

Additional 5' Exons in the *RGS3* Locus Generate Multiple mRNA Transcripts, One of Which Accounts for the Origin of Human PDZ-RGS3

John H. Kehrl,* Deepa Srikumar, Kathleen Harrison, Gaye L. Wilson, and Chong-Shan Shi

Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

*To whom correspondence and reprint requests should be addressed. Fax: (301) 402-0070. E-mail: jkehrl@niaid.nih.gov.

Regulators of G-protein signaling (RGS) proteins can be broadly divided into those that consist predominantly of an RGS domain and those that possess an RGS domain along with additional domains. *RGS3* fits into both categories, as both short and longer forms exist. Recently, a novel form of mouse *RGS3* that possesses a PDZ domain was identified. Here we show that the human PDZ-RGS3 isoform arises from 10 upstream exons along with 6 exons from the previously characterized *RGS3*. We found that 47,000 nucleotides span the last of the 10 upstream exons and the first exon used from the original cluster of *RGS3* exons. These 10 upstream exons encode 398 amino acids, which show strong conservation with those from mouse PDZ-RGS3. In addition, another isoform exists that uses 17 upstream exons, 9 of which overlap with those in PDZ-RGS3, along with the same 6 downstream exons used in PDZ-RGS3. Finally, a short form of human *RGS3* arises from an unrecognized *RGS3* exon that encodes an amino-terminal 140 amino acids. For each *RGS3* isoform, RT-PCR detected specific mRNA transcripts and immunoblot analysis identified specific bands for *RGS3* and PDZ-RGS3. *RGS3* provides an example of the complex origins of the coding regions of mammalian proteins.

Key Words: RGS, RNA splicing, G-protein, genomic

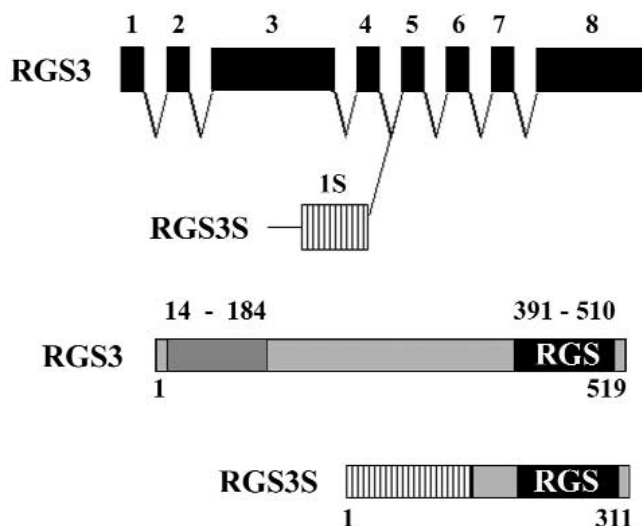
INTRODUCTION

G-protein coupled receptors (GPCRs) use heterotrimeric G-proteins to signal to downstream effectors. Receptor activation promotes exchange of GTP for GDP by the G_{α} subunit and the release of their associated $\beta\gamma$ subunits. Both GTP-bound G_{α} subunits and free $G_{\beta\gamma}$ subunits can bind and activate downstream effectors. However, G_{α} subunits possess an intrinsic GTPase activity, which returns G_{α} to its GDP-bound state and thereby limits the duration of G_{α} signaling [reviewed in 1,2]. Cells possess an additional mechanism that reduces the duration that a G_{α} subunit remains GTP bound. Members of a family of proteins termed regulators of G-protein signaling (RGS) dramatically accelerate the intrinsic rate that certain G_{α} subunits hydrolyze GTP, a property that identifies them as GTPase activating proteins (GAPs) [reviewed in 3,4].

Over 20 human genes that encode proteins with RGS domains have been identified, along with their corresponding mouse orthologs. One of them, *RGS3*, encodes a 519-amino-acid protein with a carboxy-terminal RGS domain, an unusual

amino-terminal region that shares some similarity with a domain in synapsin [5]. Northern blot analysis revealed prominent 1.8- and 3.5-kb *RGS3* mRNA transcripts, but also two less-prominent transcripts of 4.2 and 4.5 kb [5]. The 1.8- and 3.5-kb mRNA transcripts presumably encoded short and long versions of *RGS3*, but the larger transcripts had no protein-coding region assigned to them. An expression vector for a short form of *RGS3* (*RGS3T* or *RGS3CT*) was created by arbitrarily choosing an internal ATG [5-7]. Both forms of *RGS3* act as GAPs for $G_{i\alpha}$ and $G_{q\alpha}$ but not $G_{s\alpha}$ or $G_{12\alpha}$ [8]. Expression of either form of *RGS3* inhibits signaling through a variety of GPCRs that use $G_{i\alpha}$ and $G_{q\alpha}$ to transduce signals [5-9]. The shorter version of *RGS3* in part localizes in the nucleus and its expression promotes apoptosis of Chinese hamster ovary (CHO) cells [10]. The introduction of *RGS3* into B-cell lines inhibited CXCR1- and CXCR4-directed cell migration [11,12]. Perhaps accounting for the efficacy of *RGS3* as an inhibitor of chemotaxis, it, in contrast to several other RGS proteins, also inhibits signaling by free $G_{\beta1\gamma2}$ subunits [13]. Interference with $G_{\beta\gamma}$ signaling blocks chemokine-directed migration [14,15].

FIG. 1. Origin of human RGS3S and its N-terminal amino acid sequence. (A) The genomic organization of the original RGS3. The mRNA transcript that encodes human RGS3 arises from two untranslated exons and six coding exons. The N-terminal 140 amino acids of human RGS3S derive from an alternative exon (exon 1S) present in intron 4 and exons 5–8 of the original RGS3. (B) Schematic of the human RGS3 and RGS3S proteins are shown. RGS3S lacks the synapsin homology domain present in RGS3 (amino acids 41–184). The RGS domain is shown in black. (C) Amino acid sequence of the N terminus of human RGS3S. Of note a putative Akt phosphorylation site in the N terminus of human RGS3S (underlined) is conserved with a similar site in RGS3. The portion of the RGS3S amino acid sequence derived from chromosome 9 nucleotide sequence is boxed.



**MVTRRPVTNSWDWLPAGAAPEAVPCRHMPLSRLPLRVGQKEFFFLP
LLVPPISWLLLSSEQPRLVPGSPVIRPGFQRACVAAACTVAARCPGRGV
GDRSQSGASRRPIGGPKVGGPTEMLRGMYLTRNGNLQRRHTMKE 140**

RESULTS

A partial cDNA for the mouse long form of RGS3 and for a short version of RGS3 with a unique N-terminal 21-amino-acid extension [12] have been identified, as well as a cDNA for the purported full-length version of the long form of mouse RGS3 (GenBank acc. no. AK004648). A partial cDNA clone for another isoform of mouse RGS3, which possesses a 398-amino-acid N-terminal extension containing a PDZ domain, was isolated by interaction cloning using the intracellular portion of ephrin-B1 as bait [16]. Ephrin-B1 is a cell surface anchored ligand for the EphB receptor. A full-length coding sequence was assembled and named PDZ-RGS3 because the C-terminal 532 amino acids of the predicted 930-amino-acid protein shared considerable homology with human RGS3. PDZ-RGS3 mediated the inhibition of SDF-1 induced cerebellar granule cell chemotaxis triggered by the engagement of ephrin-B1 by EphB2 [16].

