RGSZ1 and Ret RGS: Two of Several Splice Variants from the Gene RGS20

Sheryll A. Barker, Jun Wang,* David A. Sierra, and Elliott M. Ross†

Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390-9041, USA

*Present address: Merck & Co., Inc., R80Y-205 P.O. Box 2000, Rahway, New Jersey 07065, USA

†To whom correspondence and reprint requests should be addressed. Fax: (214) 648-2994. E-mail: ross@utsw.swmed.edu.

RGSZ1 and Ret RGS, members of the regulator of G-protein signaling (RGS) family, are GTPase-activating proteins (GAPs) with high selectivity for Gαz. We show here that RGSZ1 and Ret RGSZ1 are products of two of several splice variants of one gene, RGS20. RGS20 spans ~107 kb and contains at least seven exons. Five exons account for RGSZ1, including a single exon distinct to RGSZ1 that encodes a newly identified amino-terminal region. The previously described open reading frame (ORF) and 3′ untranslated region are encoded by four downstream exons that also encode about half of Ret RGS. The 5′ end of the RGSZ1 ORF contains several in-frame ATG codons (3–5 depending on the species), and multiple translational start sites may help explain the molecular weight heterogeneity of purified bovine brain RGSZ. Ret RGS replaces the 24 N-terminal amino acid residues of RGSZ1 with a large, N-terminal region that initially distinguished the bovine Ret RGS from human and mouse RGSZ1. This N-terminal domain is encoded by two distinct 5′ exons that are variably combined with the four downstream exons shared with RGSZ1 to produce at least six mRNAs. They encode proteins with N termini that vary in size, hydrophobicity, and the presence of a cysteine string. At least two mRNAs that include the exon that encodes the N-terminal region unique to RGSZ1 were found in brain and a few other tissues, but not retina. RGS20 thus can account for multiple Gαz-selective GAPs in different tissues.

Key Words: RGS protein, G protein, GTPase activating protein, mRNA splicing, RGSZ1, Ret RGS

INTRODUCTION

Regulators of G protein signaling, RGS proteins, are a family of about 20 homologous proteins that act as GTPase-activating proteins (GAPs) for heterotrimeric G proteins of the Gi and Gq classes. By accelerating the rate of hydrolysis of GTP bound to the Gα subunit, they modulate both the speed and amplitude of G-protein signaling [reviewed in 1]. RGS proteins are distinguished by an RGS domain, a globular and mostly α-helical core that binds to the switch regions of Gα subunits to exert GAP activity. In addition, RGS proteins contain multiple other identifiable domains that, along with patterns of similarity in the RGS domains, distinguish RGS subfamilies. Functions of these additional domains include alteration of GAP activity and specificity, membrane targeting, and binding to proteins other than Gα substrates [1–3]. Other possible functions are topics of active study. In the case of RGS9 and RGS12, the presence of such ancillary domains is determined by alternative splicing of mRNA precursors [4–7]. According to the overlay of EST contigs on genomic sequence, RGS8 and RGS18 also produce splice products that eliminate the amino-terminal amphipathic helix (D.A.S. and Thomas M. Wilkie, unpublished data). Such N-terminal variation may be particularly important because this region is involved with membrane or protein binding, subcellular targeting, and receptor-selective regulation [5,8–11]. RGS19, which encodes GAIP, also yields multiple mRNA molecules that differ in the 5′ noncoding region [12].

The RGSZ subfamily, which is composed of GAIP [13] (RGS19), RGSZ1 [14,15] (RGS20), RGSZ2 (RGS17) [16], and Ret RGS (RGS20) [17], is distinguished by a functionally hydrophobic domain N terminal to the RGS domain (encoded by exon D; Fig. 1A). Even though this region is not strongly predicted to span phospholipid bilayers, RGSZ family members behave as tightly bound integral membrane proteins [14,18] and their binding to phospholipid bilayers depends on the N-terminal domain [19]. The N-terminal domain also contains a cysteine string, a site of probable multiple palmitoylation that targets at least one member, GAIP, to specific
intracellular membranes [11]. The importance of the region N terminal to the cysteine string remains uncertain. In addition, the mRNA for human RGSZ1 includes five closely spaced methionine codons in this region, and the physiological translational initiation sites are not known. Multiple sites may contribute to the observed heterogeneity of RGSZ1 purified from brain [14]. One of these corresponds to the presumed start codon in GAIP [13] and another corresponds to the start codon in RGSZ2. RGSZ proteins are highly active GAPs for G$z$ (Gi$z$) family. They accelerate hydrolysis of G$z$-bound GTP over 600-fold with a $K_m$ of about 2 nM [14]. RGSZ1 and Ret RGS are also quite selective for G$z$ over other Gi family members, whereas GAIP and RGSZ2 are nonselective GAPs for the Gi family ([14]; and S.A.B., unpublished data). RGSZ proteins are weak GAPs for the Gq family.

