Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 351 (2005) 119 – 130

GENE
SECTION
FUNCTIONAL GENOMICS

www.elsevier.com/locate/gene

Cloning, cellular localization, genomic organization, and tissue-specific expression of the TGFβ1-inducible *SMAP-5* gene

Katrin Stolle[a,b], Michael Schnoor[a,b], Georg Fuellen[c,d], Michael Spitzer[c,d], Thomas Engel[a], Friedrich Spener[b], Paul Cullen[a], Stefan Lorkowski[a,b,*]

[a]*Institute of Arteriosclerosis Research, University of Münster, Münster, Germany*
[b]*Institute of Biochemistry, University of Münster, Münster, Germany*
[c]*Department of Medicine, AG Bioinformatics, University of Münster, Münster, Germany*
[d]*Division of Bioinformatics, Biology Department, University of Münster, Münster, Germany*

Received 8 November 2004; received in revised form 10 February 2005; accepted 14 March 2005

Received by A.J. van Wijnen

Abstract

SMAP-5 is a member of the five-pass transmembrane protein family localizing in the Golgi apparatus and the endoplasmic reticulum. These proteins have been implicated in intracellular trafficking, in secretion and in vesicular transport. Phylogenetic analyses revealed that SMAP-5 is a member of a small Rab GTPase interacting factor protein family. The human *SMAP-5* gene spans about 12.5 kb and comprises 6 exons on chromosomal locus 5q32. The proximal 5′-flanking region of the gene lacks a TATA box and is highly GC rich. Consistent with this, the *SMAP-5* gene is expressed in all tissues. The highest level of expression was found in coronary smooth muscle cells, in which expression of the *SMAP-5* gene was induced by transforming growth factor β1, thus indicating that this protein may play an important role in inflammation.

Keywords: SB140; FinGER5; Yip1; Transforming growth factor β1

1. Introduction

The multifunctional cytokine transforming growth factor β1 (TGF-β1) regulates cell proliferation, apoptosis and immune responses, participates in disorders such as fibrosis and acute or chronic inflammation and plays an essential role in normal immune function. TGF-β1-deficient mice that survive to adulthood exhibit severe multifocal inflammation, show massive infiltration of lymphocytes and macrophages in several organs, but also show activation of lymph nodes and hyperproliferation of lymphoid cells (Christ et al., 1994; Kulkarni et al., 1993; Shull et al., 1992). Recent studies suggest also an interaction with immune-mediated defense mechanisms (Chen and Wahl, 2003; Roes et al., 2003).

On the cellular level, TGF-β1 actuates a complex intracellular signaling by binding to, and activating, specific cell surface receptors. Three types of receptors involved in TGF-β1 binding and activation of intracellular downstream signaling are known. Type I and II TGF-β receptors are serine/threonine kinases that form a heterodimer upon binding of TGF-β1 resulting in phosphorylation of the cytoplasmic domain of the type I receptor (Topper, 2000). Formation and activation of the heterodimer is modulated by the type III receptor, a membrane-anchored proteoglycan lacking kinase activity, whose precise role is unclear.

Abbreviations: DMEM, Dulbecco's modified Eagle's medium; ER, endoplasmic reticulum; EST, expressed sequence tag; FinGER, five-pass transmembrane proteins localizing in the Golgi apparatus and the endoplasmic reticulum; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MTN, multiple tissue Northern; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription polymerase chain reaction; SMAP-5, smooth muscle cell associated protein 5; SMC, smooth muscle cell; SRP14, 14 kDa signal recognition particle; TGF-β1, transforming growth factor β1.

* Corresponding author. Institut für Arterioskleroseforschung, Domagkstraβe 3, 48149 Münster, Germany. Tel.: +49 251 83 52072; fax: +49 251 83 52062.

E-mail address: stefan.lorkowski@uni-muenster.de (S. Lorkowski).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.03.012

A well-characterized downstream signaling pathway activated by TGF-β1 is transmitted by proteins that belong to the Smad family. Activation of this cascade occurs by phosphorylation of Smad2 or Smad3 through the phosphorylated type I TGF-β receptor kinase domain after binding of TGF-β1. The phosphorylated Smad2 (or Smad3) binds to Smad4, a common Smad, which is shared by different TGF-β superfamily pathways. The resulting heteromeric complex translocates into the nucleus where it associates with DNA-binding proteins and other transcriptional regulatory proteins prior to activating gene transcription (Topper, 2000).

Despite major advances in understanding the mechanisms of TGF-β1 signaling, little progress has been made in the identification of direct target genes (Pisano et al., 2003; Verrecchia et al., 2001). In particular, little information is available on genes that are involved in intracellular transport. In our study, we focused on the human members of a small protein family called FinGER (five-pass transmembrane proteins localizing in the Golgi apparatus and the endoplasmic reticulum) (Shakoori et al., 2003). These proteins may be involved in the vesicular transport between the endoplasmic reticulum (ER) and the Golgi apparatus. Loss of function of the essential yeast FinGER proteins Yip1p and Yif1p causes a block of ER to Golgi transport and accumulation of ER membrane (Matern et al., 2000; Yang et al., 1998). Currently, it remains unclear whether Yip1p and Yif1p (and their mammalian orthologues) function in vesicle budding and/or fusion steps (Shakoori et al., 2003). The yeast FinGER proteins Yip2p and Yip3p are non-essential gene products that interact with Yip1p and several Ypt/Rab GTPases (Calero et al., 2001; Calero and Collins, 2002). A mammalian homologue of Yip3p, called PRA1, is a dissociation factor for Rab-GDI complex (GDF) that also interacts with a SNARE called VAMP2 (Gougeon et al., 2002; Sivars et al., 2003). Loss of function of this protein inhibited the ER to Golgi transport and disrupted Golgi morphology. The yeast proteins Yip4p and Yip5p are non-essential gene products shown to interact with a several Ypt proteins and with Yip1p and Yif1p (Calero et al., 2002).

We investigated if TGF-β1 modulates the expression of the human FinGER mRNAs. We show that the mRNA of SMAP-5 (smooth muscle cell associated protein 5; also called SB140 or FinGER5) is up-regulated in SMCs by treatment with TGF-β1. As TGF-β1 is involved in several disease conditions, SMAP-5 may therefore also play a role in these processes. Finally, we describe the genomic organization of the *SMAP-5* gene and present a phylogenetic analysis of the FinGER protein family and the tissue-specific expression of SMAP-5 mRNA.

2. Materials and methods

2.1. Cell culture

Human coronary SMCs were from BioWhittaker. Cells were cultured using the SmGM2 Bullet Kit in medium containing 5% fetal bovine serum (BioWhittaker) as previously described (Stolle et al., 2004). When 80% confluent, cells were grown for 48 h in MCDB media (Sigma) supplemented with 100 units/ml penicillin, 100 μg/ml streptomycin and 4 mM L-glutamine (ICN Biomedicals), 100 mM sodium hydrogen carbonate (Sigma), 10 μl/ml ITS (BD Biosciences), 30 μg/ml heparin (Merckle), 50 μg/ml ascorbic acid and 1% fetal bovine serum (Cytogen). Using the supplemented MCDB medium, cells were incubated with or without 10 ng/ml TGF-β1 for different times. HeLa cells were obtained from ATCC and human skin fibroblasts cultured from biopsies of adult human hip skin were isolated as described previously (Walter et al., 1994). HeLa cells and fibroblasts were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (Biochrom), 100 units/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine.