Here we report the identification of the human PDZ-RGS3 and show that it arises by the use of 10 upstream exons and 6 exons from the originally characterized RGS3. The 10 upstream exons overlap a previously characterized coding region for a mouse protein termed C2PA [17]. Another RGS3 isoform that uses 17 upstream coding exons along with the same 6 downstream coding exons also exists. Furthermore, we show that an alternative exon located in intron 4 of the originally characterized human RGS3 encodes the N-terminal 140 amino acids of a short form of RGS3. We propose that these four isoforms of RGS3 be called C2PA-RGS3, PDZ-RGS3, RGS3, and RGS3S.

Identification of Short Forms of Human RGS3 and Mouse RGS3

To create a targeting construct for *Rgs3*, we isolated genomic clones that spanned the known coding region for RGS3. We isolated, subcloned, and sequenced two *HindIII* fragments. Based on the nucleotide sequence of the identified *Rgs3* exons, we assembled the coding region of the long and short versions of RGS3, RGS3, and RGS3S. The predicted coding regions agreed with the previously published short form of mouse RGS3 [12] and, when it became available in the database, the predicted amino acid sequence derived from a *Rgs3* cDNA (GenBank acc. no. AK004648). The N-terminal 21 amino acids of mouse RGS3S originated from an alternative exon, which created a coding region for a 192-amino-acid protein that shared its C-terminal 171 amino acids with RGS3. The presence of a stop codon upstream of the predicted initiating ATG indicated that no further 5' coding sequence likely exists.

Next, we used nucleotide sequence information from the unique portion of the mouse cDNA that encodes for RGS3S to search for a similar human cDNA. A BLASTN search of the human expressed sequence tags (ESTs) identified such an EST (GenBank acc. no. BF37203). However, it contained 168 nucleotides of open reading frame 5' of the assigned initiation codon for mouse RGS3S, although it lacked an upstream initiation codon. Next, we used the BF37203 EST to carry out a BLASTN search of the NCBI human genome

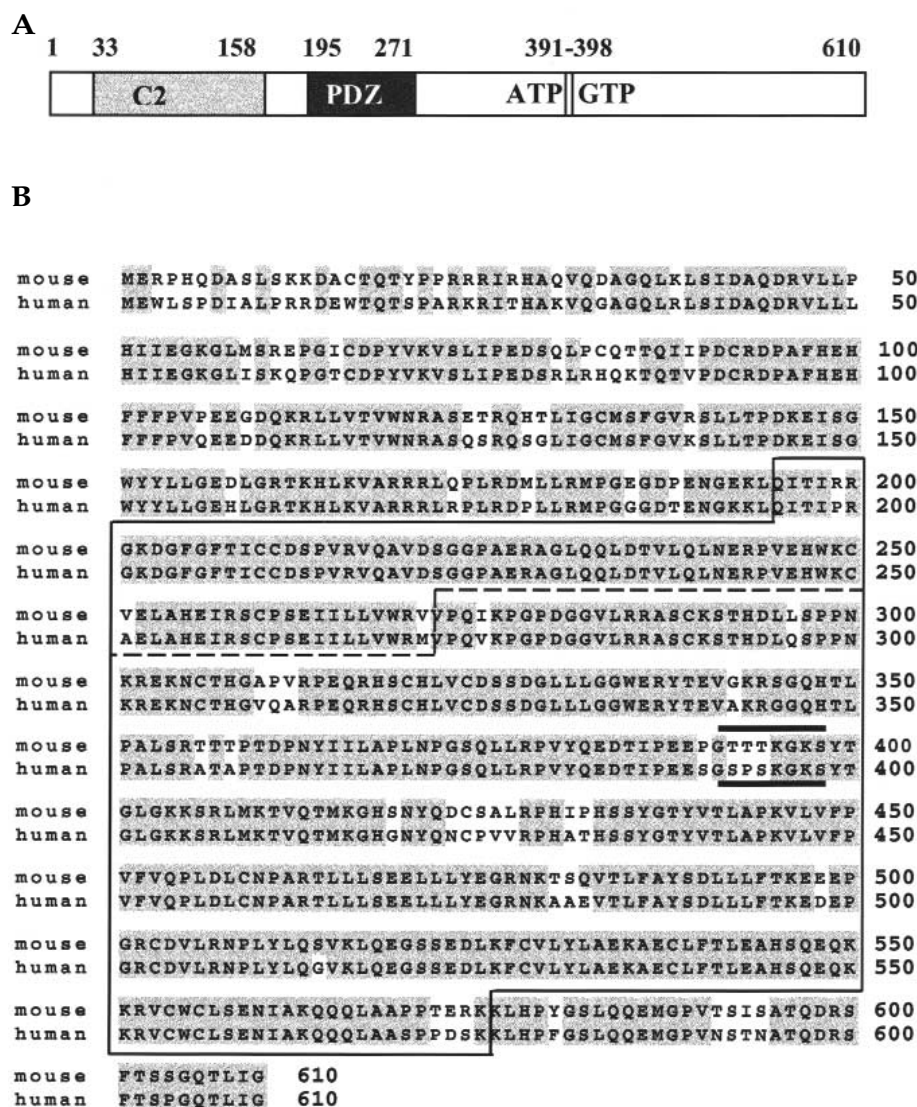


FIG. 2. Comparison of mouse and human C2PA amino acid sequences. (A) Structure of C2PA. (B) Alignment of the mouse and human proteins. The amino acid sequence of C2PA was assembled from the NCBI genomic sequences using the mouse C2PA amino acid sequence as a guide. Conserved amino acids are shown in gray. The amino acids shared with the PDZ-RGS3 protein are boxed. A dotted line delimits the PDZ domain and the putative ATP/GTP binding site is over- and underlined.

database. We identified an exon in intron 4 of the original *RGS3* exons located 91 bases upstream of exon 5 and found an upstream in-frame ATG at position 624,263 in the NT_017568 contig, a portion of the working sequence for human chromosome 9 (Fig. 1). The genomic sequence extended the open reading frame 189 nucleotides beyond the BF37203 sequence. Together they added an additional 140 amino acids of coding sequence, resulting in a human *RGS3S* protein of 311 amino acids with a predicted molecular mass of 35 kDa. Using PCR we amplified a cDNA that encodes such a protein from reverse-transcribed human spleen mRNA (GenBank acc. no. AF490838).

