ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Gene xx (2003) 1–13

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Comparative sequence and expression analyses of four mammalian *VPS4* genes

Andreas Beyer[a], Sibylle Scheuring[a], Sibylle Müller[a], Antoaneta Mincheva[b], Peter Lichter[b], Karl Köhrer[a,*]

[a]*Biologisch-Medizinisches Forschungszentrum, Heinrich-Heine-Universität Düsseldorf, Moorenstraße 5, D-40225 Düesseldorf, Germany*
[b]*Deutsches Krebsforschungszentrum, Abt. Organisation komplexer Genome, DKFZ, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany*

Received 19 August 2002; received in revised form 21 November 2002; accepted 4 December 2002
Received by B. Dujon

Abstract

The *VPS4* gene is a member of the AAA-family; it codes for an ATPase which is involved in lysosomal/endosomal membrane trafficking. *VPS4* genes are present in virtually all eukaryotes. Exhaustive data mining of all available genomic databases from completely or partially sequenced organisms revealed the existence of up to three paralogues, *VPS4a*, *-b*, and *-c*. Whereas in the genome of lower eukaryotes like yeast only one *VPS4* representative is present, we found that mammals harbour two paralogues, *VPS4a* and *VPS4b*. Most interestingly, the Fugu fish contains a third *VPS4* paralogue (*VPS4c*). Sequence comparison of the three *VPS4* paralogues indicates that the Fugu *VPS4c* displays sequence features intermediate between *VPS4a* and *VPS4b*. Using complete mammalian *VPS4a* and *VPS4b* cDNA clones as probes, genomic clones of both *VPS4* paralogues in human and mouse were identified and sequenced. The chromosomal loci of all four *VPS4* genes were determined by independent methods. A BLAST search of the human genome database with the human *VPS4A* sequence yielded a double match, most likely due to a faulty assembly of sequence contigs in the human draft sequence. Fluorescent *in situ* hybridization and radiation hybrid analyses demonstrated that human and mouse *VPS4A/a* and *VPS4B/b* are located on syntenic chromosomal regions. Northern blot and semi-quantitative reverse transcription analyses showed that mouse *VPS4a* and *VPS4b* are differentially expressed in different organs, suggesting that the two paralogues have developed different functional properties since their divergence. To investigate the subcellular distribution of the murine VPS4 paralogues, we transiently expressed various fluorescent VPS4 fusion proteins in mouse 3T3 cells. All tested VPS4 fusion proteins were found in the cytosol. Expression of dominant-negative mutant VPS4 fusion proteins led to their concentration in the perinuclear region. Co-expression of VPS4a-GFP and VPS4b-dsRed fusion proteins revealed a partial co-localization that was most prominent with mutant VPS4a and VPS4b proteins. A physical interaction between the mouse paralogues was also supported by two-hybrid analyses.

Keywords: AAA family; Endocytosis; Vacuolar protein sorting; VPS4-paralogue

1. Introduction

One of the most important features of the eukaryotic cell is its capability to transport cargo to defined destinations. During the last two decades, protein targeting and membrane traffic mechanisms have been dissected to a large extent (Lemmon and Traub, 2000). Those investigations made clear that not only the whole process but also each step of it is very complex. For example, in vacuolar/lysosomal biogenesis two transport pathways converge, namely the biosynthetic (Golgi-to-lysosome) and the endocytic one (plasma membrane-to-lysosome). A crucial step is the formation of multivesicular bodies (MVBs) which occurs by invagination of vesicles into the lumen of late endosomes (Felder et al., 1990). During their maturation, their cargo is sorted for either recycling back to the plasma membrane or transport to the lysosome (Gruenberg and Maxfield, 1995). The latter pathway is further subdivided. When MVBs fuse with lysosomes, their membrane proteins are delivered to the limiting lysosomal

Abbreviations: aa, amino acid(s); AAA, ATPases associated with diverse cellular activities, http://yeamob.pci.chemie.uni-tuebingen.de/AAA/; bp, base pair(s); EST(s), expressed sequence tag(s); FISH, fluorescence in situ hybridization; FITC, fluorescein isothiocyanate; GFP, green fluorescent protein; kb, kilobase(s); MVB, multivesicular body; nt, nucleotide(s); ORF, open reading frame; RT–PCR, reverse transcription–polymerase chain reaction; vps, vacuolar protein sorting defect; WT, wild-type; Hs, *Homo sapiens*; Mm, *Mus musculus*.

* Corresponding author. Tel.: +49-211-811-3165; fax: +49-211-811-1922.
E-mail address: koehrer@uni-duesseldorf.de (K. Köhrer).

0378-1119/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0378-1119(02)01205-2

membrane while internal vesicles are released to the lysosomal lumen (Futter et al., 1996). In this way, membrane proteins, too, are directed to vacuolar/lysosomal degradation. Hence, MVB sorting provides a key step for down regulation of membrane receptors, which is an important means of controlling signalling pathways (Futter et al., 2001). In yeast, genetic screens yielded mutants defective in vacuolar protein sorting (vps). They are grouped into the so-called classes A to E, dependent on the vacuolar morphology (Banta et al., 1988; Raymond et al., 1992). It appears that class E mutants – 15 such genes are known to date – are somehow defective in MVB formation. Mutant cells exhibit a vps phenotype, transport of proteins out of a prevacuolar endosomal compartment is compromised (Conibear and Stevens, 1998). Exaggerated prevacuolar compartments, likely corresponding to the defective MVB compartment (Odorizzi et al., 1998), accumulate vacuolar precursor material and endocytosed cargo destined to the vacuole for degradation (Babst et al., 1997; Finken-Eigen et al., 1997). Moreover, soluble vacuolar proteins are secreted, and since retrograde protein traffic to the *trans*-Golgi complex is also compromised, a late Golgi defect is observed, too (Nothwehr et al., 1996). Further investigations of the class E mutants revealed very interesting features, namely the fact that phosphatidylinositol phosphate as well as ubiquitin are involved in vacuolar protein sorting. In mammals and yeast it has been shown that phosphatidylinositol 3,5-bisphosphate is essential for MVB formation (Gary et al., 1998; Fernandez-Borja et al., 1999, respectively). Knocking out the responsible PI kinase, Fab1p, causes a class E vps phenotype. Further, there are several reported interactions between class E vps proteins (Babst et al., 2000). Recently, it has been demonstrated that ubiquitination is a sorting signal needed during MVB formation. The ubiquitin receptor in yeast was identified and called ESCRT-I complex (Katzmann et al., 2001), consisting of Vps23p, Vps28p, and Vps37p. Vps4p, which is also a class E vps protein, acts downstream of ESCRT-I in the context of formation/maturation of MVBs. Vps4p is a member of a subgroup of Walker-type ATPases, the so-called AAA-family (ATPases associated with diverse cellular activities) (Beyer, 1997). It harbours a single AAA-cassette following an N-terminal coiled-coil domain (Babst et al., 1997; Scheuring et al., 2001). Database searches for Vps4p-related proteins yielded murine SKD1 as closest mammalian relative (59% sequence identity to Vps4p) (Périer et al., 1994). It was shown to fulfil a similar function like Vps4p (Scheuring et al., 1999). However, by EST-clustering of mammalian cDNA sequences, surprisingly two different paralogues of yeast Vps4p could be identified (Scheuring et al., 2001). We had chosen the names *VPS4a* and *VPS4b* for these two mammalian paralogue genes, which, as a pair, are orthologue to yeast *VPS4*. *SKD1* is identical to murine *VPS4b*, but since 'SKD' refers to the screening procedure in which it was identified and does not refer to its biological function, we prefer the VPS4a/b nomenclature. Several features of the VPS4 protein have been worked out already. The ATPase activity is essential for its function (Babst et al., 1997). The vps phenotype of *vps4* null mutant strains is also achieved by dominant-negative *vps4* alleles with single amino acid exchanges in the conserved AAA-cassette (Scheuring et al., 2001). For human VPS4A it has been shown that mutant protein variants defective in ATP hydrolysis display a punctate localization at lysosomal class E compartment membranes, whereas the WT form mainly locates to the cytosol (Bishop and Woodman, 2000). Strong evidence has been provided that the VPS4 protein regulates the membrane association of other class E vps proteins (Babst et al., 1998; Bishop and Woodman, 2000). Very recently, another aspect has been added to vacuolar/lysosomal biogenesis: It has been demonstrated that human VPS4A interacts with the RND2 protein (a Rho family GTPase) and both partly co-localize to early endosomes (Tanaka et al., 2002).

Here we report the molecular cloning, chromosomal localization, sequencing, comparative sequence analysis and expression analyses of four mammalian *VPS4* genes, i.e. both paralogues in human and mouse.[1]

2. Materials and methods

2.1. Materials

Amersham Pharmacia Biotech (Freiburg, Germany) was the supplier for all DNA restriction and modifying enzymes, deoxynucleotides, and Taq-DNA polymerase. Oligonucleotides for DNA sequencing and for molecular cloning purposes were ordered from Gibco. Kits for purification of DNA and RNA were purchased from Qiagen (Hilden, Germany). Standard molecular biology techniques (DNA manipulation, basic PCR protocols, Northern blot analyses) were performed using standard protocols (Ausubel et al., 1990). Northern blots containing 2 μg poly(A)$^+$-RNA/lane were purchased from Clontech. *Escherichia coli* cells were transformed employing standard electroporation procedures with a BioRad Gene Pulser (Becker and Guarente, 1991).