2.2. RNA isolation

Total cellular RNA was isolated from cells using the RNeasy Mini kit (Qiagen) as previously described (Lorkowski et al., 2001a,b). Contaminating genomic DNA was removed using DNase I (Qiagen).

2.3. Reverse transcription

About 5 μg of total RNA were reverse transcribed using 500 ng/μl oligo(dT) primer (Invitrogen), 1 mM dNTPs (Fermentas), 20 units of ribonuclease inhibitor (Promega) and 200 units of M-MuLV reverse transcriptase RevertAid (Fermentas) in a total volume of 20 ml. After diluting the cDNA tenfold, aliquots were immediately used or stored at −20 °C.

2.4. RACE assays

RACE analyses were performed using the GeneRacer Kit (Invitrogen) and different reverse or forward primers specific for the SMAP-5 mRNA (Table 1). For 5′-RACE assays, total RNA isolated from human coronary SMCs was incubated with calf intestinal phosphatase in order to eliminate truncated mRNAs and non-mRNAs. Thereafter, the dephosphorylated mRNA was treated with tobacco acid pyrophosphatase to remove the 5′-cap. The full-length mRNA was ligated to the GeneRacer RNA oligonucleotide using T4 RNA ligase. After reverse transcription using the GeneRacer oligo(dT) primer, PCRs were performed using the GeneRacer 5′-forward primer and SMAP-5-specific reverse primers (Table 1). Nested PCRs were performed using the GeneRacer 5′-nested forward primer and nested SMAP-5-specific reverse primers (Table 1). For 3′-RACE analyses, total RNA isolated from human coronary SMCs was reverse transcribed using the GeneRacer oligo(dT) primer. PCRs were performed using the GeneRacer 3′-reverse primer and SMAP-5 mRNA specific forward

Table 1
Primers used in this study (all PCR and RACE primers were also used for sequencing)

Primer	Sequence	Exon
5′-RACE A reverse	5′-CGG TGT ATG GCT GTT GTG GCT GCA TCA TGT-3′	3
5′-RACE B reverse	5′-GCC AGT AGC AAT GTG GCT CCA AAA GCA AGG-3′	4
5′-RACE C reverse	5′-CAA TTG CAC TGA TCC CGT AT-3′	5
3′-RACE A forward	5′-AGC CAC ATT GCT ACT GGC T-3′	4
3′-RACE B forward	5′-TGG CAG GTC CAA TGG TTT TTT GCC TTG CTT-3′	4
3′-RACE C forward	5′-TTT TGG AGC CAC ATT GCT ACT GGC TGG CA-3′	4
3′-RACE D forward	5′-AGT GCA ATT GGA TGT CTA GGA ATG TT-3′	4
3′-RACE E forward	5′-TGT AGA CAT GAA ATG CGA ACA CTT-3′	6
3′-RACE F forward	5′-CTT GCC AGA GAG AAT CTG TAG GAA-3′	6
3′-RACE G forward	5′-CTT GCC AGA GAG AAT CTG TAG GAA AAT ACT GTA TC-3′	6
Real-time PCR forward	5′-TGC CCA TGA TCC TAC TTT CCA G-3′	5
Real-time PCR reverse	5′-AGT TGC TGT CCT TCC ATG GCT-3′	6
MTN probe forward	5′-TTA GTG AAG CTA CTG CCT TTG C-3′	1
MTN probe reverse	5′-CAA TTG CAC TGA TCC CGT AT-3′	5
Splicing intron 1 forward	5′-TTA TTT GCA ATG TCA GGC TTT GAA-3′	1
Splicing intron 1 reverse	5′-AGT CAT AGC CAG CAT ACT GTT TGC TA-3′	2
Splicing intron 2 forward	5′-TAT GCT GGC TAT GAC TAT TCG CA-3′	2
Splicing intron 2 reverse	5′-GGC TGA GGT GAA GCT GGA GTA T-3′	3
Splicing intron 3 forward	5′-CAA CTT TGA GGA TGA GCC ACC T-3′	3
Splicing intron 3 reverse	5′-ACC TGC CAA ATC AGT TTC ATT CA-3′	4
Splicing intron 4 forward	5′-AGT GCA ATT GGA TGT CTA GGA ATG TT-3′	4
Splicing intron 4 reverse	5′-GCA AAG CTG GAA AGT AGG ATC ATG-3′	5
Splicing intron 5 forward	5′-TTT TCT TTG CAA GGA ATG GTA GGA-3′	5
Splicing intron 5 reverse	5′-AAA GAC TCC ATA TAA CAA AGC GCA AG-3′	6
Exon 1a forward	5′-CCC GTG GCA CAG TGT TCA G-3′	1a
Exon 4b reverse	5′-CAA AGA CTT GAG′ GTG TAA AGC GG-3′	4b
Sequencing 1 forward	5′-TGA AGC TAC TGC CTT TGC CGC CAG CGC-3′	1
Sequencing 2 forward	5′-GCA CTC CAT CCA GCC TGG-3′	6
Sequencing 3 forward	5′-GCT ATA ATC GTT GGA TCG CCA CAT TTC CC-3′	6
Sequencing 4 forward	5′-GGA GAC ACA AGA CTT ACT GCA-3′	6
Sequencing 5 forward	5′-CGT TGG ATC GCC ACA TTT CCC-3′	6
Sequencing 1 reverse	5′-CAA CTC CAC TGC ATA GCT GTT TTG-3′	6
Sequencing 2 reverse	5′-CCA AAG CAT CTT CTA CCG TTT ATT AG-3′	6
Sequencing 3 reverse	5′-TTC CTA CAG ATT CTC TCT GGC AAG A-3′	6
Sequencing 4 reverse	5′-GGA AAT GTG GCG ATC CAA C-3′	6
Sequencing 5 reverse	5′-TGC AAC TCC ACT GCA TAG CTG TTT TGA CA-3′	6
Sequencing 6 reverse	5′-CCA GGA TGG TCT CGA TCT CCT-3′	6
Sequencing 7 reverse	5′-GGA TCA CTT TTT TGC TTT GAG TTA CC-3′	6
Sequencing 8 reverse	5′-GAG ACA GAG TCT CAC CCT GTC ACC CAG G-3′	6
Sequencing 9 reverse	5′-TGG AGT GCA GTG GTG TGA T-3′	6
C1-GFP SMAP-5 forward	5′-ATG AGC TCT GCA ATG TCA GGC TTT GAA AAC-3′	1
C1-GFP SMAP-5 reverse	5′-TAG GAT CCT CAA AAG ACG GAA ATC AGG GCA AA-3′	6

primers (Table 1). Nested PCRs were performed using the GeneRacer 3′-nested reverse primer and nested SMAP-5 specific forward primers (Table 1). In both assays, PCR products were checked by agarose gel electrophoresis and subjected to sequencing.

2.5. DNA sequencing

DNA sequencing was performed using the BigDye 2.0 or 3.1 Terminator Cycle Sequencing kit (Applied Biosystems) in a PTC-200 thermal cycler (MJ Research) as described (Stolle et al., 2004). PCR products for sequencing were either gel purified and/or cloned into pCRII-TOPO vector (Invitrogen) using the TOPO TA Cloning kit (Invitrogen). The primers used for sequencing PCR products were identical to the primers used for amplification of corresponding targets. Sequencing of PCR fragments cloned into pCRII-TOPO vector was performed using M13/pUC reverse and M13/pUC forward primers (Invitrogen). Products of the sequencing reaction were purified by gel filtration using Sephadex G-50 Superfine (Amersham Biosciences) prior to electrophoretic separation and fluorimetric detection on an ABI Prism 3700 DNA sequencer (Applied Biosystems).

