGTCTACGCC
R7a1	GGACCCCAGTAGAAGGACAGG
R7b	GTGAAGTCTGACACAAATGCAG
$R7d_1$	CAACAACCCAAGCCACTCTG
$R7d_2$	ACCCCAGTAAACAGCCAGTC
R8	TACAGCGTGCCAACTATGC
R9	GAGCCAGTGATAGACAAGCCA
R10	GAGAAGCCAGCAAATCCAGTG
R11	ATGGTGGTTGCGGGTTATGTA

[a] Primer F7a1 is annealed on the sequence of exons 7a, 7a1 and 7a2.