2.2. Used strains and media

The *E. coli* strains used for cloning and propagation of plasmid constructs as well as PACs were XL1blue and DH10B, respectively. *E. coli* cells were grown in standard LB, LB-Amp (supplemented with 100 μg/ml ampicillin) or LB-Kan (supplemented with 50 μg/ml kanamycin) media at 37 °C (Miller, 1972).

[1] Names were chosen according to the nomenclature guidelines of the HUGO/GDB Nomenclature Committee (http://www.gene.ucl.ac.uk/nomenclature/) and the mouse Locus Nomenclature Committee (http://www.jax.org/): *VPS4A* and *VPS4B* for human and *VPS4a* and *VPS4b* for murine *VPS4* genes.

2.3. Isolation and sequencing strategy of genomic clones

Screening of genomic PAC-libraries (human library RPCIP704/vector pPAC2 and murine library RPCIP711/vector pPAC4) was performed by the RZPD, Berlin, Germany (http://www.rzpd.de). For that purpose, we have provided the RZPD with the complete ORFs of all four *VPS4* genes, amplified from respective cDNA clones by PCR, to serve as probes. We checked the selected genomic clones for presence of the complete genes by two methods. First, fragments located in the middle of the gene as well as in both non-coding regions were amplified by PCR (data not shown). Then the positive clones were verified by partial sequencing. DNA sequencing reactions were carried out with the BigDye RR Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Weiterstadt, Germany) and analyzed with an ABI Prism DNA Analysis System 377. Sequence assembly and processing were performed with the Lasergene DNASTAR software package (GATC, Konstanz, Germany). Since with the exception of the first exon of murine *VPS4a* all exons of the genes were known, we decided to undertake a primer-walking strategy starting from the exons rather than a shotgun-approach for PAC clone sequencing. As templates, we employed PAC-DNA or we sequenced PCR fragments directly. Roughly 2% of the sequence, that could not be obtained by this strategy, was subcloned into pBluescript KS + (accession no. X52327) for sequencing. Both strands were sequenced completely, either from the PAC-clone itself, subclones in pBluescript KS + , or PCR fragments. This principle bears the risk of PCR-borne or cloning artefacts, especially in regions rich in repetitive DNA (e.g. Alu). In order to be able to detect such artefacts, we made sure to obtain a contiguous sequence – at least in single-strand coverage – from the original PAC-clones.

2.4. Comparative sequence analysis

Database searches were performed in the NBRF, PIR, SwissProt, GenBank, and EMBL databases, using the FASTA and TFASTA programs of the University of Wisconsin Genetics Computer Group sequence analysis package (Devereux et al., 1984). We also searched at the NCBI, using the BLAST network service (Altschul et al., 1990; Gish and States, 1993). Protein sequences were also searched for known protein sequence motifs with the MacPattern program (Fuchs, 1991), the PROSITE dictionary (Bairoch, 1991), and the BLOCKS network service (Henikoff and Henikoff, 1994). Pairwise sequence aligns and dot matrix analysis were performed with GENEPRO (Genepro Riverside Scientific, t.A., Seattle, WA, USA), multiple alignments and neighbour-joining tree construction with ClustalW (Thompson et al., 1994). Repetitive sequence elements were detected with RepeatMasker (http://www.genome.washington.edu/UWGC/analysistools/repeatmask.htm). Promoter analysis was performed with Promoter Scan Service (http://bimas.dcrt.nih.gov/molbio/signal/) and GeneFinder (http://dot.imgen.bcm.tmc.edu) as well as a comparison to the Eukaryotic Promoter Database at the NIH (http://www.ncbi.nlm.nih.gov/).

2.5. Alignment of genomic sequences

Since comparison of long sequences containing greater portions without detectable similarity (i.e. in introns) is not a simple task, alignments were done semi-manually. First, regions of significant similarity ($>$ 40% sequence identity in $>$ 100 nt) were detected in a dot-matrix and subsequently local alignments were performed and saved. Then, a program written in C++ was employed to fit the local aligns into the genomic sequences and to export the align as TIF-file. In that way sequence similarity, intron–exon-structure, and distribution of repeats could be visualized.

2.6. Protein alignments

For protein sequence comparison, all available VPS4 protein representatives were aligned using ClustalW. The ones from *Xenopus laevis* and *Gallus gallus* were assembled completely by manual EST clustering. The putative Fugu fish and *Caenorhabditis elegans* proteins had obviously been predicted incorrectly, hence they were assembled manually by comparison to the other VPS4 representatives from BLAST-searches in the Fugu genome draft and the *C. elegans* genome, respectively.

2.7. Chromsomal localization

Radiation hybrid assays were performed to determine chromosomal localization (Research Genetics, Huntsville, AL, USA, T31-RH04.02 mouse panel, http://lena.jax.org/resources/documents/cmdata/, http://www.informatics.jax.org/menus/map_menu.shtml). The oligonucleotide pairs used were m1a29/m1a30 (GAGTGTGGACCTGCAGAGAC/CGTGGAGGGAACTGCCTGCTGGC) and m1a42/m1a45 (CAGTGAGTACCTGTCAGAG/CTCTCACTGAAGTCTCCTG) for mouse *VPS4a*; as well as G63/G74 (CACAGTTAAGGAAGCCAGG/ACCCTTTGTTACAAAGTAGC) and G49/G40 (GCCACACAGCATGGTGAG/GAGAAGCCATAGGACAACG) for mouse *VPS4b*. Chromosomal localization was also investigated by fluorescence in situ hybridization (FISH). Clones selected for FISH were the ones taken for sequencing (see Section 3.1) except for murine *VPS4b*: Here, a pre-existing phage clone was taken (Scheuring et al., 1999). A 13.3 kb *Not*I/*Xho*I fragment (positions 15,875–27,115 of the published sequence, accession no. AF134119) covering exons 5 through 11 was cloned into vector pBluescript KS + . The respective DNAs were biotin-labelled by nick translation and used as probes for chromosomal in situ suppression hybridization to metaphase chromosomes as described previously (Lichter et al., 1990). Seventy nanograms of each labelled probe was

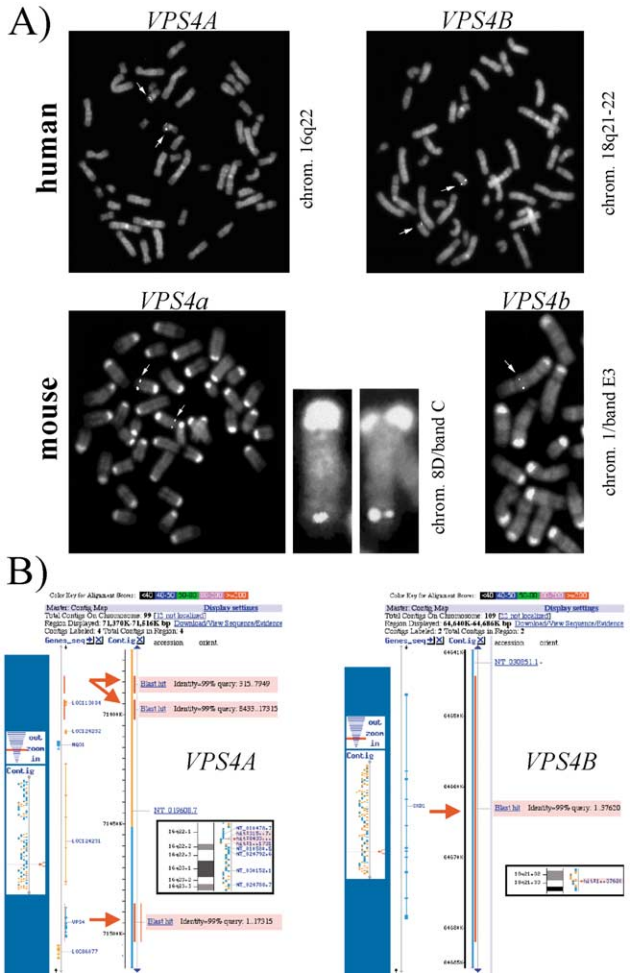


Fig. 1. Localization of the *VPS4*-paralogues in the genomes of human and mouse. (A) Genomic mapping of mammalian *VPS4A/a* and *-B/b* by FISH analyses. Hybridized probes were detected via FITC (arrows) and chromosomes were counterstained with DAPI. Top: Mapping of human *VPS4A* to chromosome band 16q22 and *VPS4B* to chromosome band 18q21–22. Bottom: Mapping of mouse *VPS4a* to chromosome 8D (on the left panel, one complete metaphase spread is shown, whereas on the right two examples of chromosome 8 are depicted to illustrate the band assignment) and mouse *VPS4b* to chromosome 1E3 (section of a metaphase spread is shown). (B) Blasting of human *VPS4A* and *-B* against the human genome draft sequence (April 2002). Screenshots of BLAST-outputs are shown. From left to right: Contig maps, ruler, hits (red arrows), percentage identity and band pattern are given.

combined with 3 μg of mouse Cot-1 or human COT-1 DNA and 7 μg salmon sperm DNA in a 10 μl hybridization cocktail, and hybridized to mouse chromosomes prepared from spleen cells of a female Balb/c animal (according to Sawyer et al., 1987) or to human lymphocyte metaphase chromosomes, respectively. The hybridized probes were detected via fluoresceinated avidin (FITC conjugated to avidin, arrows in Fig. 1A). To confirm the murine chromosomal mapping, probe *VPS4a* (mouse) was co-hybridized with a probe specific for mouse chromosome 8, labelled with digoxigenin and detected via rhodamine. In all cases, chromosomes were counterstained with 4,6-diamino-2-phenylindole-dihydrochloride (DAPI). Digitized images of emitted DAPI, FITC and rhodamine fluorescence were recorded separately by a CCD camera (Photometrics), electronically overlaid and carefully aligned. Finally, chromosomal localization was also determined by searching of the human genome draft at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/).

