Mol Reprod Dev. Author manuscript; available in PMC 2017 January 17.

Published in final edited form as:
Mol Reprod Dev. 2008 October ; 75(10): 1578–1589. doi:10.1002/mrd.20896.

A Novel Group of Multi-GAP-Domain Proteins

YIBING RUAN, **CHRISTOPH W. SENSEN**, and **FRANS A. VAN DER HOORN**[*]
Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada

Abstract

The Rho GTPase-activating proteins (RhoGAPs) play an essential role in regulating various cellular processes. Rat tGAP1 is the first reported protein that has multiple GAP domains. It is exclusively expressed in male germ cells. However, tGAP1 does not possess GAP activities in vitro. No tGAP1 homology has been identified in other species. In this study, we searched the genomic databases and identified many genes whose protein products possess 2–4 GAP domains in rat, mouse and dog. These genes all showed sequence similarity to tGAP1. The rat tGAP gene loci all locate on chromosome 2 and are all expressed in testes in RT-PCR analysis. The mouse tGAP gene loci also clustered on chromosome 3 but RT-PCR analysis showed most are pseudogene loci. Multiple sequence alignment showed that many conserved residues of the "arginine finger" motif within the GAP domains of predicted tGAP proteins have mutated, suggesting that tGAP proteins do not possess GAP activity. We also elucidated the evolutionary relations among the rat tGAP genes. Based on the phylogenetic analysis data, we proposed that tGAP genes and Arhgap20 genes have a common ancestor.

Keywords

spermatogenesis; RhoGAP proteins; pseudogenes; phylogenetic analysis

INTRODUCTION

The Rho GTPase-activating protein (RhoGAP) family of proteins is defined by the presence of a 150-amino-acid RhoGAP domain, which is necessary for enhancing the intrinsic hydrolytic activity of Rho GTPase (Peck et al., 2002). To date, a large number of RhoGAP proteins have been identified. RhoGAP proteins are important regulators of the Rho GTPases (Tcherkezian and Lamarche-Vane, 2007). The Rho GTPases operate as molecular switches that cycle between an active GTP-bound and an inactive GDP-bound form. When in GTP-bound form, Rho GTPases serve as regulators of a wide variety of biological processes, such as cell dynamics, gene transcriptions, cell cycle regulation, and membrane trafficking (Jaffe and Hall, 2005). RhoGAPs regulate these processes by placing an "arginine finger" motif into the phosphate-binding pocket of Rho GTPases and drastically increasing their GTP-hydrolytic activity, hence quickly turning Rho GTPases into the inactive form

[*]Correspondence to: Frans A. van der Hoorn, 3330 Hospital Dr NW, Calgary, Alberta, Canada T2N 4N1. fvdhoorn@ucalgary.ca.
This article contains supplementary material, which may be viewed at the Molecular Reproduction and Development website at http://www.interscience.wiley.com/jpages/1040-452x/suppmat/index.html.

(Moon and Zheng, 2003). Mutation of the conserved arginine residue reduces the enzymatic activity, but has no significant effect on the binding of the GAP protein to its substrate (Rittinger et al., 1997). However, not all RhoGAP domain-bearing proteins possess GAP activity toward GTPases. For instance, the p85 regulatory subunit of PI3-kinase binds specifically to Cdc42 and Rac1 in a GTP-dependent manner, but does not display any catalytic activity (Zheng et al., 1994).

Little is known about the role of Rho GTPases and RhoGAPs in male germ cells (Kusano et al., 2001; Quimby and Dasso, 2003). The three best characterized Rho GTPases, RhoA, Rac1, and Cdc42, are not expressed in spermatocytes and early spermatids (Naud et al., 2003). On the other hand, Rnd2, a Rho family GTPase-deficient member, is found in the mid-zone of meiotic cells, and in the Golgi-derived pro-acrosomal vesicle (Naud et al., 2003). Rnd2 binds MgcRacGAP, which was observed in human and mouse male germ cells and associates with the plasma membrane. It also binds an anion transporter Tat1 in spermatocytes and spermatids (Toure et al., 2001; Naud et al., 2003). Moreover, Ran GTPases, a member of small GTPases superfamily, have been shown to be intimately involved in male germ cell development (Kierszenbaum et al., 2002). Ran and its regulators, RanGEF and RanGAP, have been well characterized in nuclear transport and mitotic spindle assembly (Steggerda and Paschal, 2002). A mutant *Drosophila* protein SD-RanGAP, which lacks one of the nuclear export signals (NES) but retains enzymatic activity, is mislocalized to nuclei. This form decreases RanGTP levels, and causes transmission distortion in male flies (Kusano et al., 2001). Moreover, overexpression of RanGAP in males results in the same phenotype. It was shown that sperm dysfunction can result from defects in nuclear transport (Kusano et al., 2002). These findings demonstrated that in male germ cells the GAP proteins can assume unique, specific roles.