RGSZ1, identified as the major G$z$ GAP in bovine brain and subsequently cloned as a human cDNA [14,15,18], is quite similar in sequence to Ret RGS, which was identified as a bovine cDNA whose expression is limited to retinal interneurons [17,20]. Although the coding sequences were closely related over the length of RGSZ1, the Ret RGS cDNA encoded an additional N-terminal domain of 156 amino acids. The N termini of RGSZ1 and GAIP were originally thought to be of similar size. Molecular weights of RGSZ1 and Ret RGS in their tissues of origin were consistent with those predicted from the cDNAs, although G$z$ GAP purified from brain was heterogeneous in size over the 22–29 kDa size range [18]. It was thus not clear whether Ret RGS and RGSZ1 are orthologs, perhaps with an additional domain in the bovine protein, or paralogs distinctive to retina and brain. We show here that both RGSZ1 and Ret RGS are mRNA products of a single gene, RGS20. They represent two of six RGS20 mRNAs that differ in the sizes and sequences of their N-terminal domains and that should therefore differ in their attachment to membranes and their subcellular distributions.

**RESULTS AND DISCUSSION**

RGSZ1 and Ret RGS [14,15,20] are closely related members of the RGSZ subfamily of RGS proteins, which also includes RGSZ2 (RGS17) [16] and GAIP (RGS19) [13]. The amino acid sequences of both human and mouse RGSZ1 [14] (NM_021374) are highly similar to residues 156–374 of bovine Ret RGS [20]. This region includes both the N-terminal cysteine string and the RGS domain. It was therefore unclear whether RGSZ1 and Ret RGS were different paralogous genes or whether the cloned cDNAs represented orthologs. A search of human genomic databases for Ret RGS and RGSZ1 genes found unordered sequence fragments of a BAC insert containing both the previously described N terminus of RGSZ1 and an ORF similar to the N terminus of Ret RGS [20]. This region includes both the N-terminal cysteine string and the RGS domain. It was therefore unclear whether RGSZ1 and Ret RGS were different paralogous genes or whether the cloned cDNAs represented orthologs. A search of human genomic databases for Ret RGS and RGSZ1 genes found unordered sequence fragments of a BAC insert containing both the previously described N terminus of RGSZ1 and an ORF similar to the N terminus of Ret RGS. Another BAC (RP11-410C17, AC079335.1) contained sequence that encoded the RGS domain of RGSZ1. The insertion of RP11-421C10 maps to the
end of chromosome 8, and the other is not mapped. These
data indicate either that a single, alternatively processed gene
encodes the two proteins or that unique genes for each pro-
tein lie nearby on chromosome 8. Mouse interspecies back-
cross mapping indicated that a probe for the RGS domain of
mouse RGSZ1 also hybridizes to a single site, 0.0 cM on mouse chro-
mosome 1, a region syntenic to human chromosome 8 (D.A.S.
and Thomas M. Wilkie, unpublished data). Because the RGS domain of mouse RGSZ1 is
more than 80% identical to that of bovine Ret
RGS, this probe would have reported two sites if RGSZ1 and Ret RGS were different
genes separated by more than 106 base pairs. These data suggested that RGSZ1 and Ret
RGS are actually products of a single gene. We therefore used RT-PCR to confirm that the
two mRNAs each contained sequence identical to each other in shared regions and iden-
tical to the genomic DNA at this single locus. With this confirmation, we proceeded to
assemble the DNA sequence of the common
gene.