2.8. Plasmid construction

A mouse cDNA pool was used to amplify the complete ORF of mouse *VPS4a*, which was subcloned into the *Eco* RI and *Asp*718 sites of the yeast expression vector pJD302 (Dohmen et al., 1995). To introduce a single point mutation in the Walker B motif at amino acid 228 (from glutamic acid to glutamine) PCR-based site directed mutagenesis was performed with the QuikChange Site-Directed Mutagenesis Kit from Stratagene according to the supplied manual using oligonucleotides m1a-d228f CCATCATCTTCATTGATGAGGTAGATTCCCTCTGTGGGTC and m1a-d228r GACCCACAGAGGGAATCTACCTCATCAATGAAGATGATGG, and confirmed by DNA sequencing. Mouse *VPS4b* expression constructs were generated in a similar way, which has been described previously (Scheuring et al., 1999).

In order to trace cellular localization of VPS4 proteins in murine cells, we constructed GFP and dsRed fusion constructs of both WT and dominant negative *VPS4* alleles. Using the TOPO technology (Invitrogen), PCR products of WT *VPS4a/b* as well as *VPS4a* ^{E228Q}/*VPS4b* ^{E235Q} ORFs were cloned into the mammalian expression vectors pcDNA3.1CTGFP and pcDNA3.1NTGFP, yielding constructs of WT and mutant VPS4a/b with GFP fused to their C- or N-terminal ends. The expression vector pdsRedN1 (Clontech Laboratories) was employed to obtain dsRed fusion constructs. PCR products of *VPS4a/b* and *VPS4a* ^{E228Q}/*VPS4b* ^{E235Q} ORFs were ligated to the *Eco* RI and *Asp*718 sites of the vector, to generate pVPS4a-Red, pVPS4b-Red, pVPS4a^{E228Q}-Red and pVPS4b^{E235Q}-Red, all carrying a C-terminal dsRed.

2.9. Cell culture and transfection

NIH/3T3 cells (mouse embryonic fibroblasts, DSMZ GmbH, Braunschweig, Germany) were grown in Dulbecco's modified Eagle's Medium (DMEM) containing 10% foetal calf serum (Jainchill et al., 1969) under standard cell culture conditions. For transient transfections, cells were seeded on coverslips in 12-well plates. Next day, the cells were transfected with 300 ng of plasmid DNA (single transfection) or 200 ng each (for co-transfection) employing Effectene (Qiagen) according to the supplied manual. 48 h after transfection, cells were washed with PBS and fixed in PBS/4% paraformaldehyde for 10 min. Coverslips were mounted with Vectashield (Vector Laboratories) and examined by confocal microscopy (Zeiss LSM 510).

2.10. Semiquantitative RT–PCR

RNA expression levels of endogenous mouse *VPS4a* and -*b* in NIH/3T3 cells were determined by semi-quantitative RT–PCR. RNA of NIH/3T3 cells (log-phase) was prepared with the RNeasy Protect Kit (Qiagen). Reverse transcription was performed with gene specific primers for mouse *VPS4a* (GAAGTCCCATGAATTTTCTC) and mouse *VPS4b* (CTGCGTAGACAAGGCAAAGATG) using the Sensiscript RT Kit (Qiagen). Resulting cDNAs were amplified with mouse *VPS4a* and -*b* specific primers (m*VPS4a*-1/2: CATGCAAACCACAGGCTC/GTTTGAGCTGGCCCGGC; m*VPS4b*-1/2: GGAAGCTCTTCAACTCTAC/AACCACACAGAGAATCAATC) leading to a 576 bp *VPS4a* and 640 bp *VPS4b* specific PCR product. Endogenous β-actin served as a control.

2.11. Anchored vector PCR of murine Vps4a cDNA

All cDNA sequences had previously been assembled, as far as possible, in silico by EST-clustering. In mouse *Vps4a*, however, just the first five codons, representing the coding part of exon 1, were not covered by ESTs. Since such a small part of the ORF hardly can be detected in the genomic sequence, we decided to amplify it from a mouse cDNA library. Tissue-specific mouse cDNA pools cloned in pSport were obtained from the RZPD, Berlin. The first amplification of these cDNA pools was performed with a distal vector primer (vector-reverse: AGCGGATAACAATTTCACACAGG or vector-forward: CCCAGTCACGACGTTGTAAAACG) and a gene-specific primer for mouse *VPS4a* (m1a4: CAGCTACATCATTCCACCG), located in exon 5, and followed by a nested PCR with internal vector (T7 or SP6 primer) and gene-specific primer (m1a14: GCCTTCACTCTGGTTCTC). The cDNA pools of fetal thymus yielded a 300 bp amplicon. The sequence of this PCR product completed the ORF, providing the first five missing codons.

3. Results

3.1. Isolation of genomic clones of the human and murine VPS4 paralogues

Starting from the cDNA clones, we decided to investigate the genomic organization of the corresponding genes. The genomic sequence of murine *VPS4b* has been described previously (Scheuring et al., 1999). In contrast to yeast, database analyses suggested that mouse and man possess a paralogue pair of *VPS4* genes. Based on clustered EST contigs, we designed oligonucleotides and generated PCR fragments of the complete ORFs. These probes were employed to identify genomic clones in murine and human PAC libraries at the RZPD (Berlin, Germany). Between six and 14 genomic clones were obtained for each of the four *VPS4* representatives. Almost all of them contained the respective genes at least in part. The clones chosen after all for complete sequence determination were H20506 Q2 (human *VPS4A*, accession no. AF282903), K081120 Q2 (human *VPS4B*, accession no. AF282904) and P10528 Q2 (mouse *VPS4a*, accession no. AF530161). The genomic sequence of murine *VPS4b* has been described previously (Scheuring et al., 1999). Nevertheless, in order to obtain complete PAC clones of that gene, library RPCIP711 had been screened for it, too. The clones P19548 Q2 and B06240 Q2 were verified by partial sequencing, both contain *VPS4b* completely.