We have isolated a novel rat testis-specific gene called tGAP1 (Modarressi et al., 2004). The tGAP1 gene is expressed in spermatocytes and spermatids, with low but detectable expression in spermatogonia. The tGAP1 protein localizes to the spermatocyte cytoplasm and plasma membrane and the cytoplasm and nucleus of round and elongating spermatids. Strikingly, tGAP1 contains four ARHGAP20 domains, a subtype of Rho-GAP domains (Modarressi et al., 2004). To date, no other RhoGAP proteins have been found to have more than one RhoGAP domain (Tcherkezian and Lamarche-Vane, 2007). tGAP1 does not activate RhoA, rac1 or cdc42 in vitro (Modarressi et al., 2004). However, expression of tGAP1 in some somatic cells results in apoptosis (Modarressi et al., 2004).

The ARHGAP20 domain was first described in human Arhgap20 protein, which is also known as KIAA1391 and RA-RhoGAP (Katoh and Katoh, 2003; Curry et al., 2004; Yamada et al., 2005). Human Arhgap20 protein contains a RhoGAP, an RA (Ras association), and a low-similarity Pleckstrin-homology (PH) domain. The Arhgap20 protein is expressed in brain, liver and male germ cells (Curry et al., 2004). Mouse and rat orthologs of human Arhgap20 genes have also been characterized (Katoh and Katoh, 2003). The RhoGAP domain is active toward RhoA, but not Rac1 or Cdc42 (Yamada et al., 2005). Interestingly, the GAP activity of the RhoGAP domain is regulated by the RA domain, which binds to the GTP-bound active form of Rap1, a member of GTPase superfamily, and enhances its GAP activity (Yamada et al., 2005). Arhgap20 is involved in Rap1-induced neurite outgrowth,

mainly through down-regulating RhoA (Yamada et al., 2005). The function of Arhgap20 in male germ cells remains unknown.

In this study, we have undertaken a comprehensive analysis of publicly accessible databases and discovered several new tGAP1-like genes, which appear to contain more than one RhoGAP domain. All of the rat tGAP genes are expressed in testes, suggesting a unique function. Using multiple alignment and phylogenetic analysis, we were able to characterize several unique and intriguing features of tGAP proteins.

MATERIALS AND METHODS

Sequence Searches and Mapping Data Retrieval

The nucleotide sequence of tGAP1 (GI:76443659) was used in BLAST searches (Altschul et al., 1997) at the NCBI web site (http://www.ncbi.nlm.nih.gov/blast/) against the rat genome database available as of Jan 3, 2007 (RGSC v3.4). Genomic sequences with significant similarity to tGAP1 were downloaded and initially analyzed using the OMIGA software (Oxford Molecular Ltd.). Sequences of hypothetical rat tGAP genes were either retrieved from the NCBI web site, when available, or derived by comparing the genomic sequences of tGAP genes with the tGAP1 mRNA-based sequence. Gene loci and mapping data were retrieved using the NCBI MapViewer (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=10116). As a complementary approach, a 94-amino-acid sequence between the first and second ARHGAP20 domain of tGAP1 was used in BLASTP searches on thenonredundant databases as of April 2007.

A consensus of the amino acid sequences of the first ARHGAP20 domains (GAP-D1) from all rat tGAP genes was generated using ClustalW (http://www.ebi.ac.uk/clustalw/). The consensus sequence (LFGRELSSICQDGBLPSAILDMLSLLKRKGPTTEGVFLIRPSITLCQTIKBKLDSGEEVDINKQSVHVVAWIFKDFLQNIEGSLMTSKLYDEWIAVTEKVBDEEKLAAVQSLLDKLPPENAALLRQLFRILYEIKSNSSVNEMSSYHLSVGIAPCLLFLPSYCNNGLTNDIAKKISLVTFMIEN) was used in BLAST searches on all genomic databases available as of March 5, 2007. The search parameters were set as following: E value = 0.1; filter = none; program = tblastn. The search results were further analyzed using the OMIGA software. Hypothetical mRNA sequences of tGAP genes from other species were either retrieved from the NCBI web site, when available, or predicted by the GENSCAN program (http://genes.mit.edu/genscan.html).

RNA Isolation

Total RNA was isolated from 100 mg of selected rat tissues or C57BL/6 mouse testes. Tissue was homogenized at 4°C in 1 ml of TRIzol (Invitrogen Canada, Ontario, Canada) to isolate RNA following the manufacturer's instructions. Contaminating DNA was removed by incubation of RNA preparations with 10 μg/ml of RNasefree DNase I in conditions suggested by the manufacturer (GE Healthcare, Piscatway, NJ).

RT-PCR

RNA was transcribed into cDNA using the M-MLV RT kit (Invitrogen); for each tissue, 5 μg of RNA was used with 100 ng of random decamers to generate cDNA. Primers specific for each gene or locus were synthesized at the University of Calgary Core DNA Services (All primer sequences are listed in Table 1). For each primer set, 30 cycles of PCR were carried out under the following conditions: a denaturing step for 20 s at 94°C, an annealing step for 30 s at the temperature specific for each primer pair, and an elongation step for 1 min at 72°C. PCR fragments were sequenced at the University of Calgary Core DNA Services.