We used coding sequence from these
BACs to search for additional human genomic DNA fragments and ESTs that con-
tained RGSZ1/Ret RGS sequence. We combined the resulting DNA fragments to deline-
eate seven exons of RGS20, which are either identical in their coding regions to the previ-
ously identified human RGSZ1 cDNA or similar (> 70% identical) to the bovine Ret RGS
cDNA. All exons begin with canonical splice
acceptor sites (AG) and end with splice donor
sites (GT), and adjacent introns end with
pyrimidine-rich regions. Exons A and B, with
sequence similar to the 5' coding region of
bovine Ret RGS, are 5' of the RGSZ1 exons
(Figs. 1A and 1B) and their coding sequence
is oriented in the same direction. We found no
other genomic sequences similar to RGSZ1 or
Ret RGS. Further, we found no other genomic
sequence within 1 cM of
RGS20 that encodes
any RGS domain, supporting the idea that the
Ret RGS-like exons and RGSZ1 exons consti-
tute a single gene.

These data indicate that Ret RGS and
RGSZ1 are two splice variants of a single
gene, now named RGS20. It is also possible that some
mRNAs arise from multiple transcriptional initiation sites. To assemble the entire RGS20 sequence, all fragments iden-
tified to this point were overlayed onto the Celera
Chromosome 8 human genome assembly
GA_x54KRCCD58L to generate the ~ 107-kb
RGS20 structure (Figs. 1A and 1B).

RGSZ1 Is Encoded by Exons C-G
The 5' end of the RGSZ1 ORF is encoded by exon C. It contains
the most 5' ATG codon, which was not identified in the
originally cloned RGSZ1 cDNA [14,15]. This extension of the ORF adds a fifth, closely spaced, potential translational start site to the four originally described for human RGSZ1 [14]. Notably, initiation at the first of these ATG codons would produce a protein 24 amino acid residues longer than that originally described. We have not detected an in-frame stop codon 5′ of the first ATG in human RGSZ1 cDNA, but there is an in-frame stop codon 30 bp 5′ of this ATG in mouse RGSZ1 cDNA (NM_021374). The additional 5′ cDNA sequence reported here for exon C (Fig. 1C) was derived from human amygdala cDNA and agrees with sequence of four independent genomic DNA fragments and five ESTs (BE389290, BF110442, AL525981, AL546580, and AL545424 from brain, placenta, and lung and endometrial carcinomas).

The N-terminal sequence of RGSZ1 protein is fairly well conserved among mice, pigs, and humans, but the presence of these multiple potential start sites varies. Mice replace the second and third methionines with glutamic acid and threonine, but have an additional methionine in place of an arginine. Another in-frame methionine codon is polymorphic in mice, methionine in 129/B6 but arginine in Balb/C (Fig. 1E). Pigs also lack the second of the human in-frame ATG codons according to the sequence of one EST (Fig. 1E). Overall, the RGSZ1 transcript may contain between three and five translational start sites, which allows synthesis of proteins with Mr = 23,300–27,000. This size range is consistent with the molecular weight heterogeneity detected in RGSZ1 purified from bovine brain (21,000–29,000). Although some of the size heterogeneity of the brain RGSZ1 may reflect palmitoylation [11], phosphorylation [21,22], or other modification, it is likely that multiple start sites are used in the various cell types present in the brain. Functional differences among these N-terminal variants remain to be investigated.

Exons D–G contain the coding region of the originally described RGSZ1 cDNA [14,15]. Exons D and E encode most of the N-terminal region, which is required for binding to phospholipid bilayers [14,19], including a cysteine string that is probably multiply palmitoylated [11]. Exons F and G encode the RGS domain, the site of GAP activity. Exon G also includes the complete 3′ UTR of human RGSZ1 mRNA, including the polyadenylation site. Exon G is thus the most 3′ exon of RGS20.