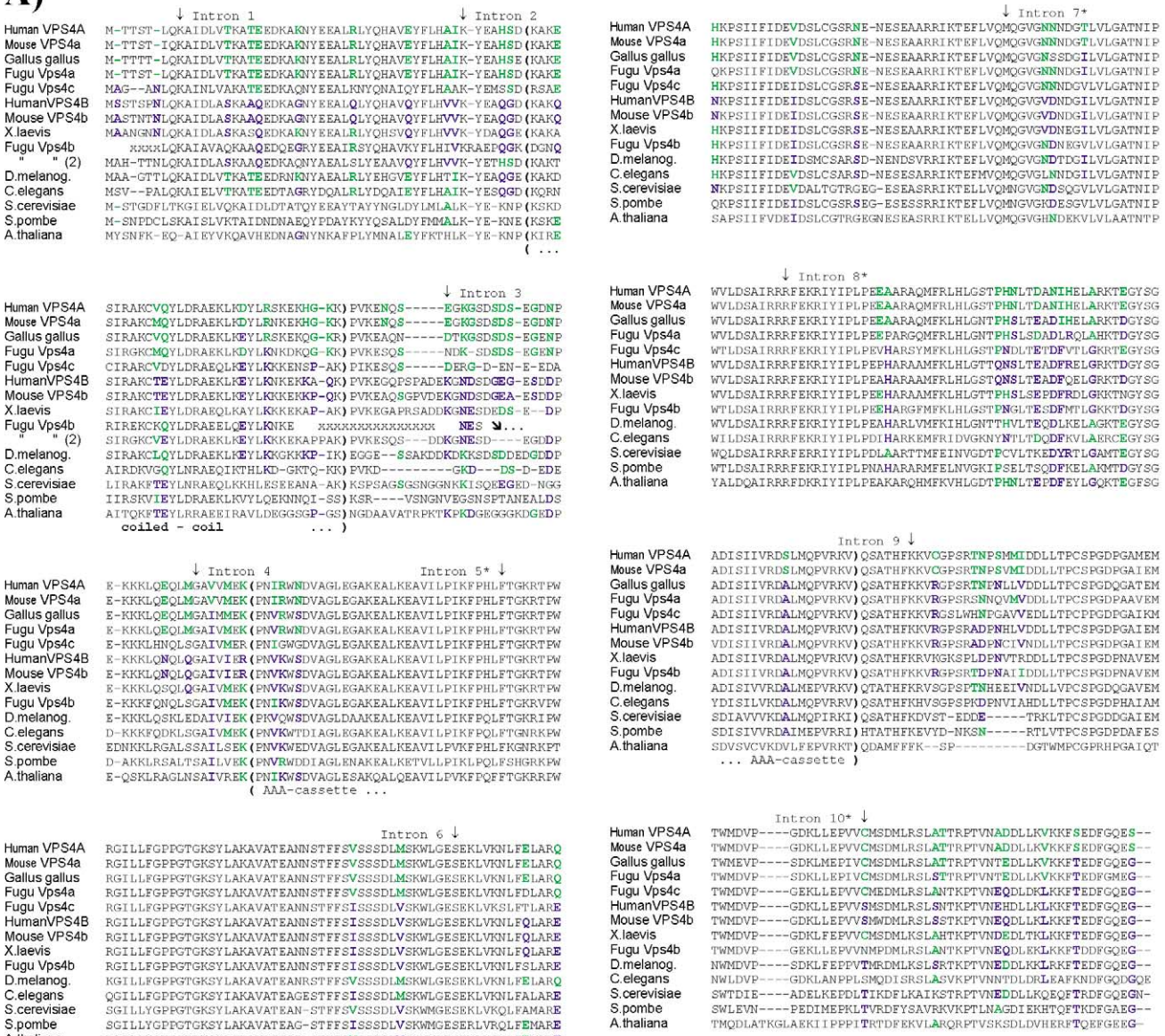
3.2. Chromosomal mapping of the VPS4 PAC clones in human and mouse

To identify the genomic context of the *VPS4* paralogues, i.e. in order to investigate human-mouse synteny, we determined the chromosomal loci of the four genes. The genomic clones chosen for sequencing (see Section 3.1) served as probes in fluorescence in situ hybridization (FISH) analyses of human and mouse chromosomes. Microscopic evaluation revealed highly specific signals of *VPS4a* on mouse Chromosome 8D in the distal part of band C and with probe *VPS4b* on mouse chromosome 1 in band E3 (Fig. 1A). The efficiency of the hybridization was high, as the signals were found on both homologue chromosomes in 93% (26/28) for *VPS4a* and in 74% (20/27) for *VPS4b*. Chromosome assignment of the *VPS4b* probe was also confirmed by dual colour FISH with a simultaneously hybridized probe specific for chromosome 8 (data not shown). FISH experiments of human *VPS4A* and *VPS4B* probes revealed signals on chromosome 16q22 and 18q21–22, respectively. In 90% (27/30) of the metaphase spreads hybridized with the *VPS4A* probe, fluorescent signals were detected in 16q22 on both homologues (Fig. 1A). In both experiments no additional signals in other regions of the human genome were observed.

Chromosomal locations were confirmed by two other methods: Fig. 1B shows the results of blasting the two human genes (with repetitive portions of the sequences being masked) against the human draft sequence. Both hits agree with the results of the FISH analyses: assignments of human *VPS4A* and *VPS4B* to chromosomes 16q22 and 18q21, respectively. However, while *VPS4B* produces only one high-scoring hit, *VPS4A* has got two in close vicinity, one of which is imperfect. Nevertheless, Southern blot analyses with several restriction enzymes did not provide any evidence for two *VPS4A* copies in the human genome (data not shown). Radiation hybrid analyses of mouse *VPS4a* and *VPS4b*, too, supported the given localization (see http://www.jax.org/resources/documents/cmdata/rhmap/rh.html) (Rowe et al., 2000).

Meanwhile, the analysis of the murine genome has proceeded to > 90% coverage in draft quality (see http://www.ncbi.nlm.nih.gov/genome/guide/mouse/), so that enough data is available to trace synteny between man

A)

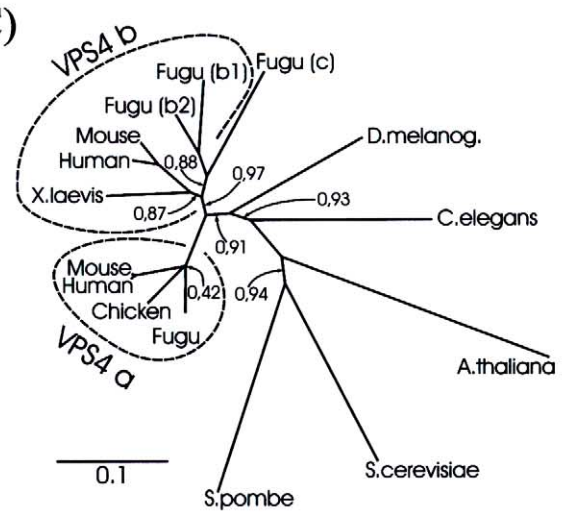


B)

Assignment of positions to mammalian a:b paralogues

Positions typical of	VPS4a : VPS4b	
human VPS4A	72 : 0	
mouse VPS4a	72 : 0	
Gallus gallus (Vps4a)	51 : 15	
Fugu Vps4a	47 : 14	(Scaffold 2013)
Fugu Vps4c	19 : 24	(Scaffold 145)
human VPS4B	0 : 67	
mouse VPS4b	1 : 68	
X.laevis (Vps4b)	14 : 44	
Fugu Vps4b(2)	13 : 37	(Scaffold 1117)
D.melanogaster	27 : 27	
C.elegans	21 : 21	
S.cerevisiae	11 : 22	
S.pombe	10 : 18	
A.thaliana	14 : 20	

C)



and mouse (Lander et al., 2001). *VPS4A/a* lies in a chromosomal region, which has not been shuffled between both organisms: roughly half of the long arm of human chromosome 16 is syntenic to a great portion of murine chromosome 8. The part of human chromosome 18, however, in which *VPS4B* is located, has obviously been rearranged several times since the divergence of primates and rodents, the syntenic region containing *VPS4B/b* is only 2 cM long and contains just ten identified genes (not shown).

3.3. Analysis of other vertebrate VPS4 ORFs

(I) There are exactly two *VPS4* paralogues in mammals: As searches in EST-databases, as well as in human and mouse genome databases reveal, all available mammalian *VPS4* sequences can be assigned to either of the two paralogues. (II) Codon usage differs remarkably in both paralogue pairs. In both organisms, there is a strong bias: *VPS4A/a* clearly prefers codons ending in G or C (human 73%; mouse 55% NNG/C-codons), while in *VPS4B/b* codons more frequently end in A or T (human 34%; mouse 41% NNG/C-codons). In *VPS4A/a*, that bias is even stronger in the ORFs than in the non-coding portions of the genes, in *VPS4B/b* it is roughly the same (see Section 3.4.IV). (III) A protein sequence alignment was employed to trace the phylogeny of the paralogues (Fig. 2). For calculation of relationships, positions typical of the mammalian a- versus b-paralogue were identified (coloured in Fig. 2A). Following the logic of maximum parsimony for deduction of phylogenetic history, the respective positions in the other VPS4 representatives were assigned to either 'a', 'b', or were excluded (Fig. 2B). The chicken protein can be clearly assigned to the VPS4a branch, while the frog protein resembles much more the VPS4b representatives. The Fugu fish genome, surprisingly, contains at least three *VPS4* paralogues one of which (*VPS4b*) is differentially spliced in its N-terminal part. The differences between the genes are high enough to exclude contig assembly errors. The identification of all three ORFs was unambiguous, only a single internal exon and in one case the very first exon were not found due to their shortness and poor sequence conservation ('xx' in alignment). Two of the Fugu proteins resemble the two paralogues VPS4a and -b, respectively, and are tentatively denoted as such. The third one, called VPS4c, seems to be somehow 'intermediate'. Though it branches together with the other VPS4b proteins (Fig. 2C), its assignment to them is not very convincing, since the relation of VPS4a-typical to b-typical positions is far less than for all the other vertebrate proteins (Fig. 2B). Plant, yeast, and worm are roughly equally distantly related to both mammalian paralogues, indicating that the VPS4a/VPS4b split is restricted to vertebrates or chordates rather than being primordial to animals or eukaryotes.