The Identification of Human PDZ-RGS3

Examination of the recently published amino acid sequence for mouse PDZ-RGS3 revealed that the C-terminal portion was identical to mouse *RGS3*, indicating that PDZ-RGS3 likely arose

by an RNA splicing event. A direct comparison of the AK004648 nucleotide sequence and the nucleotide sequence that encodes PDZ-RGS3 indicates that AK004648 is a partial cDNA sequence for the PDZ-RGS3 isoform (there are some minor disagreements between the two sequences). A TBLASTN search of the NCBI nonredundant database with the N-terminal 398 amino acids of mouse PDZ-RGS3 revealed that a portion aligned with the coding region for a previously reported open reading frame termed C2PA [17] and the predicted coding region of a human cDNA. Another TBLASTN search with the N-terminal 398 amino acids of mouse PDZ-RGS3 using the NCBI human genome database revealed partial identity with a cluster of likely exons that overlapped the identified exons from which the human C2PA cDNA arose, which just happened to lie 40 kb away from the previously identified *RGS3* cluster of exons on chromosome 9. This suggested that humans and mice possess a highly related and

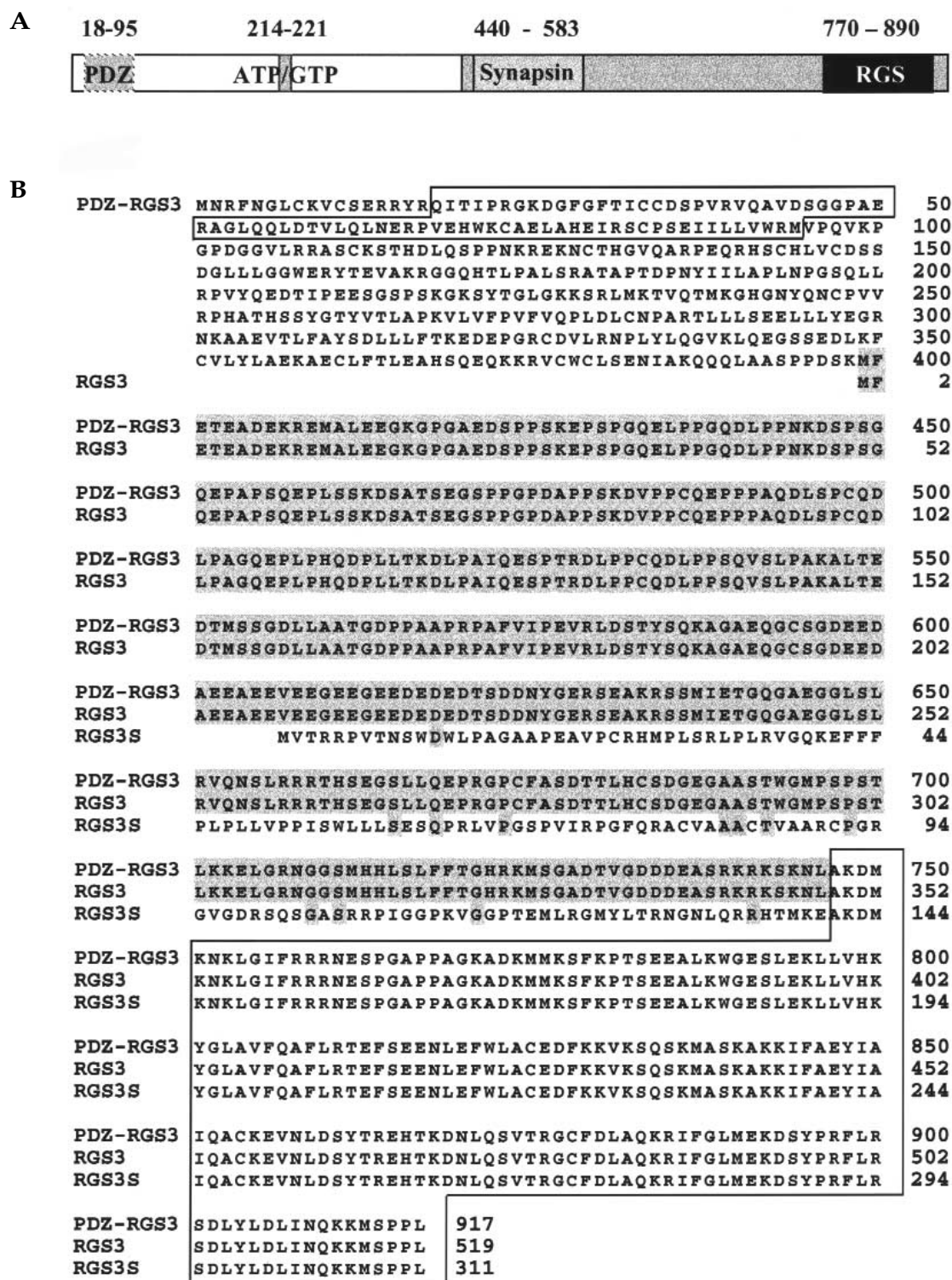


FIG. 3. Alignment of human PDZ-RGS3 with RGS3 and RGS3S. (A) A schematic of the human PDZ-RGS3 isoform. The PDZ, ATP/GTP binding site, synapsin homology domain, and the RGS domains are indicated. (B) Alignment of three human RGS isoforms. The PDZ domain of PDZ-RGS3 is boxed, those amino acids shared between human RGS3 and PDZ-RGS3 are highlighted in gray, and the amino acids shared by all three isoforms are boxed.

complex coding region, which spanned what had been presumed to be two separate genes. As the reported human C2PA coding region only encompassed a portion of mouse C2PA, we used results from BLASTN searches of the NCBI human genome sequences with cDNA sequences that encode human C2PA and mouse C2PA to assemble a predicted C2PA coding region. Like the mouse C2PA [17], the human ortholog has 610 amino acids, a C2 membrane binding domain, an ATP/GTP binding site, and a PDZ domain. The human and mouse proteins share 90% amino acid identity (Fig. 2). Alignment of mouse C2PA and mouse PDZ-RGS3 revealed that amino acids 18–498 of PDZ-RGS3 derived from the region that encoded C2PA, but the N-terminal 17 amino acids did not align with any region of the C2PA protein. Another TBLASTN search of the NCBI genome at reduced stringency with the mouse N-terminal 17 amino acids identified an exon predicted to encode an identical sequence in an intron of the human C2PA coding region. Thus, we could assemble the complete amino acid sequence of the human counterpart of mouse PDZ-RGS3. Figure 3 shows a schematic of the human PDZ-RGS3 protein and compares the amino acid sequences of human PDZ-RGS3, RGS3, and RGS3S.

The coding region for human C2PA spans 18 exons and approximately 80 kb on chromosome 9. The first coding exon of human PDZ-RGS3 is located in the intron between exons 8 and 9. PDZ-RGS3 amino acids 18–398 are derived from exons 9 through 17 and the remainder of the protein from exons 3–8 of the original RGS3. The ATG for the initiator methionine for RGS3 lies precisely at the beginning of exon 3, thereby allowing the splicing of the upstream exons in-frame with the coding region for RGS3. The coding region of human PDZ-RGS3 does not use exons 1–8 or exon 18 of the coding sequence for C2PA (Fig. 4). Therefore, both the human and mouse PDZ-RGS3 lacks the C2 domain present in C2PA and the C-terminal 35 amino acids.