2.6. Primer design

PCR primers (Table 1) were designed using the Primer Express v2.0.0 software (Applied Biosystems) as previously described (Stolle et al., 2004). Primers were from Invitrogen and forward and reverse primers locate in different exons.

2.7. Multiple tissue Northern (MTN) blot

SMAP-5 tissue expression was analyzed using Human 12-lane MTN Blots (BD Bioscience). To prepare a SMAP-5-specific probe, total RNA was isolated from human coronary SMCs and reverse transcribed. MTN probe forward and reverse primers (Table 1) were used to PCR amplify a 530 bp product that was cloned into a pCRII-TOPO vector (Invitrogen). A digoxigenin-labeled RNA probe was synthesized using the DIG RNA Labeling kit (Roche) and the linearized vector as template. The MTN blot was hybridized with 50 ng/ml of digoxigenin-labeled probe in ExpressHyb solution (BD Bioscience) for 2 h. Standard methods and CDP-Star chemiluminescence reagent (Roche) were used for detection.

2.8. RT-PCR

RT-PCR mixes of 20 μl included 0.5 μl cDNA, 1.25 units HotStar *Taq* polymerase and PCR buffer (both from Qiagen) containing 1.5 mM $MgCl_2$. In some cases, $MgCl_2$ up to 3.5 mM was added. PCR cycling conditions were as follows: Initial denaturation at 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C, 30 s at the primer melting temperature and 72 °C for 1 min per 1000 base pairs. Products were checked by agarose gel electrophoresis and sequenced.

2.9. Real-time RT-PCR

We performed real-time RT-PCR using the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) and the QuantiTect SYBR®Green PCR kit (Qiagen). Cycling parameters were: initial denaturation at 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C and 60 s at 60 °C. Each reaction contained cDNA corresponding to between 0.2 and 7 ng total RNA and 200 nM of forward and reverse primer (Table 1). For tissue-specific mRNA expression analyses, total RNA of different tissues was obtained from BD Bioscience or Ambion. For each assay, cDNA samples were assayed in duplicate and analyzed using the Sequence Detection System v2.1 software (Applied Biosystems). We performed relative mRNA quantification using the ΔΔCt method and normalized samples as described below. PCR efficiencies were calculated using cDNA dilutions derived from appropriate cells and were taken into account for calculating fold changes. Fold changes were finally normalized to the three most stably expressed housekeeping genes, 14 kDa signal recognition particle (SRP14; NM_003134), DNA directed RNA polymerase II polypeptide K (NM_005034) and family 3A histone H3 (NM_002107), using GeNorm (Vandesompele et al., 2002). Following PCR amplification, samples were subjected to melting temperature analyses. The identity of PCR products was verified by sequencing. Controls containing all constituents but the template were performed to rule out contamination of reagents. For TGF-β1 dependent expression analyses, the procedure was identical, but normalization was performed using glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and efficiencies were calculated using cDNA dilutions derived from coronary SMCs.

2.10. Cloning of cDNA, transient transfection and confocal imaging

Complementary DNA was produced by reverse transcribing total RNA purified from SMCs as described above. The open reading frame of the SMAP-5 transcript was amplified by PCR using appropriate oligonucleotides (Table 1) and a proof-reading *Pfu* DNA polymerase (Promega). The PCR product was inserted between *Bam*HI und *Sac*I sites into the pEGFP-C1 vector (BD Biosciences) to allow transient expression of a C-terminal GFP-tagged SMAP-5 protein. Cloned cDNA fragments were sequenced as described above to prove identity of the sequence. For transfection, HeLa cells were passaged and cultured in DMEM supplemented with 10% fetal calf serum for 24 h. Transfection with the SMAP-5 containing pEGFP-C1 vector was performed using a Gene Pulser II electroporator (Bio-Rad) according to the manufacturer's instructions. After different times of culturing, transfected cells were washed and immunofluorescence staining was performed as described previously (Lorkowski et al., 2001b) using a monoclonal mouse anti-GM130 antibody (BD Transduction Laboratories) as primary antibody and AlexaFluor 555 goat anti-mouse IgG (Molecular Probes) as secondary antibody. The stained cells were observed using a confocal microscope (Axiovert 200 M LSM510, Carl Zeiss).

2.11. Database searches and multiple sequence alignments

GenBank entry NM_030799 containing the putative human SMAP-5 mRNA sequence was used to search the GenBank NR database to generate an aligned SMAP-5 mRNA/EST (expressed sequence tag) dataset. Human sequences showing a significant number of matches to the putative human SMAP-5 mRNA sequence were extracted. Single nucleotide insertions as well as regions of low quality were removed from each entry using human chromosome 5 sequence (NT_029289) as reference. Expressed sequence tags in the wrong orientation were made reverse complementary. Next, sequences not spanning at least one intron were removed, because these sequences likely represent DNA contaminations. A multiple alignment of the processed sequences was generated using DiAlign (Morgenstern, 1999) and visualized using VisCoSe (Spitzer et al., 2004). The locations and sizes of exons and introns were extracted by BLAST-based

comparison of the consensus of the alignment to the sequence of human chromosome 5.

2.12. Sequence retrieval and phylogenetic inference

For phylogenetic analyses, an improved version of RiPE was applied to automatically retrieve proteins that are putative homologues of human SMAP-5 (Fuellen et al., 2003). The main improvements of RiPE are as follows. (i) Advanced PSI-BLAST post-processing was implemented to automate retrieval of sequence fragments that are homologous to the search profile generated from the human SMAP-5 homologues (FinGER1 to FinGER8; see Fig. 3 for GenBank accession numbers) by searching the database entries separately for high-and low-conserved parts of the profile and merging the retrieved fragments afterwards. (ii) This procedure was repeated once using the fragments retrieved from the first step as the new search profile to increase number and size of retrieved homologous sequence fragments. (iii) Redundancy of the protein dataset was minimized by implementing splice variant detection using support vector machines (Byvatov and Schneider, 2003). (iv) An automated detection of protein family boundaries based on E-value distribution was used to define an E-value threshold to distinguish between SMAP-5 homologues and non-homologous proteins.

We performed the PSI-BLAST search on the NCBI NR GenBank protein database (PSI-BLAST v2.2.4; database download 8th March 2004) using the eight known human members of the FinGER protein family. After completing the PSI-BLAST search, all non-RefSeq reference sequences were automatically removed yielding a dataset of 829 sequences. By minimizing redundancy and using the E-value-based family threshold for the FinGER proteins, the number of proteins was reduced to 59 representing the non-redundant set of the SMAP-5 (FinGER) protein family in the NR GenBank protein database. Definition lines of the sequence entries were reduced to the GenBank accession code and taxon. The resulting dataset was subjected to phylogenetic tree inference utilizing the QuickTree neighbor-joining implementation using 1.000 bootstrap trials (Howe et al., 2002; Saitou and Nei, 1987).

2.13. In silico analysis of predicted protein sequences and core promoter sequence

Predicted amino terminal protein sequences were analyzed for the presence of signal peptides, the position of cleavage sites and cellular localization using SignalP (http://www.130.225.67.199/services/SignalP/index.html) (Nielsen et al., 1997), the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), SMART (http://www.smart.embl-heidelberg.de/) (Letunic et al., 2004), and PSORT II (http://www.psort.ims.u-tokyo.ac.jp/) (Nakai and Horton, 1999). The core promoter of the human *SMAP-5* gene was analyzed using the Transcription Element Search System (http://www.cbil.upenn.edu/tess/).