3.4. Alignment of the genomic sequences

(I) Sequence similarity between mouse and human is strongest within the coding regions (both pairs 89%), and less in the 3′ and 5′ non-coding regions (*VPS4A/a*: 64%; *VPS4B/b*: 54%), while significant homology is observed only in less than half of intronic sequence portions. Detectable similarity between *VPS4A/a* and *VPS4B/b* even is restricted to the ORF and small parts of the neighbouring introns just adjacent to the intron–exon boundaries. (II) The B-paralogues are significantly longer than the A-paralogues (human by 2.4-fold; mouse by 1.8-fold). This obviously is due to more frequent insertions of interspersed repeats into the introns of the B-paralogues, which now comprise a greater portion of these genes (Table 1 and Fig. 3). And indeed, there is only a single case of a *VPS4A/a* intron being longer than the homologous *-B/b* introns (namely intron 6 in mouse *VPS4a*). (III) The majority of those insertions seems to have happened after the mammalian radiation. Only in few cases (four in *VPS4A/a* and five in *VPS4B/b*), the repetitive elements obviously have been present in the ancestor of rodents and primates (arrows in Fig. 3); the remainder has been inserted after their divergence. (IV) GC-content differs: *VPS4A/a* is, compared to the average, slightly GC-rich (54 and 48% GC, respectively), while *VPS4B/b* is AT-rich (60% AT both). (V) Promoter analyses did not reveal any clear incidents of TATA- or CAAT-boxes. Moreover, in comparison of both orthologue pairs similarities upstream of exon 1 hardly can be detected. Search for palindromes and employment of common gene-finding programs available via internet yielded only poor results, too. Searches in databases with transcription factor binding sites also gave no high-scoring hits. The only obvious feature is the existence of CpG islands of typical length (Table 2).

Comparison of the four mammalian genes with the three Fugu representatives was performed only quantitatively. Since the repeats typical of our own genome have been inserted long after fish-tetrapod divergence (Lander et al., 2001), such elements are absent in Fugu. All three Fugu genes are of comparable length and much shorter than the

Fig. 2. Comparison of VPS4 proteins. (A) Protein sequence alignment of all available VPS4 representatives. Intron positions are marked above alignment, asterisks denote those conserved between animals and *Arabidopsis thaliana*. Identified regions (domains) are indicated at the bottom line. Positions in which the mammalian paralogues differ are green (typical of VPS4a) vs. blue (VPS4b). (B) All VPS4 representatives were counted out for positions resembling either mammalian VPS4a or -b. In cases where only one of both paralogues was found in the databases, its most likely (yet tentative) assignment is given in brackets. (C) A neighbour-joining dendrogram of the proteins shown in (A). The length of the bar corresponds to 10% sequence divergence. Figures denote bootstrap probabilities (5000 trials); branches without figures have bootstrap values of 1.0.

Table 1
Distribution of repeats in the four *VPS4* genes

	Human VPS4A (13.64 kb[a], 54% GC)			Human VPS4B (33.27 kb[a], 40% GC)		
	No. of elements	Length occupied (bp)	Percentage of sequence	No. of elements	Length occupied (bp)	Percentage of sequence
SINEs:	19	4572	26.4	31	7572	20.1
ALUs	**16**	**4086**	**23.6**	**27**	**6964**	**18.5**
MIRs	**3**	**486**	**2.8**	**4**	**608**	**1.6**
LINEs:	4	704	4.1	11	3713	9.9
LINE1	**1**	**269**	**1.6**	**5**	**1578**	**4.2**
LINE2	**3**	**411**	**2.4**	**6**	**2135**	**5.7**
LTR elements:	1	352	2.0	1	156	0.4
MaLRs	**1**	**352**	**2.0**			
Retroviral				**1**	**156**	**0.4**
DNA elements:	1	249	1.4	5	654	1.7
MER1_type	**1**	**249**	**1.4**	**3**	**409**	**1.1**
MER2_type				**1**	**160**	**0.4**
Simple repeats	2	137	0.8	3	134	0.4
Low complexity	1	25	0.2	1	21	0.1
Total		6042	34.9		12250	32.5

	Mouse VPS4a (14.43 kb[a], 48% GC)			Mouse VPS4b (25.58 kb[a], 40% GC)		
	No. of elements	Length occupied (bp)	Percentage of sequence	No. of elements	Length occupied (bp)	Percentage of sequence
SINEs:	26	3488	20.8	14	2243	7.6
B1s	**11**	**1144**	**6.8**	**4**	**468**	**1.6**
B2–B4	**13**	**2052**	**12.2**	**9**	**1640**	**5.6**
IDs	**1**	**72**	**0.4**			
MIRs	**1**	**220**	**1.3**	**1**	**135**	**0.5**
LINEs:	1	487	2.9	6	2081	7.1
LINE1	**1**	**487**	**2.9**	**5**	**1972**	**6.7**
LINE2				**1**	**109**	**0.4**
LTR elements:	0	0	0.0	6	2186	7.5
MaLRs				**5**	**1770**	**6.0**
Retroviral				**1**	**416**	**1.4**
DNA elements:	0	0	0.0	1	104	0.4
MER1_type				**1**	**104**	**0.4**
MER2_type						
Simple repeats	6	249	1.1	9	421	1.4
Low complexity	1	21	0.2	4	182	0.6
Total		4237	25.0		7217	24.5

Absolute values and percentage are given, main families are bold. Mariners, Mer4, small RNAs and satellites were not detected.
[a] Sequenced regions 5′ of first and 3′ of last exon were not taken into account.

mammalian representatives: *VPS4a*, 3.27 kb; *VPS4b* (differentially spliced, without Exon 1), 4.24 kb; *VPS4c*, 3.21 kb. Obviously a bias in genomic dynamics like observed for mammalian *Vps4* representatives does not exist in Fugu. GC-content also does not vary, nor the codon usage: all three ORFs use NNG/C-codons exactly twice as often as NNT/A-codons. There is no pyrimidine to purine bias.

3.5. Conservation of intron positions

Introns at conserved places must be primordial and hence can provide information on the genes' phylogenetic histories as well as clues on their domain structure. The position of introns in relation to the domain structure of *VPS4* has been discussed previously (Scheuring et al., 1999). The availability of more genomic sequences now allows to compare different species: In Fig. 2A, introns conserved between *Arabidopsis thaliana* and animals (*C. elegans*/vertebrates) are marked; however, only few of them really are. Because of the availability of only one non-animal sequence it unfortunately is not possible to distinguish recent insertion from recent loss. For an exhaustive phylogenetic analysis of intron–exon-structure, one obviously has to wait for more genomes to be finished.

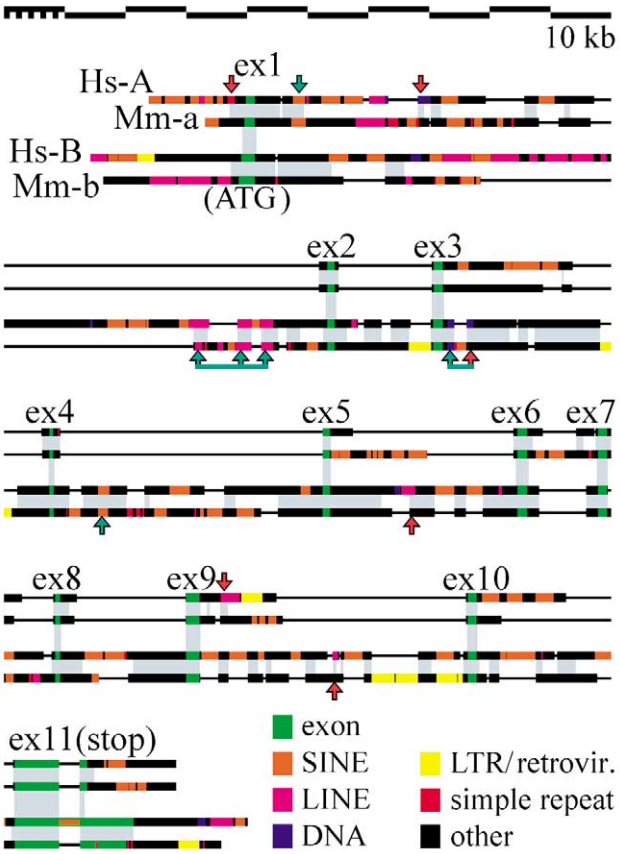


Fig. 3. Schematic alignment of both *VPS4* paralogues in human and mouse. Colours denote different classes of sequence: SINE, small interspersed repeats; LINE, long interspersed repeats; DNA, DNA-elements (transposons); LTR, long terminal repeats. Detectable homology is indicated by grey bars connecting the sequences. Hs-A/-B, Mm-a/-b, human and mouse VPS4A/a and -B/b, respectively. Arrows denote homologous repeats at corresponding places; red arrows: one of them has not been recognized by the repeat masker program.

3.6. Gene expression analyses

To find out if the expression rates of the two murine *VPS4* paralogues in different cell types are at comparable level and moreover, if their expression is interrelated, we hybridized a mouse multi-tissue Northern blot. As probes we employed digoxigenin-labelled PCR fragments containing the complete ORFs, actin served as control (Fig. 4). Cross-hybridization of the probes to the paralogue mRNAs was not observed under the stringent hybridization and washing conditions applied. We noticed that expression rates differed remarkably and moreover, that most tissues showed a strong bias for one of either *VPS4* representatives. The abundance ranged from 'high' (liver, testis), to 'almost absent' (skeletal muscle, spleen). The ratio reached from 'mainly *VPS4a*' (testis) over 'balanced' (heart) to 'mainly *VPS4b*' (liver). As judged by semi-quantitative RT–PCR of endogenous mouse *VPS4a* and *-b* in NIH/3T3 cells, *VPS4b* is expressed slightly higher than *VPS4a* (data not shown). A comparable expression ratio was also observed in liver and kidney in Northern blot analysis (Fig. 4).