The splice donor and acceptor sequences involved in creating the alternative mRNAs did not seem to be unusual. The exon 17 splice donor sequence is used for the mRNA transcripts that encode human C2PA and PDZ-RGS3. It conformed to the consensus sequence for a 5' splice donor site. The exon 18 splice acceptor site (TCTGGTCCAC/AG) located approximately 4.4 kb away is not optimal as it deviates slightly from the consensus site for a 3' splice acceptor (ideally the two G's should be replaced by either T or C). In contrast the 3' splice acceptor site preceding the original RGS3 exon 3 conforms to the consensus site and is preceded by a long, pyrimidine-rich sequence. Whether these variations have any significance remains unclear.

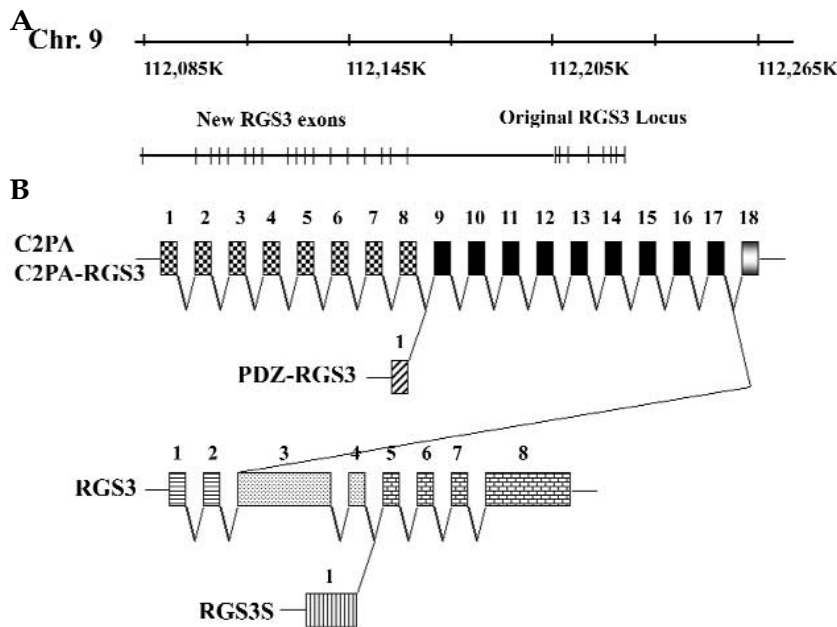


FIG. 4. Origins of human C2PA-RGS3 and PDZ-RGS3 from the extended RGS3. (A) Schematic of RGS3 on chromosome 9. The schematic is derived from master contig map of human chromosome 9. The region displayed is between 112,085K–112,265K base pairs. Only a portion of the human C2PA coding region had been previously localized. Coding exons are designated by vertical lines. (B) The splicing pattern presumed to generate four human RGS3 isoforms. The introns and exons are shown schematically and not scaled according to their size. Shown are the only known unique C2PA coding exon (texture); the two noncoding exons uniquely used by RGS3 (horizontal stripes); the unique RGS3S exon (vertical stripes); the unique PDZ-RGS3 exon (slanted stripes); shared exons between C2PA, C2PA-RGS3, and PDZ-RGS3 (black); exons used in all the RGS3 isoforms (bricks); and exons used in all the RGS3 isoforms except RGS3S (dots).

Verification of the Human PDZ-RGS3 Coding Region and Identification of a Fourth RGS3 Isoform

We confirmed our construction of human PDZ-RGS3 by PCR amplification of the coding region of PDZ-RGS3 from a Quick Screen human cDNA library panel. We identified the appropriate size PCR fragment, and nucleotide sequencing revealed that it encoded human PDZ-RGS3 (GenBank acc. no. AF490840). Next, we tested whether a fourth isoform might exist that included the coding regions for amino acids 1–575 of C2PA and 1–519 of RGS3, which we termed C2PA-RGS3. We amplified such a cDNA from reverse-transcribed testis mRNA confirming its existence (GenBank acc. no. AF490839). Furthermore, a search of the human EST database identified a germinal center B cell EST (GenBank acc. no. AA766285), which on initial inspection appeared to encode C2PA, but complete nucleotide sequencing revealed that it contained C2PA sequence 5' of the exon 8/9 junction and RGS3 exon 3 nucleotide sequence 3' of the expected junction between exons 17 and 18 (data not shown). Such an mRNA would be predicted to encode a 1094-amino-acid protein with a C2 domain, a PDZ domain, a GTP/ATP binding site, a synapsin homology domain, and a RGS domain.

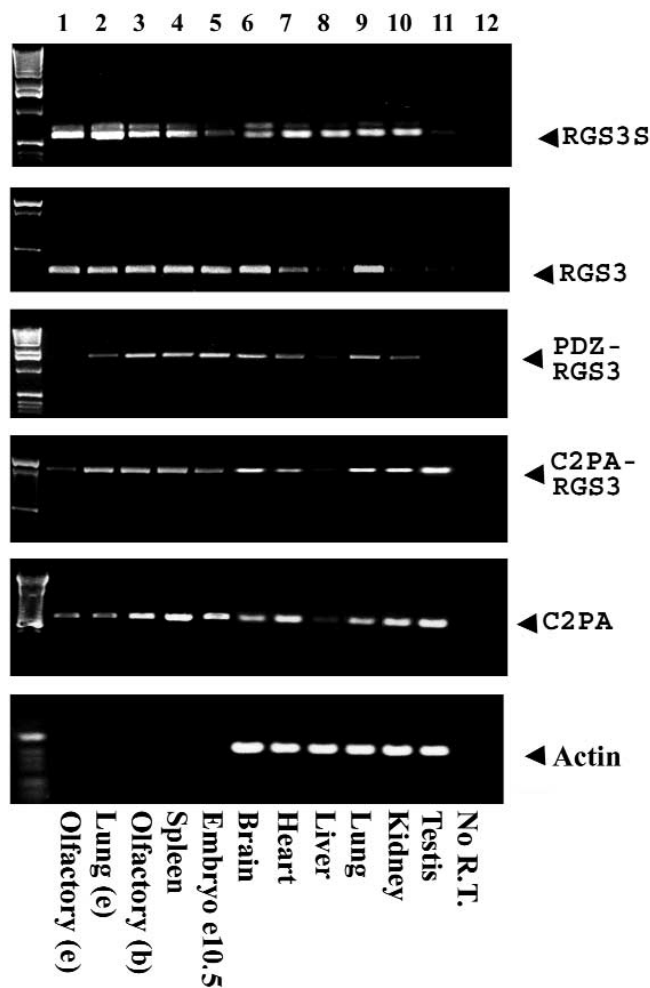


FIG. 5. RT-PCR analysis of mRNAs arising from the regions that encode mouse RGS3S, RGS3, PDZ-RGS3, C2PA-RGS3, and C2PA. cDNAs from RNA prepared from the following mouse tissues: olfactory epithelium (lane 1), respiratory epithelium (lane 2), olfactory bulb (lane 3), spleen (lane 4), E10.5 embryo (lane 5), brain (lane 6), heart (lane 7), liver (lane 8), lung (lane 9), kidney (lane 10), and testis (lane 11). Tissues were analyzed by PCR with primers specific for mouse RGS3S, RGS3, PDZ-RGS3, C2PA-RGS3, C2PA, and β -actin. Olfactory epithelium mRNA was used for the no RT control PCR reaction (lanes 12). One kb (RGS3S and PDZ-RGS3) and 100-bp ladders (RGS3, PDZ-RGS3, and C2PA) are shown. Lanes 1–5 and lanes 6–12 are results from two separate experiments. The β -actin control for the lanes 1–5 is not shown.