2.14. Calculation of conserved sequence motifs

Consensus sequences of the human FinGER proteins and of each cluster shown in Fig. 2 were calculated and visualized using VisCoSe (http://www.viscose.ifg.uni-muenster.de) (Spitzer et al., 2004). To avoid misalignment and miscalculation of consensus sequences, entries NP_704350 (cluster III), NP_915653 (cluster IV) and NP_505775 (cluster VI) were removed because they contained insertions which are not part of the other homologous proteins.

3. Results

3.1. mRNA sequence, genomic organization and alternative splicing

RACE assays were performed to identify the human SMAP-5 mRNA sequence and to predict the genomic organization of the human *SMAP-5* gene. RACE analyses revealed a previously unknown 5′-end of the SMAP-5 mRNA which is 32 bp longer and a 3′-end which is 412 bp longer than that of the previously published SMAP-5 mRNA sequence (GenBank entry NM_030799.4). 3′-RACE of the same RNA and analysis of database sequences revealed SMAP-5 transcripts with at least four different 3′-ends (termed oligo(dA) I–IV) as shown in Fig. 1. The previously identified 3′-end of the hypothetical SMAP-5 mRNA (oligo(dA) sequence number II in Fig. 1) is located in the middle of exon 6, a position at which a genomic oligo(dA) DNA sequence is present. In addition, prediction of a SMAP-5 transcript comprising exons 1 to 6 using this oligo(dA) motif revealed a transcript of 1554 bp (transcript B). This is not in accordance with our findings from the MTN blot which revealed a SMAP-5 transcript of about 2 kb in size (Fig. 2). We therefore performed RT-PCRs using a forward primer located in exon 5 and different reverse primers located downstream of this oligo(dA) stretch to check if this is the natural occurring 3′-end of the SMAP-5 transcript or if mispriming of the oligo(dT) primer to the oligo(dA) stretch appeared. Using this approach, we identified transcript E and fragments that showed an extension of exon 6 downstream of the oligo(dA) sequence such as transcript A. To identify the 3′-end of the transcripts containing an exon 6 that is extended downstream of the genomic oligo(dA) sequence, we performed 3′-RACE assays using a forward primer located downstream of the genomic oligo(dA) sequence of exon 6. This approach revealed 3′-ends contained in the wild-type transcript and in transcript A (Fig. 1). No corresponding genomic oligo(dA) stretch was found for these 3′-ends, so

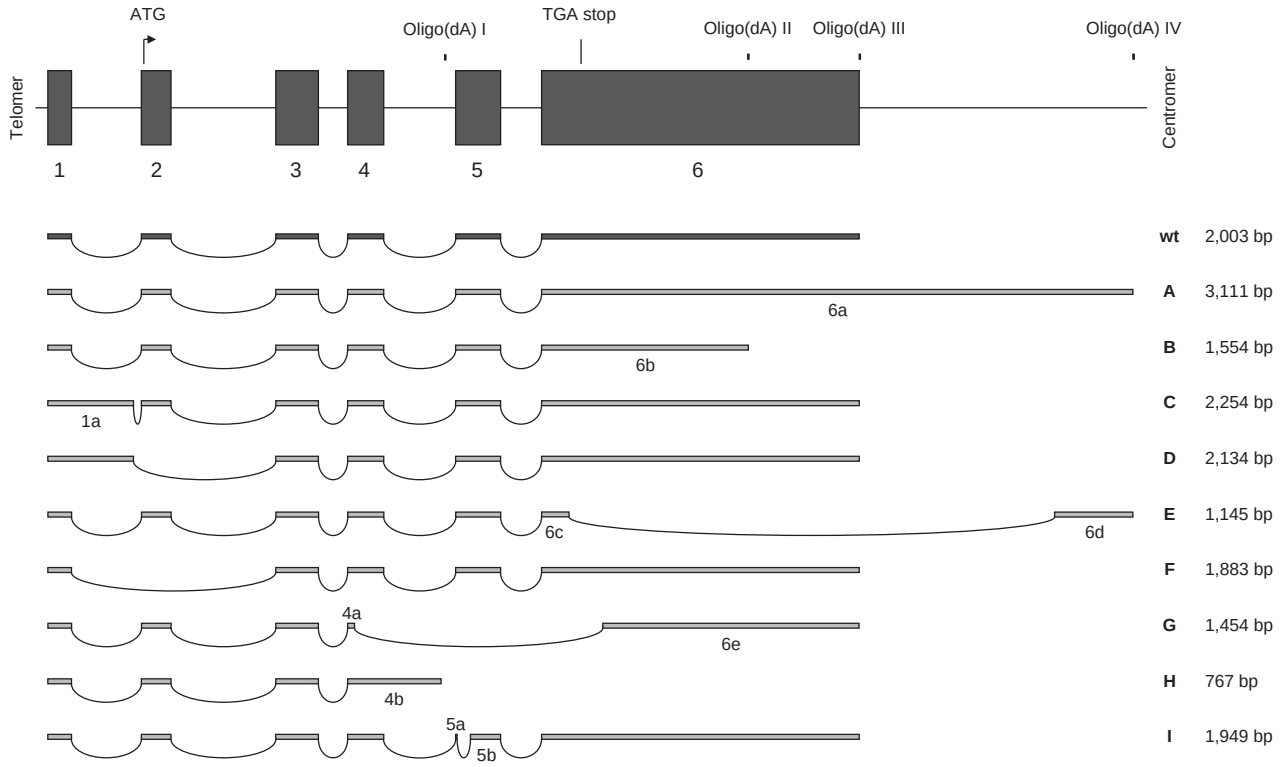

Fig. 1. Schematic representation of the human SMAP-5 mRNAs and genomic organization of the human *SMAP-5* gene. RACE and RT-PCR revealed six exons encoding a transcript with a total length of 2003 bp (see Fig. 2). The wild-type transcript is 32 bp longer at the 5′-end and 412 bp longer at the 3′-end as the GenBank reference transcript (entry NM_030799.4). Nine splice variants (named A to I) with lengths between 767 and 3111 bp were found. Transcripts A to E were experimentally identified in our studies, whereas transcripts F to I were derived from EST sequences included in the GenBank database (see text for details). The ATG start and TGA stop codons are located in exons 2 and 6, respectively. Exact positions of exons and introns are listed in Table 2.

that these ends are likely the natural occurring 3′-ends. Our MTN blots revealed a transcript size of about 2 kb. We therefore conclude that the wild-type sequence shown in Fig. 1 (2.003 bp) is the most prominent transcript in vascular coronary SMCs and not transcripts A, B and E. Transcripts C and D were reproducibly identified by RT-PCR with primers located in exon 1 and 3 but both transcripts were not detected by 5′-RACE using a reverse primer located in exon 5.

Overall, the human *SMAP-5* gene spans about 12.5 kb on chromosome 5q32 and comprises six exons encoding at least six different transcripts which were identified in our study. The positions and sizes of exons and introns are listed in Table 2. Exons comprising the wild-type transcript are termed 1 to 6 whereas splice variants were indexed with lower case letters a to e. The smallest exon of the wild-type SMAP-5 transcript is exon 1 with 96 bp and exon 6 is the largest with 1.286 nucleotides. Introns range between 569 bp (intron 1) and 4.241 bp (intron 2). All splice sites of the wild-type transcript have canonical boundaries (Breathnach and Chambon, 1981), starting the intron with ‘gt’ and ending with ‘ag’. However, several transcripts derived from alternative splicing show non-canonical boundaries (see Table 2). Transcripts F to I were derived from GenBank database sequences (entries AV661198, AV713741, AV714350, BX338306, CA389327, CD695794, AA654812, AI418308, BQ269248, BQ285890, CA771736 and CA772577) and were not verified in our own experiments. Therefore, we cannot exclude that these sequences containing variants of exons in the protein coding region were obtained because of cloning or PCR artifacts. Sequences of the different SMAP-5 transcripts identified in our experiments have been submitted to the GenBank database (entries AY640925–AY640929 and AY640934 corresponding to the wild-type transcript and transcript variants A to E).