Table 2
CpG-islands in VPS4 genes

	CpG-islands: CpG/100 bp			
	(kb 5′)	**(-bp...ATG)**	**(ATG...bp)**	(kb 3′)
Hs VPS4A	3.1 (1.3)	**12.6 (413)**	**11.1 (470)**	2.6 (15.1)
Mm VPS4a	0.2 (0.3)	**9.8 (471)**	**11.8 (550)**	1.1 (15.5)
Hs VPS4B	1.5 (1.9)	**8.3 (775)**	**8.3 (230)**	1.0 (34.7)
Mm VPS4b	0.8 (2.1)	**9.6 (334)**	**11.0 (205)**	1.2 (26.7)

Data indicate CpG incidences/100 nt. Numbers in parentheses denote the length of sequence measured. The two columns in bold refer to the CpG island itself 3′ and 5′ of the ATG start codon, respectively. The far left and right columns refer to the remainder of the sequence on either side of the CpG-island.

To date, no expression studies in mammalian cells have been performed considering both paralogues. Hence, we decided to investigate both VPS4 proteins in a 'homologue system', that is, WT and dominant-negative protein versions of murine *VPS4a* and *VPS4b* in mouse 3T3 cells. For that purpose we constructed plasmids expressing fluorescent GFP or dsRed fusion proteins of WT or dominant negative mutant VPS4 (see Section 2.8). Subcellular distribution analyses were performed with 3T3 cells, transfected transiently with these fusion constructs (Fig. 5A). To further examine a possible co-localization of the paralogues, we analyzed the simultaneous expression of these constructs in co-transfected cells (Fig. 5B). Both dsRed-tagged WT VPS4 proteins, but especially VPS4a, were distributed in a more or less fine punctate pattern near the nucleus and over the cytosol (Fig. 5A, a,b,d,e). GFP-tagged WT VPS4a and -b displayed a very similar pattern, except that they appeared more evenly distributed (Fig. 5B, b,f,j). Interestingly, the perinuclear aggregation was more prominent in cells expressing dominant-negative mutant VPS4a/b fusion proteins (Fig. 5A, g,h,j,k). In mouse fibroblasts co-transfected with a dominant-negative and a WT VPS4

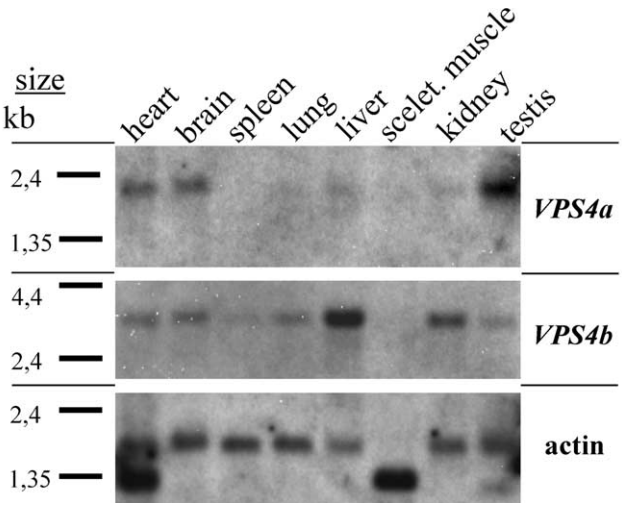


Fig. 4. Multi-tissue northern-blot of mouse *VPS4a* and *VPS4b*. Eight different tissues were tested (head line), digoxigenin-labelled complete cDNAs were employed as probes, actin served as control (lower panel).

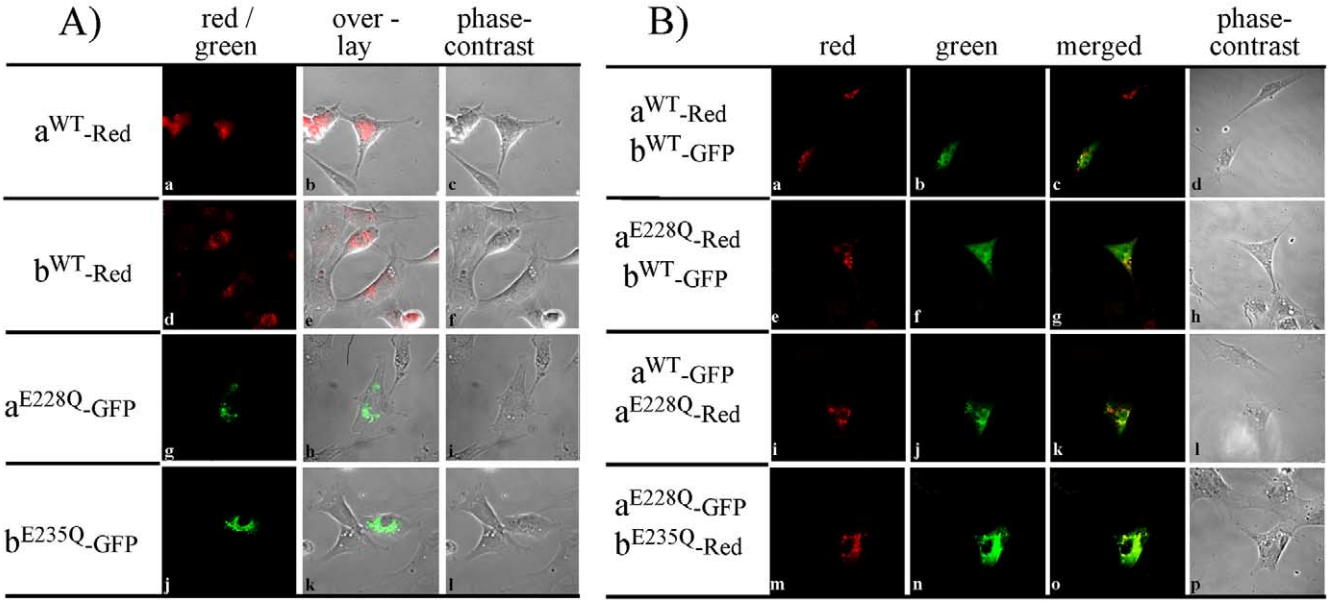


Fig. 5. Mouse 3T3 cells transiently transfected with mammalian expression vectors encoding dsRed or GFP tagged VPS4a/b or VPS4a^{E228Q}/VPS4b^{E235Q} proteins. Subcellular distribution of single (A) or co-transfected (B) mouse fibroblasts were analyzed by confocal microscopy for red or green fluorescence. Merged images are shown for co-localization. Expression constructs used (abbreviated on left side of panels) were pVPS4a-Red (A, a–c; B, a–d); pVPS4b-Red (A, d–f); pVPS4a-GFP (B, i–l); pVPS4b-GFP (B, a–h); pVPS4a^{E228Q}-Red (B, e–l); pVPS4b^{E235Q}-Red (B, m–p); pVPS4a^{E228Q}-GFP (A, g–i; B, m–p); pVPS4b^{E235Q}-GFP (A, j–l) (details of plasmid construction are given in Section 2.8).

fusion construct (Fig. 5B, e–l), the mutant protein version still exhibited this punctate, perinuclear concentration while the distribution of the WT protein remained more or less unaffected. Only if both mutant VPS4 proteins were expressed simultaneously a more pronounced co-localization could be observed (Fig. 5B, m–p).

Previously, we have shown that human VPS4A/B proteins interact with each other and form a heteromeric complex in two-hybrid analyses (Scheuring et al., 2001). Therefore, we wondered if the observed co-localization of mutant murine VPS4a/b tagged proteins in mammalian cells could also be supported by two-hybrid analyses. Indeed, we observed a weak heteromeric interaction between WT VPS4a and -b proteins. Similar to human VPS4A/B proteins, the interaction was much stronger if dominant-negative murine mutant VPS4a/b protein variants were involved (data not shown).

4. Discussion

4.1. Chromosomal localization of both VPS4 genes in human and mouse

Chromosomal localization of the four mammalian *VPS4* genes was achieved by two independent methods (Fig. 1A, B). The locations agree well with the data available concerning human-mouse synteny (http://www.ncbi.nlm.nih.gov/Homology/). But most interestingly, human *VPS4A* got two BLAST hits with $>$ 99.9% sequence identity in close vicinity to each other (100 kb) on chromosome 16 (Fig. 1B). Nevertheless, southern blot analysis does not support two loci for *VPS4A* (data not shown). Since in this region of the human draft sequence there are non-overlapping contigs, and moreover because the two hits of *VPS4A* in the draft begin with very similar but non-identical repeats that puzzling BLAST-result most likely is due to an assembly error in the draft genome sequence.