epithelium, liver, and testis. To specifically amplify mRNAs encoding C2PA-RGS3, we moved the 5' primer to 5' of the C2PA exon 8/9 junction and used the same RGS3 3' primer. We visualized bands at the predicted size in most of the tissues, thereby indicating widespread expression of the mRNA transcripts for C2PA-RGS3 although testis had the highest levels. Finally, we detected transcripts predicted to encode C2PA using a 5' primer based on sequence from the coding region and a 3' primer based on sequence in the 3' untranslated portion of the mRNA that encodes mouse C2PA.

To supplement the RT-PCR analysis, we carried out a northern blot analysis of 20 mg total RNA prepared from many of the same mouse tissues using a probe that spanned the coding region of RGS3, thereby allowing detection of mRNAs that encode RGS3S, RGS3, PDZ-RGS3, and C2PA-RGS3 (Fig. 6). We noted a considerable variation in the expression levels of the various isoforms among the different tissues, which correlated reasonably well with the RT-PCR analysis. Transcripts for RGS3S predominated in heart, liver, kidney, and skeletal muscle, whereas those for mouse RGS3 predominated in brain, spleen, stomach, and small intestine. Lung, thymus, and uterus had nearly equal amounts of mRNAs encoding RGS3S and RGS3. Most of the mouse tissues expressed relatively low levels of mRNAs for PDZ-RGS3, whereas testis had high levels of transcripts for C2PA-RGS3. Skin and placenta had low levels of all the isoforms.

Detection of mRNAs for Mouse RGS3 Isoforms by RT-PCR and Northern Blot Analysis

Next, we designed a set of PCR primers to detect expression of mRNAs that encode mouse RGS3S, RGS3, C2PA, PDZ-RGS3, and C2PA-RGS3. We examined cDNA derived from a variety of tissues (Fig. 5). With the primers designed to detect mRNA transcripts encoding RGS3S, we detected two bands, one of the appropriate size and one slightly larger, which based on nucleotide sequencing contained intronic sequence. All the tissues tested expressed mRNA for RGS3S, although relatively low levels were found in testis and brain. We detected low levels of transcripts for RGS3 in RNA prepared from liver, kidney, and testis. Next, using a primer pair predicted to amplify mRNAs encoding both mouse PDZ-RGS3 and C2PA-RGS3 (the 5' primer from shared coding sequence), we readily detected the appropriate size product from all the different tissues (data not shown). To specifically amplify mRNAs for PDZ-RGS3, we used a 5' primer from the unique first coding exon of PDZ-RGS3 along with the same 3' primer from RGS3. The highest level was found in embryonic day (E) 10.5 mouse and relatively low levels were seen in olfactory

Detection of RGS3 Isoforms by Western Blot Analysis

Next, we used an antibody to RGS3, raised against the N-terminal region of human RGS3, to analyze cell lysates prepared from a variety of cell lines. The antibody recognizes RGS3, PDZ-RGS3, and C2PA-RGS3, but not RGS3S, when those proteins are expressed in HEK 293 cells (data not shown). We detected prominent bands at approximately 72 and 125 kDa that migrated at a rate similar to what we have observed for RGS3 and PDZ-RGS3 expressed in HEK 293 cells. In order to easily visualize the bands, we used a biotinylated secondary antibody to amplify the signal. RGS3 ran as a doublet as we had previously observed [8]. K562 cells, an erthroleukemia cell line, expressed high levels of RGS3, but low levels of PDZ-RGS3 (Fig. 7, lane 4). NG-108 cells (a mouse glioblastoma/rat neuroblastoma hybrid cell line) and PC-12 cells (a rat adrenal pheochromocytoma cell line) expressed relatively high amounts of PDZ-RGS3 (Fig. 7, lanes 5 and 6). The bands observed at approximately 66 kDa are likely nonspecific, but

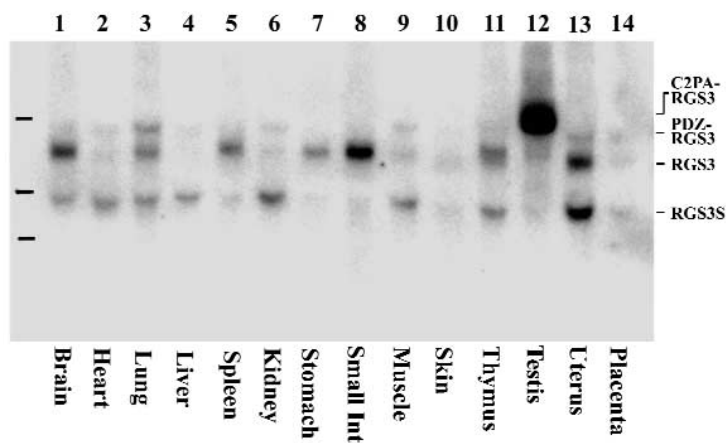


FIG. 6. Northern blot analysis of multiple mouse tissues using an RGS3 cDNA probe. A multiple tissue northern blot was probed with a cDNA probe that spanned the coding region of mouse RGS3. Assigned bands for the mRNA transcripts that encode mouse RGS3S, RGS3, PDZ-RGS3, and C2PA-RGS3 are indicated. The tissue origin of each RNA sample is indicated below. The markers indicate 4.7, 1.8, and 1.0 kb from top to bottom.

the 38- to 40-kDa bands may be a degradative product of RGS3 because we often observe similar bands when RGS3 is expressed in bacteria. In the cell lines we examined, we did not detect a band likely to correspond to C2PA-RGS3. While the approach of using a biotinylated secondary antibody significantly increased the sensitivity of the RGS3 immunoblot, when we extended our analysis to tissue extracts it produced an unacceptable level of background reactivity.

DISCUSSION

Human PDZ-RGS3 is assembled from two clusters of exons located on human chromosome 9 separated by approximately 40 kb of intervening nucleotide sequence. This assembly requires 10 upstream exons from one cluster and 6 exons from the other cluster, which span the previously known coding region of RGS3. In addition, another isoform termed C2PA-RGS3 exists that derives from 17 upstream coding exons along with the same coding exons of RGS3 as does PDZ-RGS3. RGS3S derives from an exon present in intron 4 of the original RGS3 along with the three most 3' coding exons. The four RGS3 isoforms share exons 5–8 from the original RGS3 [18], which together encode the shared RGS domain.