3.2. Predicted protein sequence, signal peptides and membrane topology

The wild-type SMAP-5 transcript encodes a putative protein of 257 amino acids. According to SMART, the protein comprises four transmembrane α-helices forming a characteristic Rab GTPase interacting factor domain closely related to the yeast Yip1 protein (PFAM domain entry PF04893). According to PSORTII, the SMAP-5 protein contains five transmembrane helices and shows type 3b membrane topology. The predicted wild-type SMAP-5 protein has a putative cleavage site near the amino terminus, but the protein contains neither a signal peptide nor motifs predictive for the subcellular localization of the protein.

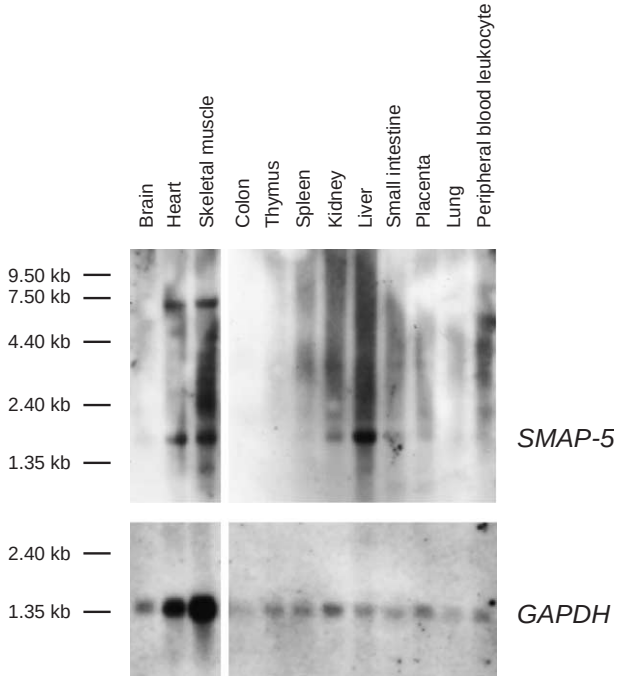

Fig. 2. Tissue specific expression of human SMAP-5 mRNA according to MTN blots. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA levels were used as a reference. Regulation of GAPDH mRNA expression is due to the fact that the mRNA amount in each lane was normalized to β-actin. SMAP-5 is present in different tissues showing highest levels of expression in liver. Prominent expression was also found in heart and skeletal muscle. In the other tissues low or no expression was observed. The size of the prominent SMAP-5 wild-type transcript is about 2 kb. The signals detected in heart and skeletal muscle appeared after 1 h, whereas the signals in the other tissues were detected after 10 min.

3.3. Putative regulatory elements in the core promoter

The core promoter of the human *SMAP-5* gene immediately upstream of exon 1 was analyzed using TESS to identify potentially regulatory elements and transcription factor binding sites. This analysis revealed that the region from the transcription start site of exon 1 to −350 bp upstream, which has a G/C content of about 60%, is a TATA box-less promoter that contains more than ten possible SP1 binding sites which are often found in promoters of TATA box-less genes (McKnight and Tjian, 1986), but only a single GC box.

3.4. Identification of homologous proteins

We used RiPE (Fuellen et al., 2003) to identify homologues of the human SMAP-5 protein in the GenBank RefSeq protein database and to calculate the phylogenetic tree shown in Fig. 3. We identified six clusters of homologous proteins containing the eight human FinGER proteins as described by Shakoori et al., 2003. Additionally, we identified a protein (GenBank entry NP_872398) showing high similarity to the profile of the eight human FinGER proteins which is, according to GenBank entry NP_872398, a hypothetical protein with experimental evidence of its mRNA. According to the NCBI Conserved Domain Database, this protein contains the Rab GTPase interacting factor domain of the FinGER proteins. The protein localizes in cluster VI of our phylogenetic tree and shows highest similarity to SMAP-5 (FinGER5). According to the nomenclature suggested by Shakoori et al., 2003, we termed this protein FinGER9. In contrast to the eight other human FinGER proteins, FinGER9 was expressed at much lower levels in human coronary SMCs. The closest homologues of SMAP-5 were identfied in the vertebrates *Mus musculus* (NP_075800) and *Danio rerio* (NP_956589) and homologous proteins were found in nearly all investigated eukaryotic proteomes.

Table 2
Positions of the exons comprising the human *SMAP-5* gene on chromosomal locus 5q32

Exon	Exon size (bp)[a]	Position[a]	Intron size (bp)[a]	Splice acceptor site[a] intron′exon	Donor acceptor site[a] exon′intron
1	96	4,713,075–4,713,035	569	′AGCAACGGGG	TCAGAGTTTG′gtgagtgacc
1a	347	4,713,075–4,712,784	318	′AGCAACGGGG	CAGTCCTAAG′gtgaaaaatc
2	120	4,712,465–4,712,346	4241	tttattacag′ATTATTTGCA	CCTATAGCAA′gtaacttttc
3	173	4,708,104–4,707,932	1175	tcatttatag′ACAGTATGCT	TTATTAGAAG′gtaagatttg
4	146	4,706,756–4,706,611	1711	ttctctttag′AGTTAGGTAT	ATTGCTACTG′gtaagatttt
4a	28	4,706,756–4,706,729	1829	ttctctttag′AGTTAGGTAT	ACCACATCTG′gcaaaaaaca
4b	378	4,706,756–4,706,379	–	ttctctttag′AGTTAGGTAT	TATTAGGGGA′
5	182	4,704,899–4,704,718	1658	ttacccatag′GCTGGCAAAA	TTTCTTTGCA′gtaagtactg
5a	6	4,704,899–4,704,894	54	ttacccatag′GCTGGC	GCTGGC′aaaatccagt
5b	122	4,704,839–4,704,718	1658	tctaggaatg′TTTTGTTTAT	TTTCTTTGCA′gtaagtactg
6	1286	4,703,059–4,701,774	–	ttgtttacag′AGGAATGGTA	TTCCCAATTA′
6a	2394	4,703,059–4,700,666	–	ttgtttacag′AGGAATGGTA	CTGTGTCTCA′
6b	838	4,703,059–4,702,222	–	ttgtttacag′AGGAATGGTA	CTCTGTCTCA′
6c	111	4,703,059–4,702,949	1966	ttgtttacag′AGGAATGGTA	TTTTAGTAGC′atatccttgc
6d	317	4,700,982–4,700,666	–	acactgttaa′TCTTCAGCCA	CTGTGTCTCA′
6e	1037	4,702,810–4,701,774	–	ctgctgcagc′GCAACTCTCA	TTCCCAATTA′

Sequences of the SMAP5 transcripts identified in this study have been published in the GenBank database with the accession nos. AY640925–AY640929 and AY640934. Exons 4a, 4b, 5a, 5b, and 6e have been predicted only from database sequences (see text for details).

[a] Derived from the GenBank *Homo sapiens* genomic DNA reference sequence of chromosome 5 contig (NT_029289.10; 5 April 2004).