4.2. Alignment of the entire VPS4 protein family

All available VPS4 representatives were compared. As expected, distances of the *A. thaliana*, yeast, *C. elegans*, and *Drosophila melanogaster* proteins to the paralogue pairs are roughly equal. But surprisingly, the patterns typical of the a/b paralogues are found in the frog and chicken proteins, too. The concordance is much too high to occur by chance (Fig. 2A–C). In particular, the deletions highly specific for VPS4a (second line of alignment, first to fifth sequence) allow definite assignment of the chicken protein as a-paralogue. Yet, only ESTs from a single *VPS4* gene could be identified in *Bos taurus* (not shown in Fig. 2A), *X. laevis*, and chicken. This is, however, not too unexpected given the high bias in expression levels – e.g. from *X. laevis*, mainly oocyte ESTs are deposited in the gene bank (compare also to Fig. 4).

Apart from this, the existence of three VPS4 proteins in Fugu fish is puzzling. Data quality is too high (only a single frameshift error had to be corrected manually in the three ORFs), and moreover the differences between them fit too well into the alignment to assume artificial duplications caused by contig assembly errors. Two Fugu paralogues can

be assigned to the VPS4a and -b branches, respectively (Fig. 2A–C). VPS4c is more similar to the b-branch, but its assignment is ambiguous (Fig. 2B,C). It clearly represents an earlier split-off, but the question remains if this had occurred before or after the separation of the two main VPS4 branches. Moreover, all of these three paralogues seem to be functional. Fugu VPS4a and VPS4b are supported by cognate ESTs from different fish (Fugu, Danio, Salmon: 11 ESTs in total). No EST from a fish could be assigned to VPS4c, but several reasons make it very unlikely that it is a pseudogene: (I) no frameshift is present in the ORF, (II) splice sites are conserved, (III) codon usage is conserved, (IV) the majority of substitutions is conservative, (V) as visualized in distance matrix/phylogenetic tree analysis, the gene has not experienced an evolutionary speed-up like pseudogenes usually do, and (VI) as the distance matrix shows, the separation of VPS4c must have happened a long time ago. Nevertheless, all the features listed still are present. This can only be due to selective pressure on a functional gene.

Concerning the origin of three Fugu paralogues, on the one hand there are hints to two subsequent genome duplications, or at least some series of partial genome duplications initiating vertebrate evolution (Gu et al., 2002). On the other hand, evidence has been provided that large scale gene duplications have occurred more recently in teleost fish (Robinson-Rechavi et al., 2001). Hence, the presence of three *VPS4* genes, roughly equally distantly related to each other, can be explained by two different scenarios: Either an ancient *VPS4* double duplication has occurred early in vertebrate radiation, and subsequently one paralogue was lost in Fugu evolution and two in the history of tetrapods, or *VPS4* has originally been duplicated only once, and a second time in the evolution of fish. It would be interesting to find out if two or even more *VPS4* genes are present in chondrocytes, agnatha, acrania, and tunicata.

4.3. Comparison of the genomic VPS4 sequences from mammals and Fugu fish

On the one hand, distances of both paralogues to other *VPS4* representatives are roughly equal (Fig. 2B,C). Therefore, no speed-up in evolution has occurred for one of them, which means that functional constraints most likely have not changed and none of both has experienced gain of novel function(s). On the other hand, there are remarkable differences between the paralogues. (I) the obvious difference in length. There are two possible explanations for this finding. One could hypothesize that functional constraints may cause a gene to shrink, e.g. when transcription has to be optimized. However, if there is any general difference in expression rates, as judged by EST-report and Northern blot analyses (Fig. 4), the longer VPS4b-paralogue is expressed to a higher level. Moreover, to our knowledge there is no report in the literature giving evidence that differences in length, like observed here, may be critical for expression. Another reason might be a structural one. Maybe the chromosomal loci have differing susceptibilities to insertion of repeats. Indeed, Venter and coworkers found such evidence for duplicated regions in the genome (Venter et al., 2001), so that this is the most likely explanation. In line with this hypothesis of a different chromosomal environment is the differing GC-content, which is conserved between human and mouse to a certain degree. (II) There is a clear bias in codon usage. In *VPS4A/a*, NNG/C-codons are more abundant, while the *B/b*-paralogues prefer NNA/T-codons. This bias is more pronounced in human, which is in line with the differing GC-content of the genes; even the stronger difference seen in the two human genes is reflected in the codon bias. *VPS4B/b* possesses the more regular codon usage, i.e. it is moderately A/T-rich. Usually, higher expressed genes exhibit such a ‘canonic’ codon usage. In the case of VPS4, however, *VPS4A/a* in general does not appear to be lowly expressed, so one would assume that the codon usage bias is rather ‘structurally caused’ than somehow functionally important.

It is interesting to note that some of the repeats inserted before the human-mouse-divergence have not been identified by the repeat masker program because one copy is too divergent (red arrows in Fig. 3). This supports the notion of the Human Genome Organization (Lander et al., 2001) that the percentage of our genome covered by interspersed repeats they determined as ≈ 50% is an underestimate.

4.4. Comparison of VPS4a : VPS4b expression

Northern blot analyses and semi-quantitative RT–PCR demonstrated a clear expression bias in different mouse tissues (see Section 3.6 and Fig. 4). Both, the total abundance and the *VPS4a* : *VPS4b* ratio vary very much. This is further substantiated by the fact that the occurrence of *VPS4* ESTs of all mammals, chicken, and frog in the databases is subjected to a strong bias, too (i.e. for frog, chicken, and cow, ESTs of one paralogue only were found in the databases, see Fig. 2A).

Expression of WT and dominant-negative (E235Q) mouse SKD1/VPS4b proteins was investigated by Yoshimori et al. (2000). The subcellular distribution analyses they performed, concentrated on the VPS4b paralogue and therefore, a direct comparison with the VPS4a paralogue was not possible. They reported murine VPS4b to be located diffusely in the cytoplasm. In their analysis, WT GFP-VPS4b was found in the cyto- and nucleoplasm, whereas the mutant GFP-VPS4b^{E235Q} showed a punctate, perinuclear pattern. This finding is consistent with yeast studies, where different groups have shown that WT yeast Vps4p and the WT human VPS4B protein are evenly distributed in the yeast cytosol (Babst et al., 1998; Scheuring et al., 2001). Dominant-negative mutant yeast Vps4p^{E233Q} and human VPS4B^{E235Q} protein, however, exhibit a punctate staining. The regions where the mutant VPS4 proteins concentrate

co-localize with endosomal membranes, which was shown in yeast with FM4-64 staining (Scheuring et al., 2001) and in HEK293 cells by co-localization with the early endosome antigen EEA1 (Yoshimori et al., 2000).

Comparing the subcellular localization of GFP-/dsRed-tagged WT as well as dominant-negative mutant protein versions of VPS4a and -b, we found both WT VPS4a and VPS4b paralogues distributed in the cytosol. A fraction of cells, however, exhibited a slightly punctate pattern. This tendency was clearly stronger for VPS4a fusion proteins, while for all dominant-negative mutant protein versions such punctate distribution pattern really was characteristic. WT and mutant mouse VPS4a and -b fusion proteins partially co-localize in transfected 3T3 cells (see Section 3.6). Moreover, when both dominant negative alleles are present, co-localization even is stronger, providing evidence for a VPS4a–b interaction in vivo. Indeed, a direct physical interaction of both mouse paralogues could be demonstrated by the two-hybrid analyses we performed. These results are in line with some other findings: (I) the yeast Vps4 protein forms multimers (Babst et al., 1998), and (II) human VPS4A and VPS4B proteins interact in a two-hybrid assay (Scheuring et al., 2001).

In our present study, we addressed the question if both paralogues display a similar or even identical distribution. Indeed, one must be conscious of the fact that tagged proteins may behave different than the untagged proteins and moreover, that expression levels may be altered. Antibodies for in-situ detection would be an alternative, but all efforts to raise paralogue-specific antibodies for the detection of endogenous VPS4a/b have failed so far. However, to minimize artefacts, we generated different fusion constructs (dsRed; GFP, C- and N-terminal fusions) and performed control experiments with the mere GFP and dsRed proteins (data not shown).

4.5. Why do we, in contrast to yeast, worm, and fly, possess two VPS4 genes?

Obviously, the paralogue *VPS4* pair is conserved in all mammals, most likely even over the majority of vertebrates. On the one hand, slight but clear and conserved differences between *VPS4a* and *VPS4b* are present. Moreover, the fact that expression in different tissues is strongly biased argue for functional diversification of both proteins. On the other hand, only few (roughly two dozen) positions in the proteins' sequences are distinct between the paralogues, and the majority of them even is conservatively substituted. This high degree of similarity between VPS4a and VPS4b makes it very probable that they perform the same molecular function and makes it really hard to argue for a difference in their molecular reaction mechanisms. Besides this, co-localization of the proteins is obvious, though not perfect.