Ret RGS and Novel RGS20 mRNAs

As first described, bovine Ret RGS differed from human RGSZ1 and GAIP most notably in having a 155-amino-acid, hydrophobic N-terminal domain before the apparent N terminus of the other two proteins [20]. The cDNA for human RGS20 mRNAs. (A) RGS20 splice products identified in human mRNA. The upper bar, shaded as in Fig. 1, shows the exons of RGS20 (without breaks) and the positions of the primers (Table 1) used to amplify human cDNA from retina or amygdala (arrows). The lower bars represent the complete or partial exons sequenced from the RT-PCR products. Each cDNA was amplified at least three times from two RNA preparations. The inset shows the upper three cDNAs plus the ~1.1-kb cDNA resolved on an ethidium-stained gel. The 755-bp product is RGSZ1 (exons CDEFG). (B) Predicted amino acid sequences of the proteins encoded by RGS20 mRNAs. The exons that contribute to each mRNA are shown at left. Where mRNAs contain multiple nearby ATGs, it is not known which translational initiation sites are used, or with what frequency.
Ret RGS, which we isolated from a retinal cDNA library, contains a similar extension that is encoded by exons A and B. Exon A includes at least 93 bp of the 5’ UTR. The combined exon AB ORF (Fig. 1C) is 70% identical to the corresponding region of bovine Ret RGS in predicted amino acid sequence (Fig. 1D). The human Ret RGS cDNA also includes two insertions compared with the bovine cDNA, one of nine codons and one of six codons. (The first 630 bp of the DNA sequence reported for the 5’ UTR of bovine Ret RGS [20] is almost identical to genomic DNA reported to lie on human chromosome 1p36.1 (AL139426), and is therefore apparently unrelated to human RGS20.)

RT-PCR using mRNA from human retina indicates that exons A and/or B are spliced in various combinations to the more 3’ exons to form several apparently mature mRNA species (Fig. 2). Four distinct RGS20 cDNAs were found that include exon A plus downstream exons: two major PCR products at 821 and 672 bp, and two more poorly amplified products at 1100 and 589 bp (Fig. 2A).

The 821-bp PCR product, amplified from human and mouse retinal mRNA, corresponded to exons ADEFG, with the predicted amino acid sequence shown in Fig. 2B. The protein product of this mRNA contains the originally described RGSZ1 sequence [14] plus an additional 55 amino acids at the N terminus. This addition is relatively hydrophilic and contains no obvious functional motifs.

The second cDNA, 672 bp in length, consists of exons AEFG (Fig. 2). It thus encodes a protein that contains the last 168 amino acid residues of RGSZ1 linked to an alternative 12-amino-acid N terminus encoded by exon A (Fig. 1C). cDNA composed of exons AEFG was also found in mouse eye. The predicted translation product of this mRNA is distinctive in several respects. First, it lacks the hydrophilic region of the RGSZ1 N-terminal domain, including the cysteine string, and would therefore be predicted to be soluble unless lipid-modified or bound to some other membrane protein. Second, the portion of the ORF contributed by exon A is a different reading frame than that used in Ret RGS and in the ADEFG mRNA, which means that the N terminus is also distinct. Third, this mRNA encodes only one plausible translation initiation site.

The third cDNA, of 589 bp, was much less abundant than the first two. It consists of exons AFG, and thus encodes a short RGS protein that consists of an 11-amino-acid N-terminal region, the RGS domain beginning with helix 1 [23], and a short carboxy-terminal domain. This protein is therefore quite small and soluble, similar to RGS10. The fourth cDNA, which could be amplified only inconsistently but was prepared from several batches of mRNA, was 1100 bp. Its 5’ end includes exons A, D, and E, but we were unable to prepare enough to clone or sequence it completely. Its identity and importance are thus uncertain.

Although we cloned the human Ret RGS cDNA from a λ-phage library, we were unable to amplify the full-length cDNA by RT-PCR using exon A sense primers. However, we did amplify a 711-bp product that contained exons BDEFG (using sense primer S2; Table 1). This product may be a fragment of the Ret RGS cDNA (exons ABDEFG), which would suggest that the full-length Ret RGS mRNA is simply difficult to detect by RT-PCR. Using cDNA from human amygdala, we also detected RGSZ1 cDNA, a 755-bp product encoded by exons C–G, as well as a 606-bp product, CEGF, that does not contain an ORF and which we did not pursue further.