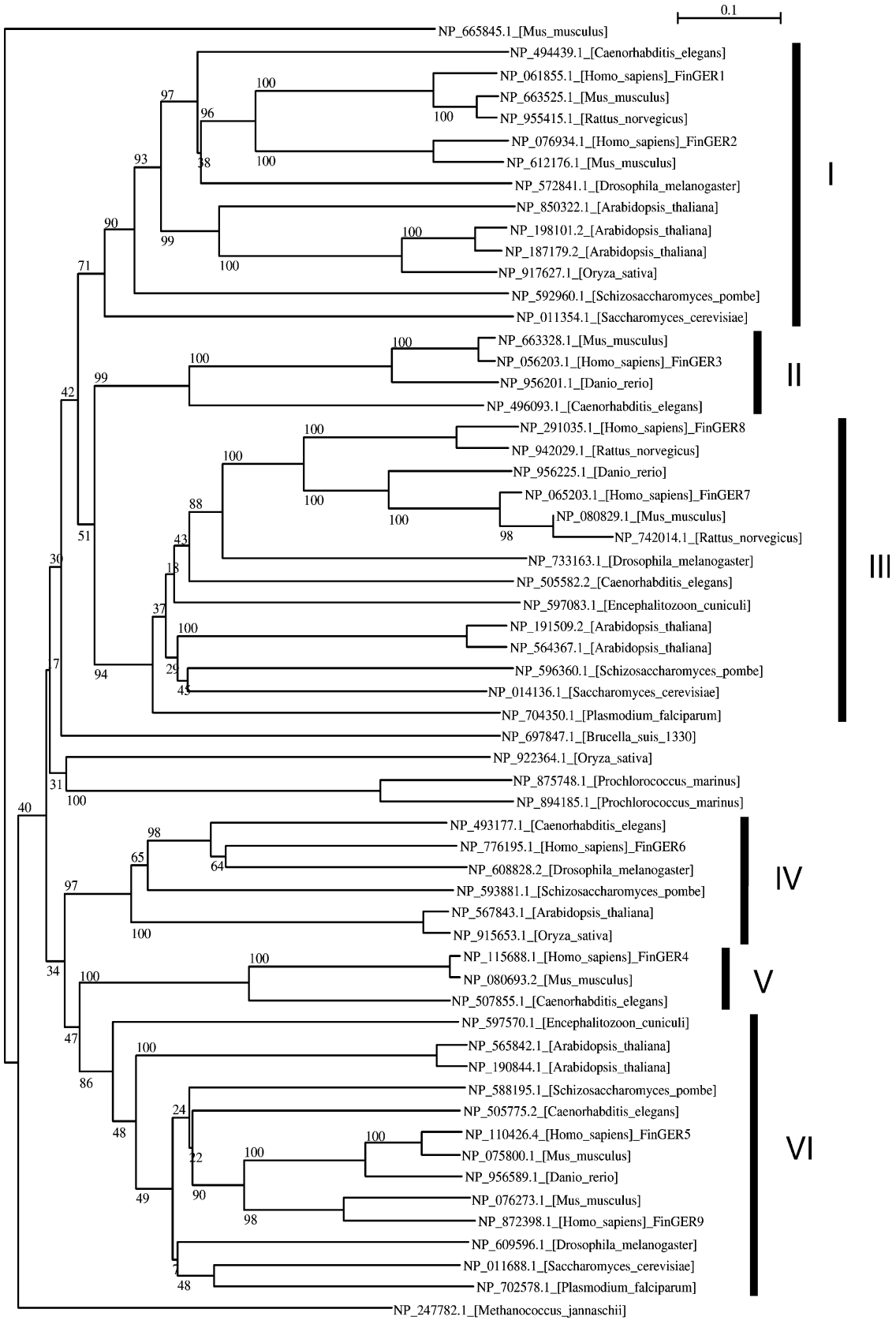

Fig. 3. Phylogenetic tree of proteins homologous to the SMAP-5 protein (FinGER5; NP_110426.4 in cluster VI). Identification of SMAP-5 homologues and calculation of the phylogenetic tree were performed using RiPE. Six different clusters containing the nine human FinGER proteins are shown. The newly identified member of the FinGER protein family termed FinGER9 (GenBank entry NP_872398) is the closest homologue of SMAP-5 and localizes in cluster VI. The mRNAs of all members of this protein family were detectable in cultured human smooth muscle cells.

Table 3
Cluster-specific motifs of the FinGER protein family

Cluster	Pattern 1	Pattern 2	Pattern 3
I	P	DLYGP	GY
II	P	ELYGP	GY
III	P	DLYIP	GY
IV	P	DLWGP	GY
V	P	DFWGP	GY
VI	P	DLAGP	GY

3.5. Conserved motifs

According to PFAM, the Yip1 domain of the FinGER proteins is characterized by the motifs DLYGP and GY. In addition to these motifs, we found a highly conserved proline 10 to 32 amino acids upstream of the DLYGP motif (amino acid position 193 in the alignment shown in supplemental Fig. S1). Although the DLYGP motif is characteristic for the Yip1 domain, each of the clusters shown in Fig. 3 has a specific variation of this motif as listed in Table 3 (see also supplemental Figs. S1 and S2).

3.6. Cellular localization of the protein

Confocal imaging of HeLa cells transiently expressing a C-terminal GFP-tagged SMAP-5 protein revealed that SMAP-5 generally colocalized with GM130, which is characteristic for the *cis*-Golgi, and to a lesser extent with the ER close to the nucleus (data not shown). Transient expression of the GFP-tagged SMAP-5 protein had no visible effects on the Golgi structure. In addition, treatment with 10 ng/ml TGF-β1 during transfection had no effect on the subcellular localization of the GFP-tagged SMAP-5 protein (data not shown).

3.7. Tissue-specific expression

We performed real-time RT-PCR and MTN blots to analyze tissue-specific expression and the correct size of the wild-type SMAP-5 transcript. In the MTN blot normalized with β-actin (Fig. 2), we found a band of about 2 kb which corresponds well to the SMAP-5 transcript as shown in Fig. 1. SMAP-5 mRNA was present at greatest extent in the liver and