So in all likelihood they form part of the same protein complex. The differing levels of expression, however, make it quite unlikely that there is a defined VPS4a:b stoichiometry. Either, the protein complexes display varying VPS4a:b contents, or there are two sorts of them: a VPS4a and a VPS4b-paralogue containing one. The latter possibility, however, is not convincing given the interactions seen in two-hybrid assays and the determined partial co-localization of fluorescence-labelled proteins. If such mixed complexes exist, the question remains to be elucidated how they differ and how such a functional difference is achieved by the two paralogues. There is only one characteristic difference between them, which may give a hint: The highly specific deletion in VPS4a in the linker region between the coiled-coil domain and the AAA-cassette. Evidence has been provided that VPS4 recruits target proteins via its coiled-coil domain (Babst et al., 1998). Now it is tempting to speculate that the AAA-domain acts on them, modulating protein-protein interactions. The difference in linker length between 'anchor' (coiled-coil) and 'actor' (AAA-cassette) may affect the spatial orientation of the AAA-domain relative to the target protein(s), which may have an influence on the molecular processes. Given the high degree of complexity of the process of MVB formation one can well imagine that such a slight difference may make sense. The in vivo prove of heteromeric/homomeric VPS4 complex formation, however, has to await specific probes for the endogenous proteins.

After all, the existence of (at least) three *VPS4* genes in Fugu fish plus (at least) one splicing variant is surprising. In general, lower vertebrates are regarded more primitive than, e.g., mammals with respect to their physiology as well as their morphology. Given the restricted current knowledge on Vps4's molecular function, one can hardly imagine why Fugu or maybe fish in general need more paralogues than higher vertebrates. It can be expected that phylogenetic and biochemical investigations on VPS4 are going to yield more interesting and unexpected results.

Acknowledgements

We thank the RZPD for their screening service and for supplying us with the murine cDNA library. The mouse cDNA pool was a gift from Jens Hirchenhain (Frauenklinik, Heinrich-Heine-Universität, Düsseldorf). This work was supported by a grant from the research committee of the Heinrich-Heine-Universität Düsseldorf, the Deutsche Forschungsgemeinschaft and the SFB 575 ('Experimentelle Hepatologie').

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G.,

Smith, J.A., Struhl, K., 1990. Current Protocols in Molecular Biology, Green Publishing Associates/Wiley Interscience, New York.

Babst, M., Sato, T.K., Banta, L.M., Emr, S.D., 1997. Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p. EMBO J. 16, 1820–1831.

Babst, M., Wendland, B., Estepa, E.J., Emr, S.D., 1998. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. EMBO J. 17, 2982–2993.

Babst, M., Odorizzi, G., Estepa, E.J., Emr, S.D., 2000. Mammalian tumor susceptibility gene 101 (TSG101) and the yeast homologue, Vps23p, both function in late endosomal trafficking. Traffic 1, 248–258.

Bairoch, A., 1991. PROSITE: a dictionary of protein sites and patterns. Nucleic Acids Res 19 Suppl., 2241–2245.

Banta, L.M., Robinson, J.S., Klionsky, D.J., Emr, S.D., 1988. Organelle assembly in yeast: characterization of yeast mutants defective in vacuolar biogenesis and protein sorting. J. Cell Biol. 107, 1369–1383.

Becker, D.M., Guarente, L., 1991. High-efficiency transformation of yeast by electroporation. Methods Enzymol. 194, 182–187.

Beyer, A., 1997. Sequence analysis of the AAA protein family. Protein Sci. 6, 2043–2058.

Bishop, N., Woodman, P., 2000. ATPase-defective mammalian VPS4 localizes to aberrant endosomes and impairs cholesterol trafficking. Mol. Biol. Cell 11, 227–239.

Conibear, E., Stevens, T.H., 1998. Multiple sorting pathways between the late Golgi and the vacuole in yeast. Biochim. Biophys. Acta 1404, 211–230.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., Varshavsky, A., 1995. An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. J. Biol. Chem. 270, 18099–18109.

Felder, S., Miller, K., Moehren, G., Ullrich, A., Schlessinger, J., Hopkins, C.R., 1990. Kinase activity controls the sorting of the epidermal growth factor receptor within the multivesicular body. Cell 61, 623–634.

Fernandez-Borja, M., Wubbolts, R., Calafat, J., Janssen, H., Divecha, N., Dusseljee, S., Neefjes, J., 1999. Multivesicular body morphogenesis requires phosphatidyl-inositol 3-kinase activity. Curr. Biol. 9, 55–68.

Finken-Eigen, M., Röhricht, R.A., Köhrer, K., 1997. The VPS4 gene is involved in protein transport out of a yeast pre-vacuolar endosome-like compartment. Curr. Genet. 31, 469–480.

Fuchs, R., 1991. MacPattern: protein pattern searching on the Apple Macintosh. Comput. Appl. Biosci. 7, 105–106.

Futter, C.E., Pearse, A., Hewlett, L.J., Hopkins, C.R., 1996. Multivesicular endosomes containing internalized EGF–EGF receptor complexes mature and then fuse directly with lysosomes. J. Cell Biol. 132, 1011–1023.

Futter, C.E., Collinson, L.M., Backer, J.M., Hopkins, C.R., 2001. Human VPS34 is required for internal vesicle formation within multivesicular endosomes. J. Cell Biol. 155, 1251–1264.

Gary, J.D., Wurmser, A.E., Bonangelino, C.J., Weisman, L.S., Emr, S.D., 1998. Fab1p is essential for PtdIns(3)P 5-kinase activity and the maintenance of vacuolar size and membrane homeostasis. J. Cell Biol. 143, 65–79.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Gruenberg, J., Maxfield, F.R., 1995. Membrane transport in the endocytic pathway. Curr. Opin. Cell Biol. 7, 552–563.

Gu, X., Wang, Y., Gu, J., 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat. Genet. 31, 205–209.

Henikoff, S., Henikoff, J.G., 1994. Protein family classification based on searching a database of blocks. Genomics 19, 97–107.

Jainchill, J.L., Aaronson, S.A., Todaro, G.J., 1969. Murine sarcoma and leukemia viruses: assay using clonal lines of contact-inhibited mouse cells. J. Virol. 4, 549–553.

Katzmann, D.J., Babst, M., Emr, S.D., 2001. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. Cell 106, 145–155.

Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Lemmon, S.K., Traub, L.M., 2000. Sorting in the endosomal system in yeast and animal cells. Curr. Opin. Cell Biol. 12, 457–466.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D., Ward, D.C., 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science 247, 64–69.

Miller, J., 1972. Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nothwehr, S.F., Bryant, N.J., Stevens, T.H., 1996. The newly identified yeast GRD genes are required for retention of late-Golgi membrane proteins. Mol. Cell Biol. 16, 2700–2707.

Odorizzi, G., Babst, M., Emr, S.D., 1998. Fab1p PtdIns(3)P 5-kinase function essential for protein sorting in the multivesicular body. Cell 95, 847–858.

Périer, F., Coulter, K.L., Liang, H., Radeke, C.M., Gaber, R.F., Vandenberg, C.A., 1994. Identification of a novel mammalian member of the NSF/CDC48p/Pas1p/TBP-1 family through heterologous expression in yeast. FEBS Lett. 351, 286–290.

Raymond, C.K., Howald-Stevenson, I., Vater, C.A., Stevens, T.H., 1992. Morphological classification of the yeast vacuolar protein sorting mutants: evidence for a prevacuolar compartment in class E vps mutants. Mol. Biol. Cell 3, 1389–1402.

Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., Laudet, V., 2001. Euteleost fish genomes are characterized by expansion of gene families. Genome Res. 11, 781–788.

Rowe, L.B., Barter, M.E., Eppig, J.T., 2000. Cross-referencing radiation hybrid data to the recombination map: lessons from mouse chromosome 18. Genomics 69, 27–36.

Sawyer, J.R., Moore, M.M., Hozier, J.C., 1987. High resolution G-banded chromosomes of the mouse. Chromosoma 95, 350–358.

Scheuring, S., Bodor, O., Röhricht, R.A., Müller, S., Beyer, A., Köhrer, K., 1999. Cloning, characterisation, and functional expression of the Mus musculus SKD1 gene in yeast demonstrates that the mouse SKD1 and the yeast VPS4 genes are orthologues and involved in intracellular protein trafficking. Gene 234, 149–159.

Scheuring, S., Röhricht, R.A., Schöning-Burkhardt, B., Beyer, A., Müller, S., Abts, H.F., Köhrer, K., 2001. Mammalian cells express two VPS4 proteins both of which are involved in intracellular protein trafficking. J. Mol. Biol. 312, 469–480.

Tanaka, H., Fujita, H., Katoh, H., Mori, K., Negishi, M., 2002. Vps4-A is a binding partner for a novel Rho family GTPase, Rnd2. Biochem. J. 365, 349–353.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 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.

Venter, J.C., et al., 2001. The sequence of the human genome. Science 291, 1304–1351.

Yoshimori, T., Yamagata, F., Yamamoto, A., Mizushima, N., Kabeya, Y., Nara, A., Miwako, I., Ohashi, M., Ohsumi, M., Ohsumi, Y., 2000. The mouse SKD1, a homologue of yeast Vps4p, is required for normal endosomal trafficking and morphology in mammalian cells. Mol. Biol. Cell 11, 747–763.