How are the human PDZ-RGS3 and C2PA-RGS3 mRNA transcripts generated? Because 47,000 bp separate exon 8 of the C2PA coding region and exon 3 of the original cluster of RGS3 exons, the generation of a pre-mRNA that spans them seems a considerable feat. Yet, very large introns exist in other human genes. Two forkhead genes, *FKHR* and *FKHRL1*, each possess a genomic organization that includes a first intron greater than 90 kb [19]. An approximately 150-kb intron separates exons 2 and 3 of the protein kinase $C\beta$ gene [20]. Alternatively, a *trans*-splicing mechanism could generate the C2PA-RGS3 and PDZ-RGS3 mRNA transcripts. Mammalian cells have the ability to generate functional mRNAs by *trans*-splicing [21], but *trans*-splicing would not require an adjacent chromosomal localization of the two clusters of exons, thereby making this an unlikely explanation.

Northern blot analysis using a probe that spans the coding region of RGS3 of a variety of adult human tissues revealed four transcripts: two smaller, prominent transcripts and two larger, less prominent transcripts [5]. RNA prepared from human heart contained all four transcripts, whereas most other tissues had either the 4.2 kb mRNA or the 4.5 kb mRNA, or neither large mRNA. Testis RNA possessed a prominent band for the 4.5-kb

mRNA without a visible 4.2-kb mRNA band. The two smaller transcripts likely encode human RGS3 and RGS3S, whereas the two larger transcripts likely encode PDZ-RGS3 and C2PA-RGS3. A northern blot analysis of *Rgs3* mRNAs in a variety of adult mouse tissues revealed an overall similar pattern of expression with some exceptions. While we detected four distinct mRNA transcripts using an *Rgs3* cDNA of approximately the same mobility as in humans, only mouse testis had a detectable 4.5-kb mRNA, whereas the majority of the other mouse tissues expressed varying amounts of the 4.2-kb mRNA. Although the human northern blot used oligo dT purified RNA and the mouse blot used total RNA, we visualized the mouse 4.2-kb mRNAs in most tissues suggesting that we should have detected the 4.5-kb mRNAs had they been present at similar levels. Nevertheless, further studies using poly(A)⁺ mouse RNA did reveal low levels of the 4.5-kb *Rgs3* mRNA in brain (K.H., unpublished data).

Despite the difficulties in finding mRNAs that account for C2PA-RGS3 in tissues other than testis by northern blot analysis, we detected them by RT-PCR. In addition we detected appropriate RT-PCR products from mRNAs for mouse PDZ-RGS3, RGS3S, RGS3, and C2PA. In addition, we amplified the correct PCR product with primers from the first exon of PDZ-RGS3 and from the 3' untranslated portion of the coding region for C2PA, indicating that another C2PA isoform exists, which lacks coding sequence for an RGS domain (D.S., unpublished data). Furthermore, by further analysis of the human EST database we have identified two alternative first exons for the coding region of human C2PA, one that is noncoding, thereby producing a protein predicted to be initiated from a downstream ATG (amino acid 59 in the C2PA coding sequence; EST acc. no. AA824242), and another that produces an alternative N terminus (MERSLHRVSLGSRRAHPDLSYLTF) that joins the remainder of the C2PA amino acid sequence at amino acid 35 (EST acc. no. AW605207). We are currently examining whether these two isoforms also connect to the coding regions for the RGS domain.

Western blot analysis with an antiserum raised against the C-terminal region RGS3 protein readily detected human RGS3. RGS3 runs as a doublet and slower on SDS-PAGE than pre-

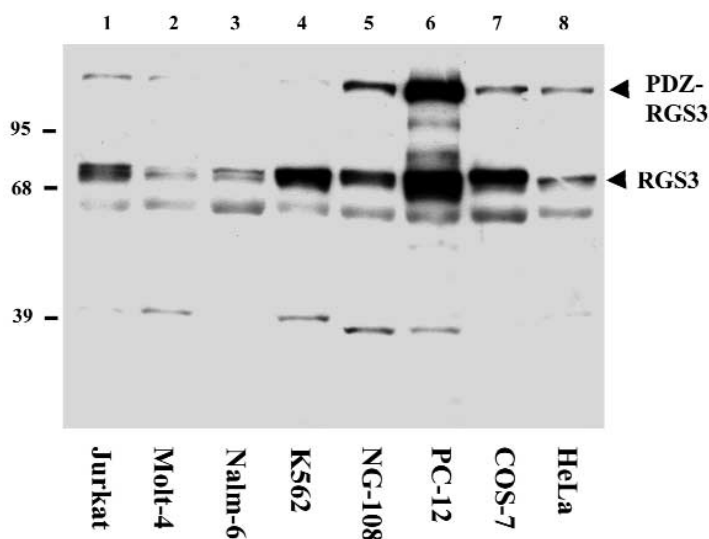


FIG. 7. Western blot analysis of lysates from various cell lines using an antibody that recognizes the N terminus of human RGS3. Lysates (50 μ g) were prepared from various cell lines (Jurkat (lane 1), Molt-4 (lane 2), Nalm-6 (lane 3), K562 (lane 4), NG-108 (lane 5), PC-12 (lane 6), COS-7 (lane 7), and HeLa (lane 8)), size fractionated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with the RGS3 antiserum. The locations of the indicated molecular weight markers are shown.

dicted from its molecular mass. The same antiserum detected a slower mobility band that likely corresponds to human PDZ-RGS3. While the estimated molecular mass of 125 kDa is consistent with the predicted molecular mass of C2PA-RGS3 (126 kDa), the aberrant mobility of RGS3 is against this assignment. Furthermore, PDZ-RGS3 expressed in mammalian cells migrates at 125 kDa, supporting the designation of the 125-kDa band as PDZ-RGS3. While we documented the expression of mRNAs predicted to encode human C2PA-RGS3 by RT-PCR, western blot did not detect the protein in any of the cell lines we examined. We are planning further studies using antibodies raised against the C2PA portion of C2PA-RGS3. The antiserum we used in this study does not detect RGS3S, but we have recently produced an antibody that specifically recognizes the unique N terminus of RGS3S. Immunoblotting with this antibody detected a 34-kDa human protein in several cell lines consistent with the predicted molecular mass of the coding region for human RGS3S derived from cDNA and genomic sequences (C.-S.S., unpublished data).

While western blot analysis can distinguish the RGS3 isoforms based on their different molecular masses, discriminating some of the isoforms in tissue sections will be impossible. Furthermore, because C2PA-RGS3, PDZ-RGS3, and C2PA share eight coding exons, RNA probes originally designed to detect mRNAs encoding mouse PDZ-RGS3 or mouse C2PA may have inadvertently detected transcripts arising from two or three other mRNAs. The probe used to detect transcripts that encode PDZ-RGS3, nucleotides 3–861 of the mouse PDZ-RGS3 cDNA, contained shared nucleotide sequence with other isoforms, and the probes used for northern blot to detect transcripts that encode C2PA (nucleotides 818–2068) and *in situ* hybridization (nucleotides 818–4712) both overlapped the coding regions for mouse C2PA-RGS3 and PDZ-RGS3 [16,17]. Also, the mRNAs that encode C2PA and PDZ-RGS3 are similar in size making their distinction on northern blot difficult unless specific probes are used.