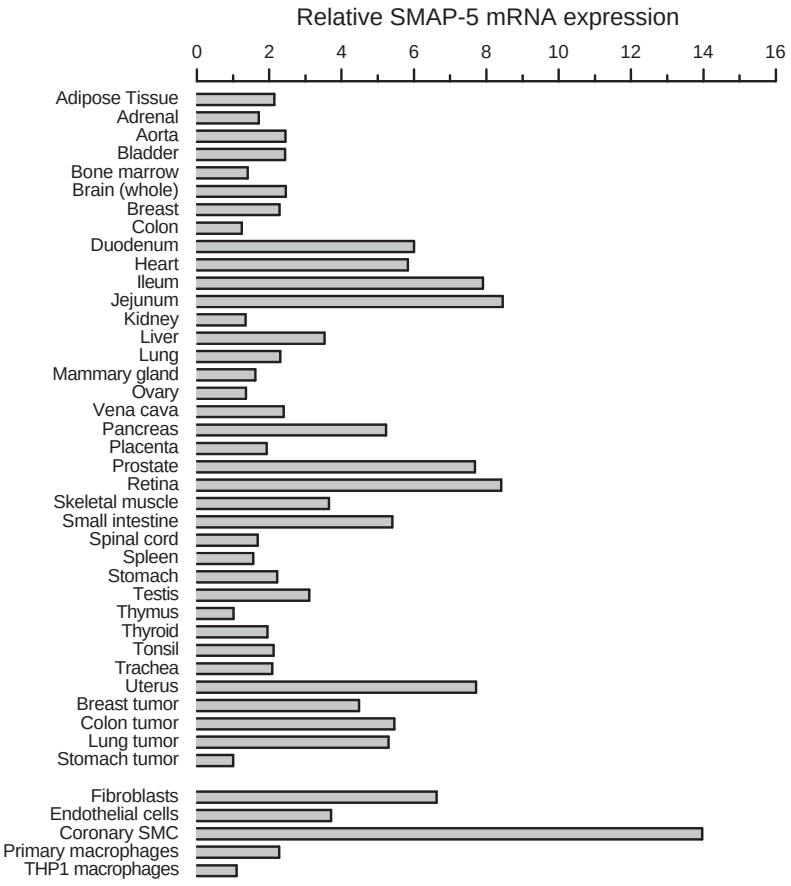

Fig. 4. SMAP-5 mRNA expression in different tissues (top) and cultured cells (bottom) according to real-time RT-PCR. Expression values were normalized to the geometric mean of the three most robust housekeeping genes using GeNorm: 14 kDa signal recognition particle, H3 histone family 3A and DNA directed RNA polymerase II polypeptide K. Fold changes were calculated using SMAP-5 mRNA expression in stomach tumor as a reference set as one. In mRNAs of cultured cells, highest abundance was detected in smooth muscle cells. In tissue mRNAs, high levels of expression of SMAP-5 mRNA was found in duodenum, heart, ileum and jejunum, in prostate and retina, in uterus and all tumor tissues except stomach tumor.

heart, skeletal muscle, kidney and small intestine showed relatively high expression levels. Low or no expression was detected in the other tissues analyzed. In heart and skeletal muscle samples, an additional band of about 7 kb appeared. It is not clear if this was due to unspecific hybridization or to the existence of an unknown splice variant. No clear bands with sizes similar to splice variants A, B and E were detected, indicating that these variants may be of minor importance in vivo. To gain more information about the tissue-specific expression of SMAP-5, we performed real-time RT-PCRs using total RNA from 37 different human tissues and five different human cell types (Fig. 4). For each RNA sample, cDNA synthesis was performed twice and samples were measured in duplicates at least three times. All relative expression levels were calculated using SMAP-5 expression in stomach tumor RNA (showing lowest abundance) as reference. The SMAP-5 transcript was most abundant in duodenum, heart, ileum, jejunum, pancreas, prostate, retina, small intestine, uterus, colon tumor and lung tumor as well as in fibroblasts. The highest level of expression (about 14-fold higher as in stomach tumor tissue) was detected in cultured vascular SMCs. Due to the different normalization methods, expression levels obtained from the MTN blot differ from that obtained by real-time RT-PCR.

3.8. TGF-β1 up-regulates SMAP-5 mRNA

Expression of wild-type SMAP-5 mRNA was analyzed in vascular SMCs treated with TGF-β1 for different times as shown in Fig. 5. RNA was isolated from at least two different experiments and cDNA synthesis was performed at least two times for every RNA sample. Real-time RT-PCR was performed in duplicates at least five times for every time point and fold changes were normalized to GAPDH. As shown in Fig. 5A, TGF-β1 induced SMAP-5 expression in a time-dependent manner. Up-regulation was more than 2-fold after 4 h and statistically significant after 8 h (3.18 ± 1.02; $p < 0.01$) and 24 h (2.70 ± 1.19; $p < 0.05$). Relative changes of GAPDH expression are shown in Fig. 5B. TGF-β1 did not change the expression of mRNAs for the other FinGER proteins in SMCs (data not shown). As expression of FinGER9 mRNA (GenBank entry NM_182592) was very low in vascular SMCs, quantification was impossible.

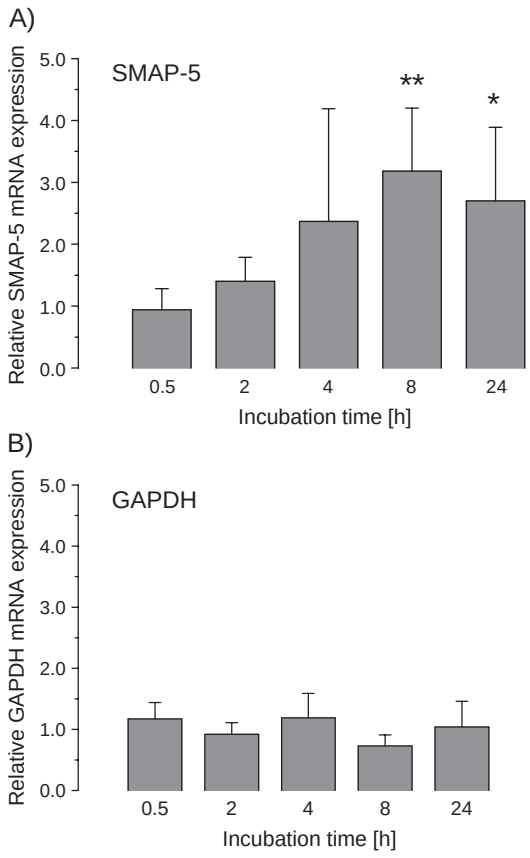

Fig. 5. Regulation of SMAP-5 mRNA expression in human vascular smooth muscle cells during treatment with 10 ng/ml transforming growth factor-β1 (TGF-β1) for different times. (A) The SMAP-5 mRNA was significantly up-regulated after 8 h (3.2 ± 1.0; $p < 0.01$) and 24 h (2.7 ± 1.2; $p < 0.05$). Control cells not treated with TGF-β1 were used as reference and were set therefore to one. Fold changes were calculated to control cells and normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA levels. (B) Expression of GAPDH was not affected by TGF-β1. Real-time RT-PCR analysis revealed that expression of the remaining human FinGER genes was not significantly regulated during TGF-β1 treatment (data not shown). Symbols used are: $*p < 0.05$; $**p < 0.01$.

4. Discussion

We cloned the full-length human SMAP-5 transcript of 2003 bp and identified nine splice variants. Based on our MTN blots, RACE and RT-PCR analyses we suggest that the splice variants are of minor importance in vivo. The identification of the 5′-end of the SMAP-5 mRNA allowed us to characterize the core promoter of the human *SMAP-5* gene. This promoter is GC rich and lacks a TATA box as is typical for promoters of ubiquitously expressed "housekeeping genes". Consistent with this, SMAP-5 was expressed in all monitored human tissues and cell types, but was expressed at highest levels in cultured human coronary SMCs letting us assume that additional transacting factors are involved in tissue- or cell-type-specific expression of SMAP-5.

A comparison of the SMAP-5 protein sequence with the protein sequences of its homologues revealed that, in addition to the known Yip1 domain specific motifs (DLYGP and GY), a highly conserved proline is present in almost all FinGER proteins. Furthermore, the Yip1 specific motif DLYGP can be used to distinguish each of the clusters shown in Fig. 3 due to a specific variation of this motif. However, the exact function of the conserved residues with respect to the FinGER proteins remains unknown.

Our RiPE pipeline allowed us to identify a new member of the FinGER family that showed highest similarity to the SMAP-5 protein. We termed this hypothetical protein FinGER9. According to the GenBank database, there is

experimental evidence for the existence of mRNA for this protein. We detected a very low level mRNA for this gene in vascular SMCs using RT-PCR. All other FinGER mRNAs were expressed at moderate levels.