### Anatomic Expression of the RGS20 mRNAs

Because retinal cDNA yielded only exon-A-containing transcripts and brain cDNA yielded only exon-C-containing transcripts, we carried out northern blots with exon-specific probes to clarify the transcripts expressed in each tissue. When northern blots were hybridized with an exon FG probe (RGS box), the major mRNA in amygdala was a broad band centered at 1600 bp (Fig. 3A) as reported previously [14]. This band was also detected with an exon C probe and is of the correct size to represent the RGSZ1 mRNA. The major transcript from human retina, about 2100 bp, was detected with probes based on exons FG, B, and (more weakly) A (Fig. 3A). It corresponds to the length of the human Ret RGS cDNA, although we cannot be sure that we isolated the entire 5’ UTR. The exon FG probe also labeled a faint band in retinal mRNA that was slightly smaller than the RGSZ1 mRNA (~1450 bp). We could not detect this species with the other probes. It may represent another splice variant shown in Fig. 2.

To clarify the anatomic distribution of RGS20 expression, we used northern blot analysis of RNA prepared from multiple mouse tissues (Fig. 3B). The major RGS20 mRNA expressed in mouse brain is the 1600–1700 bp RGSZ1 cDNA [14]. Its expression is particularly high in the olfactory bulb (Fig. 3B, lane 13), where a second species of about 3000 bp and a minor band at 3500 bp were also obvious. On separate northern blots, mRNA prepared from cortex, cerebellum, medulla, and midbrain also showed levels of the 1700-bp and 3000-bp bands comparable to those seen in olfactory bulb (data not shown). Only trace amounts, if any, of either transcript were observed in non-cerebral tissues except for eye or, at low levels, female sex organs. RNA from eye displayed these same mRNAs plus two more of approximately equal intensity at 2400 and 3900 bp (lane 14). The larger size may

### TABLE 1: Primers used to amplify segments of human RGS20 DNA

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>ATATATGAATCCACATGGCCCCAGCTTTCCCAAG</td>
</tr>
<tr>
<td>S2</td>
<td>GAAGAAGACGCCACCCCGT</td>
</tr>
<tr>
<td>S3</td>
<td>ATATATCTGACGGCTGCAATATGGAGAAGGAC</td>
</tr>
<tr>
<td>S4</td>
<td>CAGGATCCATGGAGTACAGACGGATGGAG</td>
</tr>
<tr>
<td>AS1</td>
<td>ATATATGTAACCTATGCTTCAATAGATTTCTCCG</td>
</tr>
<tr>
<td>AS2</td>
<td>CTGGGCGGCGGCGCATCTGC</td>
</tr>
<tr>
<td>AS3</td>
<td>GTGTCCTGGCGGCGGCGGCG</td>
</tr>
<tr>
<td>AS4</td>
<td>CAGCGGTGCGGCGTCTCTC</td>
</tr>
</tbody>
</table>

Copyright © 2001 by Academic Press. All rights of reproduction in any form reserved.
indicate the presence of still-unidentified mouse 5' UTR. These mRNAs may also reflect the use of alternative transcriptional start sites 5' of that used to initiate RGSZ1 mRNA.

The data described here indicate that at least two previously described RGS proteins are both derived from a single gene, RGS20. RGSZ1 is expressed at low levels in brain and in a few other tissues, and derives from a single mRNA encoded by exons C–G. RGSZ1 mRNA contains up to five nearby ATG codons, and multiple translational start sites provide a likely explanation for at least some of the size heterogeneity of purified brain RGSZ1. Based on studies of both natural and recombinant protein [14,18,19], this RGSZ1 gene product is a strongly hydrophobic G-protein GAP that is firmly membrane-bound. The N-terminal extension of 23 amino acid residues encoded by exon C contains no recognizable functional motif.

In retina, the Ret RGS protein is also a tightly membrane-bound GAP [18], and the role of the additional large N-terminal domain encoded by exons A and B is unclear. It may span the membrane [20], and protein–protein association might alter the behavior of the RGS domain. In the case of both RGSZ1 and Ret RGS, as is true of all the mRNAs that include transcripts of exon D, the palmitoylated cysteine string will contribute to hydrophobicity and may target the protein to secretory vesicles [11]. Based on studies with recombinant proteins [19], the newly recognized products of the AEFG and AFG transcripts encode soluble proteins that have the potential to be recruited to their membrane-bound targets in response to yet unknown partners. The regulatory inputs that select for the production of each mature mRNA will thus modulate how and with what affinity these GAPs function in cells.