C2PA, C2PA-RGS3, and PDZ-RGS3 share the same PDZ domain, putatively classified as a type I PDZ domain and resembling those of mouse rhophilin, a *Drosophila melanogaster* guanine nucleotide exchange factor, RGS12, and PSD95 (EMBL accession numbers in [16]). Class I PDZ domains bind a short motif, S/T/X/V, at the C terminus of many membrane proteins [22]. However, amino acid substitutions in the C2PA and PDZ-RGS3 PDZ domains at two contact residues likely alter their target sequence. Indeed, the related motif, Y/Y/K/V, at the C terminus of the known B ephrins presumably mediated the PDZ-RGS3 and Ephrin-B1 interaction [16]. Because C2PA, C2PA-RGS3, and PDZ-RGS3 share the same PDZ domain the three proteins may compete for binding to the same receptors. Consistent with that possibility, a construct designed to express PDZ-RGS3 without an RGS domain, PDZ-RGS3^{ARGS}, interfered with the activity of PDZ-RGS3 [16]. Thus, the ratio of these three proteins may critically regulate the availability of the RGS domain of RGS3.

The identification of upstream coding exons in the human RGS3 and mouse *Rgs3* loci has led to the assembly of several unexpected coding regions enriching, but complicating, our understanding of the functional importance of RGS3 isoforms. Because each of the coding regions may possess a different promoter, the expression patterns of the multiple RGS3 isoforms may differ substantially as the RGS3 northern blot analysis and western blot data attest. Gene targeting experiments to remove either individual RGS3 isoforms or different combinations should help delineate their physiologic roles.

MATERIALS AND METHODS

Identification of human RGS3S. We isolated and sequenced a 9.4-kb murine *Hind*III fragment spanning exons 5–8 of *Rgs3*. We used BLASTN searches to identify the *Rgs3* exons, including an exon predicted to encode the first 21 amino acids of a short form of mouse RGS3 (GenBank acc. no. AAF34626). A BLASTN search of the human EST database using the nucleotide sequence that encodes mouse RGS3S identified an expressed sequence tag for the human RGS3 short form (GenBank acc. no. BF684647), which was used to perform a BLASTN search of the NCBI human genome database. This identified an exon located just 5' of exon 5 in the human RGS3 gene [18], which encodes the N-terminal amino acids of RGS3S.

Identification of the human C2PA, PDZ-RGS3, and C2PA-RGS3 coding exons and the derived amino acids. A TBLASTN search of the NCBI nonredundant database using the first 398 amino acids of mouse PDZ-RGS3 demonstrated identity between amino acids 18–398 of PDZ-RGS3 and amino acids 195–575 of the coding region for mouse C2PA (NM_019492), and some identity with the coding region from a homologous human cDNA (GenBank acc. no. AK000377), which was predicted to encode a 341-amino-acid protein. We used the mouse C2PA and human AK000377 cDNA sequences in a BLASTN search of the

human NCBI genome contig sequences. Additional TBLASTN searches with the complete mouse C2PA amino acid sequence or various portions were also performed. These searches identified a cluster of exons located on the chromosome 9 master contig map, between 112,085K and 112,170K bp. We identified predicted intron/exon junctions and assembled the human C2PA amino acid sequence. We next used the first 17 amino acids of mouse PDZ-RGS3 in a TBLASTN search of the human NCBI genome database. They aligned perfectly with the predicted amino acid sequence of a nucleotide sequence within intron 8 of the human C2PA coding sequence. The initial ATG was located at nucleotide 531,369 of the chromosome 9 working draft sequence segment NT_17568, approximately 112,125K bp on the master contig map. We assembled the human PDZ-RGS3 amino acids from this first exon-predicted amino acid sequence, the predicted amino acid sequence from exons 9-17 of the human C2PA coding region, and the original RGS3 amino acid sequence. A BLASTN search of the human NCBI EST database with a mouse C2PA cDNA (GenBank acc. no. AJ250999) identified a NCI_CGAP_GCB1 cDNA clone (GenBank acc. no. AA766285). Complete nucleotide sequence of the cDNA insert (1494 nt) followed by a BLASTN search of the nonredundant database revealed that it overlapped the coding regions for human C2PA and RGS3.

Cloning human PDZ-RGS3, C2PA-RGS3, and RGS3S. The PCR primers used to amplify the coding region of human PDZ-RGS, RGS3S, and C2PA-RGS3 were as follows: PDZ-RGS3, 5' primer, 5'-ATGAACCGCTTCAATGGGCTCTGCAAGGTG-3', 3' primer, 5'-AAGCGGGGACTCATCTTCTTCTGGTTAATAAGGTC-3'; RGS3S, 5' primer, 5'-ATGGTAACGAGGAGGCCAGTCACAAATAGC-3', 3' primer, same as above; C2PA-RGS3, 5' primer, 5'-ATGGAGTGGCTAAGCCCTGATATCGTCTG-3', 3' primer, same as above.

A Quick Screen human cDNA library panel (PDZ-RGS3 and RGS3S; Clontech, Palo Alto, CA) or mouse testis mRNA, which had been reverse-transcribed (C2PA-RGS3), were subjected to 40 cycles of PCR amplification with the appropriate primers. We identified, subcloned, and determined the nucleotide sequences of the correct PCR products (GenBank acc. no. AF490838-40).

RNA preparation, RT-PCR analysis, and northern blot analysis. We isolated RNA from adult mice tissues and from E10.5 mice using the Ultraspec RNA isolation system (Biotex Laboratories, Houston, TX) and reverse transcribed 1 µg of RNA using Omniscript reverse transcriptase for first-strand cDNA synthesis (Qiagen, Valencia, CA). Next, we performed 30-35 cycles of PCR with HotStarTaq (following the manufacturer's protocol, Qiagen) using one-tenth of the cDNA for each PCR reaction. Each PCR was also performed with sham cDNA (no reverse transcriptase) to monitor for the presence of spurious DNAs. We fractionated the PCR products on a 1% agarose gel and stained them with ethidium bromide. The nucleotide sequences of the primers and their predicted PCR product sizes are as follows: RGS3S, 5' primer, 5'-TCCTGAGTCTCAAGTGGGGGAC-3' (AF215669, nt 57-84), 3' primer, 5'-CAGAGCGGAGGAAGCGAGGGTAAGA-3' (AF215669, nt 618-594), predicted product 562 nt; RGS3, 5' primer, 5'-GTGCTTATTCACCTTGGAGGCAC-3' (AK004648, nt 311-335), 3' primer, 5'-TGGGTGGGAGGTCTTGTCTACA-3' (AK004648, nt 720-696), predicted product 410 nt; C2PA-RGS3 and PDZ-RGS3, 5' primer, 5'-CCGACTCTCTGCTGTGCGGAGCA-3' (AJ250999, nt 1621-1645), 3' primer, same as RGS3, predicted product 635 nt; PDZ-RGS3, 5' primer, 5'-GGGCTCTGCAAAGTGTGTTCA-GAA-3' (AF350047, nt 66-84), 3' primer, same as RGS3, predicted product 1452 nt; C2PA-RGS3, 5' primer, 5'-CAGTGAGACCAGGCAGCATAAGCT-3' (AJ250999, nt 601-624), 3' primer, same as RGS3, predicted product 1655 nt; C2PA, 5' primer, 5'-CCGACTCTCTGCTGTGCGGAGCA-3' (AJ250999, nt 1621-1645), 3' primer, 5'-GCCAGGAAGGTGCCAGGTGCCA-3' (AJ250999, nt 2341-2317), predicted product 721 nt.