The closest yeast homologue of SMAP-5 is Yip1p (NP_011688) which has been proposed to perform a critical role in establishing the fusion competence of ER to Golgi vesicles at the time of budding (Barrowman et al., 2003). For this, Yip1p forms a complex with Yif1p (NP_014136). The human homologue of Yif1p (FinGER7, NP_065203) has also been reported to interact with SMAP-5 (Shakoori et al., 2003). This implies that SMAP-5–together with FinGER7–may have a conserved function in the vesicular transport from ER to Golgi in human cells as was reported for the yeast protein (Barrowman et al., 2003; Matern et al., 2000; Yang et al., 1998). This is consistent with the cellular localization of SMAP-5 to ER and *cis*-Golgi which was also shown by Shakoori et al., 2003. Further, the yeast Yip1p is known to bind the Ras-like GTPases Ypt1p and Ypt31p in their GDP-bound conformation (Yang et al., 1998). It is therefore possible that SMAP-5 binds to the human homologues of Ypt1p and Ypt31p (Rab1a and Rab11a, respectively). However, no experimental evidence exists to support this hypothesis.

In coronary SMCs, but not in fibroblasts, expression of SMAP-5 mRNA was significantly induced by TGF-β1, whereas expression of the other FinGER genes was not affected. The function of the *SMAP-5* gene is not known, but evidence exists that it is part of the ER-to-Golgi transport pathway. Due to the ability of TGF-β1 to induce the expression of SMAP-5, we speculate that SMAP-5 may be involved in the secretion of components of the extracellular matrix such as collagens because synthesis and secretion of these proteins are also induced by TGF-β1. However, further experiments are required to elucidate the role of SMAP-5 in this respect. Nevertheless, SMAP-5 may be involved in TGF-β1-mediated processes and may therefore be an interesting target in conditions such as inflammation, cancer, and arteriosclerosis.

Acknowledgements

We thank Thomas Böking and Christine Fabritius for their excellent technical assistance. The study was supported by the Deutsche Forschungsgemeinschaft and the European Union (Bellosta et al., 2002). This work forms part of the PhD thesis of Katrin Stolle which was performed at the University of Münster.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.03.0127.

References

Barrowman, J., Wang, W., Zhang, Y., Ferro-Novick, S., 2003. The Yip1p–Yif1p complex is required for the fusion competence of endoplasmic reticulum-derived vesicles. J. Biol. Chem. 278, 19878–19884.

Bellosta, S., et al., 2002. Macrophage function and stability of the atherosclerotic plaque: progress report of a European project. Nutr. Metab. Cardiovasc. Dis. 12, 3–11.

Breathnach, R., Chambon, P., 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. 50, 349–383.

Byvatov, E., Schneider, G., 2003. Support vector machine applications in bioinformatics. Appl. Bioinformatics 2, 67–77.

Calero, M., Collins, R.N., 2002. *Saccharomyces cerevisiae* Pra1p/Yip3p interacts with Yip1p and Rab proteins. Biochem. Biophys. Res. Commun. 290, 676–681.

Calero, M., Whittaker, G.R., Collins, R.N., 2001. Yop1p, the yeast homolog of the polyposis locus protein 1, interacts with Yip1p and negatively regulates cell growth. J. Biol. Chem. 276, 12100–12112.

Calero, M., Winand, N.J., Collins, R.N., 2002. Identification of the novel proteins Yip4p and Yip5p as Rab GTPase interacting factors. FEBS Lett. 515, 89–98.

Chen, W., Wahl, S.M., 2003. TGF-beta: the missing link in CD4+CD25+regulatory T cell-mediated immunosuppression. Cytokine Growth Factor Rev. 14, 85–89.

Christ, M., et al., 1994. Immune dysregulation in TGF-beta 1-deficient mice. J. Immunol. 153, 1936–1946.

Fuellen, G., Spitzer, M., Cullen, P., Lorkowski, S., 2003. BLASTing proteomes, yielding phylogenies. In Silico Biol. 3, 313–319.

Gougeon, P.Y., Prosser, D.C., Da-Silva, L.F., Ngsee, J.K., 2002. Disruption of Golgi morphology and trafficking in cells expressing mutant prenylated rab acceptor-1. J. Biol. Chem. 277, 36408–36414.

Howe, K., Bateman, A., Durbin, R., 2002. QuickTree: building huge neighbour-joining trees of protein sequences. Bioinformatics 18, 1546–1547.

Kulkarni, A.B., et al., 1993. Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death. Proc. Natl. Acad. Sci. U. S. A. 90, 770–774.

Letunic, I., et al., 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 32, D142–D144.

Lorkowski, S., et al., 2001a. Genomic sequence and structure of the human ABCG1 (ABC8) gene. Biochem. Biophys. Res. Commun. 280, 121–131.

Lorkowski, S., et al., 2001b. Expression of the ATP-binding cassette transporter gene ABCG1 (ABC8) in Tangier disease. Biochem. Biophys. Res. Commun. 283, 821–830.

Matern, H., Yang, X., Andrulis, E., Sternglanz, R., Trepte, H.H., Gallwitz, D., 2000. A novel Golgi membrane protein is part of a GTPase-binding protein complex involved in vesicle targeting. EMBO J. 19, 4485–4492.

McKnight, S., Tjian, R., 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell 46, 795–805.

Morgenstern, B., 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15, 211–218.

Nakai, K., Horton, P., 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24, 34–36.

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

Pisano, M.M., Mukhopadhyay, P., Greene, R.M., 2003. Molecular fingerprinting of TGFbeta-treated embryonic maxillary mesenchymal cells. Orthod. Craniofac. Res. 6, 194–209.

Roes, J., Choi, B.K., Cazac, B.B., 2003. Redirection of B cell responsiveness by transforming growth factor beta receptor. Proc. Natl. Acad. Sci. U. S. A. 100, 7241–7246.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Shakoori, A., et al., 2003. Identification of a five-pass transmembrane protein family localizing in the Golgi apparatus and the ER. Biochem. Biophys. Res. Commun. 312, 850–857.

Shull, M.M., et al., 1992. Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease. Nature 359, 693–699.

Sivars, U., Aivazian, D., Pfeffer, S.R., 2003. Yip3 catalyses the dissociation of endosomal Rab-GDI complexes. Nature 425, 856–859.

Spitzer, M., Fuellen, G., Cullen, P., Lorkowski, S., 2004. VisCoSe: visualization and comparison of consensus sequences. Bioinformatics 20, 433–435.

Stolle, K., Weitkamp, B., Rauterberg, J., Lorkowski, S., Cullen, P., 2004. Laser-microdissection-based analysis of mRNA expression in human coronary arteries with intimal thickening. J. Histochem. Cytochem. 52, 1511–1518.

Topper, N.J., 2000. TGF-beta in the cardiovascular system: molecular mechanisms of a context-specific growth factor. Trends Cardiovasc. Med. 10, 132–137.

Vandesompele, J., et al., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, research0034.1–research0034.11.

Verrecchia, F., Chu, M.L., Mauviel, A., 2001. Identification of novel TGF-beta/Smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. J. Biol. Chem. 276, 17058–17062.

Walter, M., Gerdes, U., Seedorf, U., Assmann, G., 1994. The high density lipoprotein-and apolipoprotein A-I-induced mobilization of cellular cholesterol is impaired in fibroblasts from Tangier disease subjects. Biochem. Biophys. Res. Commun. 205, 850–856.

Yang, X., Matern, H.T., Gallwitz, D., 1998. Specific binding to a novel and essential Golgi membrane protein (Yip1p) functionally links the transport GTPases Ypt1p and Ypt31p. EMBO J. 17, 4954–4963.