**MATERIALS AND METHODS**

**Analysis and assembly of the RGS20 gene.** Sequence from human RGSZ1 (NM_003702) and bovine Ret RGS (U89254) cDNA was used to search NCBI and Celera databases using BLASTN and TBLASTN programs. For initial assembly of the RGS20 gene, we used human genomic DNA that was at least 95% identical to human RGSZ1 or at least 70% identical to bovine Ret RGS throughout the sequence. An incompletely sequenced human genomic BAC (clone RP11-421C10; AC023132.4) was found to contain DNA sequence corresponding to both the extended 5' region of Ret RGS cDNA and the first two previously identified exons of RGSZ1. Sequences from this BAC were used to search for other DNA sequences; over 110 fragments of genomic DNA that contained RGS20 sequence were found. We then assembled the human RGS20 gene based on Celera genome assembly scaffolds GA_x54KRCDDL 3.0-3.5 (bases 457517-500,000) and 3.5-3.64 (bases 1-64750), and used the following fragments to fill gaps: GA_24226608, GA_29259491, GA_23050835, GA_31692761, GA_30468882, GA_25349490, GA_23933044, GA_990379, GA_19895268, GA_28120294, GA_7060888, GA_35860658, GA_22535561, GA_13152016, GA_2463065, GA_24618825, GA_25713992, GA_35068144, GA_13086250, GA_12439514, GA_14988175, GA_15302579, BAC RP-11421C10 (AC023132.4 bases 2983-4543, 6058-7685, 9375-11213, 11314-12577, 12678-14655, 14756-17774, 24895-28615, 33265-37827, 71758-79804, 79905-87292, 101594-110504, 110605-123442, 152791-169160-169261-186107) and BAC RP-11 410C17 (AC079335.1 bases 3363-3383); GSSs B44338 and AQ281160; and ESTs BE389290 and BF110442 (Fig. 1B). Contigs shown as Celera fragments have the following accession numbers (Fig. 1B): 1, GA_25713992; 2, GA_30468882; 3, GA_23050835, GA_31692761; 4, GA_15302579; 5, GA_24618825; 6, GA_35068144, 7, GA_19352709, GA_13666439, GA_18935380; 8, GA_35090155, GA_25349490, GA_19610023, GA_35860658.
the coding region and 3'UTR of mouse RGSZ1 (bp 51–999, NM_021374; exons CDEFG). The 821-, 672-, and 589-bp products generated with the S1-AS1 primer pair, the 1100-bp product from the S1-AS1 primer pair, and the 1423-bp product were amplified bi-directionally for human opioid-receptor-like gene ORL1 and its 5'UTR of mouse RGSZ1 (bp 51–999, NM_021374; exons CDEFG). The 821-, 672-, and 589-bp products generated with the S1-AS1 primer pair were all cloned into pBluescript. The 711-bp product generated with the S2-AS1 primer pair, the 1100-bp product from the S1-AS1 primer pair, and the 606-bp and 755-bp products from the S3-AS1 primer pair were sequenced directly. Longer PCR products were not obtained by increasing the time of the extension reaction. The full-length cDNA that corresponds to human RGSZ1 [20] was isolated from a human retinal cDNA library in 1 Digit (Strategene; five isolates, each unique but overlapping and all encoding the RGSZ1 box). Polymorphism in the 5' coding region was confirmed by sequencing cDNAs from two commercial Balb/C libraries and from an RT-PCR amplification of 129/b6 brain RNA.

ACKNOWLEDGMENTS

We thank Waw-Heng Wang and lok-Hou Pang, Alcon Research, Ltd., for human retinas and samples of cDNA; Jeremy Nathans (Johns Hopkins University) for the human retina cDNA library; and Thomas Wilke for discussions, advice, and support. This study was supported by U.S.P.H.S. grant GM30335 and postdoctoral fellowship GM20576, and R. A. Welch Foundation grant l-0892.

RECEIVED FOR PUBLICATION MARCH 28; ACCEPTED SEPTEMBER 21, 2001.

REFERENCES