A mouse multiple tissue blot was purchased from Seegene (Seoul, Korea) and probed with a RGS3 cDNA that spanned the coding region of mouse RGS3. The blot was hybridized and washed using standard high-stringency conditions. Equal RNA loading was verified by ethidium bromide staining.

Western blot analysis of human RGS3 proteins. We analyzed cell lysates prepared from the following cell lines: Jurkat (human T cell leukemia), Molt-4 (human T cell leukemia), Nalm-6 (human pre-B cell), K562 (erythrocyte leukemia), NG-108 (rat/mouse hybrid neuroblastoma), PC-12 (rat

pheochromocytoma), Cos-7 (monkey kidney), and HeLa (human epitheloid carcinoma). After blocking the membrane, we incubated it with an N-terminal specific rabbit anti-RGS3 antibody (1:400 [8]) followed by a biotinylated goat anti-rabbit antibody (Dako Corporation, Carpinteria, CA). The complexes were detected using streptavidin conjugated to horseradish peroxidase (Dako) and enhanced chemiluminescence (Amersham, Piscataway, NJ).

ACKNOWLEDGMENTS

We thank Mary Rust (NIAID, NIH) for editorial assistance; Astrid Scheskonka (Georg-August Universität, Göttingen, Germany) for preparation of the N-terminal RGS3 antibody and the initial studies with the antiserum; John Yuen for the isolation of the mouse RGS3 genomic clone; and Anthony Fauci (NIAID, NIH) for support

RECEIVED FOR PUBLICATION JUNE 22, 2001;
ACCEPTED MARCH 14, 2002.

REFERENCES

- Hepler, J. R., and Gilman, A. G. (1992). G proteins. *Trends Biochem. Sci.* **17**: 383-387.
- Neer, E. J. (1995). Heterotrimeric G proteins: organizers of transmembrane signals. *Cell* **80**: 249-257.
- Kehrl, J. H. (1998). Heterotrimeric G protein signaling: roles in immune function and fine-tuning by RGS proteins. *Immunity* **8**: 1-10.
- Hepler, J. R. (1999). Emerging roles for RGS proteins in cell signalling. *Trends Pharmacol. Sci.* **20**: 376-382.
- Druey, K. M., Blumer, K. J., Kang, V. H., and Kehrl, J. H. (1996). Inhibition of G-protein-mediated MAP kinase activation by a new mammalian gene family. *Nature* **379**: 742-746.
- Dulin, N. O., et al. (1999). RGS3 inhibits G protein-mediated signaling via translocation to the membrane and binding to Gα11. *Mol. Cell. Biol.* **19**: 714-723.
- Chatterjee, T. K., Eapen, A. K., and Fisher, R. A. (1997). A truncated form of RGS3 negatively regulates G protein-coupled receptor stimulation of adenylyl cyclase and phosphoinositide phospholipase C. *J. Biol. Chem.* **272**: 15481-15487.
- Scheschonka, A., et al. (2000). RGS3 is a GTPase-activating protein for Gαi and Gαq and a potent inhibitor of signaling by GTPase-deficient forms of Gαq and G11α. *Mol. Pharmacol.* **58**: 719-728.
- Neill, J. D., et al. (1997). Potential role for a regulator of G protein signaling (RGS3) in gonadotropin-releasing hormone (GnRH) stimulated desensitization. *Endocrinology* **138**: 843-846.
- Dulin, N. O., et al. (2000). Regulator of G protein signaling RGS3T is localized to the nucleus and induces apoptosis. *J. Biol. Chem.* **275**: 21317-21323.
- Bowman, E. P., et al. (1998). Regulation of chemotactic and proadhesive responses to chemoattractant receptors by RGS (regulator of G-protein signaling) family members. *J. Biol. Chem.* **273**: 28040-28048.
- Reif, K., and Cyster, J. G. (2000). RGS molecule expression in murine B lymphocytes and ability to down-regulate chemotaxis to lymphoid chemokines. *J. Immunol.* **164**: 4720-4729.
- Shi, C. S., et al. (2001). RGS3 inhibits Gβ1γ2 induced inositol phosphate production, mitogen activated protein kinase activation, and Akt activation. *J. Biol. Chem.* **9**: 24293-24300.
- Arai, H., Tsou, C. L., and Charo, I. F. (1997). Chemotaxis in a lymphocyte cell line transfected with C-C chemokine receptor 2B: evidence that directed migration is mediated by βγ dimers released by activation of Gαi-coupled receptors. *Proc. Natl. Acad. Sci. USA* **94**: 14495-14499.
- Neptune, E. R., and Bourne, H. R. (1997). Receptors induce chemotaxis by releasing the βγ subunit of Gi, not by activating Gq or Gs. *Proc. Natl. Acad. Sci. USA* **94**: 14489-14494.
- Lu, Q., Sun, E. E., Klein, R. S., and Flanagan, J. G. (2001). Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. *Cell* **105**: 69-79.
- Linares, J. L., Wendling, C., Tomasetto, C., and Rio, M. C. (2000). C2PA, a new protein expressed during mouse spermatogenesis. *FEBS Lett.* **480**: 249-254.
- Chatterjee, T. K., Eapen, A., Kanis, A. B., and Fisher, R. A. (1997). Genomic organization, 5'-flanking region, and chromosomal localization of the human RGS3 gene. *Genomics* **45**: 429-433.
- Anderson, M. J., Viars, C. S., Czepak, S., Cavenee, W. K., and Arden, K. C. (1998). Cloning and characterization of three human forkhead genes that comprise an FKHR-like gene subfamily. *Genomics* **47**: 187-199.
- Greenham, J., Adams, M., Doggett, N., and Mole, S. (1998). Elucidation of the exon-intron structure and size of the human protein kinase C β gene (PRKCB). *Hum. Genet.* **103**: 483-487.
- Eul, J., Graessmann, M., and Graessmann, A. (1995). Experimental evidence for RNA trans-splicing in mammalian cells. *EMBO J.* **14**: 3226-3235.
- Songyang, Z., et al. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **275**: 73-77.