Genomic DNA Isolation and PCR

Mouse genomic DNA was isolated from mouse tails as described by Sambrook and Russell (2001). The PCR was performed using Expand Long Template PCR System (Roche Applied Science, Quebec, Canada) with the conditions recommended by the manufacturer.

Multiple Sequence Alignments

All amino acid alignments were performed using the program T_COFFEE (Notredame et al., 2000) with default settings. The alignments were then adjusted and shaded using the multiple sequence alignment editor GENEDOC (Nicholas et al., 1997).

Phylogenetic Analysis

We used PHYLIP version 3.66 (Felsenstein, 1989) for our phylogenetic analysis. Maximum likelihood, distance, and parsimony analyses were performed using the protein alignment as input. Bootstrap values were obtained using SEQBOOT and creating 1,000 delete-half Jackknife data sets. The maximum likelihood analysis was performed using PROML with the “S” and “G” options invoked. The distance analysis was performed by using PROTDIST and subsequently NEIGHBOR with standard parameters, and the parsimony analysis was performed using PROTPARS with standard parameters. In all cases, the “M” option for the analysis of the multiple data sets created with SEQ-BOOT was invoked.

RESULTS

The Rat tGAP Gene Family

Rat tGAP1 protein contains four ARHGAP20 domains. (For precise description, the four domains will be referred to as GAP-D1, GAP-D2, GAP-D3, and GAP-D4, respectively.) To analyze whether tGAP1 is the only multi-RhoGAP domain protein in the rat genome, we performed BLAST searches against rat genomic databases using both tGAP1 mRNA and protein sequences. We identified a list of potential tGAP gene loci. Surprisingly, all significant blast hits clustered within a 150-Mb region on rat chromosome 2. By further analyzing these potential loci using both bioinformatics and biochemical approaches, we discovered six loci which contain genes that are homologous to tGAP1. The products of these genes all consist of two or more ARHGAP20 domains (Fig. 1). We named the six loci tGAP1 to tGAP6, respectively. The tGAP1, tGAP3, tGAP4, and tGAP6 loci all contain one tGAP gene, whereas the tGAP2, and tGAP5 loci code for more than one tGAP-like gene (Fig. 1A). These genes are named by adding a character after the name of the gene locus.

For instance, tGAP2a and tGAP2b are two separate genes present in the tGAP2 locus. In addition, tGAP2a and tGAP5b appear to have two copies on rat chromosome 2 (Fig. 1A). We investigated whether other genes proximal to the rat tGAP genes are also present in multiple copies. A few genes have 2–5 copies in this region. However, no genes have replications like tGAP genes.

The expression of tGAP genes was analyzed by RT-PCR. Similar to tGAP1, all investigated tGAP genes are expressed in rat testis (Fig. 1B). We also designed primers for the 80.9, 138.2, 154, and 154.2 M loci. However, no gene expression could be detected in RT-PCR (data not shown). We analyzed the rest loci in Figure 1A with bioinformatics approaches and found these loci could either be pseudogenes or noncoding transcripts.

Mouse tGAP Exist as a Single Gene and Many Pseudogenes

To identify tGAP1 homologs in other species, we first used BLAST to search the mouse and human genomic databases using the rat tGAP1 mRNA sequence. However, these searches only returned nonsignificant hits, which imply that there may be no tGAP1 homologs in mouse and human. To further investigate whether tGAP genes are exclusive to rat, we generated a consensus amino acid sequence of the first ARHGAP20 domain (consensus D1) using ClustalW and searched all available genomic databases for all genes bearing regions similar to GAP-D1 domain. The results are summarized in Table 2. In all the genomic databases searched, only the mouse genome yielded a number of blast hits comparable to that of the rat genome. After detailed analyses, these hits were found to represent three types of genes: tGAP-like genes that have multi-Arhgap20 domains; Arhgap20 genes that bear one ARHGAP20 domain flanked by a PH and/or RA domain; and T-cell activation RhoGAP genes that possess one ARHGAP20 domain and no other recognized conserved domains. Interestingly, all tGAP-like genes are clustered on mouse chromosome 3 in a region syntenic to the rat tGAP clusters (Fig. 2A). Moreover, most of these genes are clustered in a 22 Mb chromosomal region. Surprisingly, unlike the rat tGAP gene clusters, only two of seven selected potential tGAP gene loci showed their expression in mouse testis with RT-PCR analysis (Fig. 2B). These two loci (124.8 and 29.1 M) already had entries in NCBI protein database (XP_995294 for 124.8 M and BAE21461 for 29.1 M). We propose to name the XP_995294 gene as mouse tGAP (mmtGAP), because it encodes a 544-amino acid protein with two individual ARHGAP20 domains. The BAE21461 gene product is highly similar to rat tGAP genes up to the sequence before the GAP-D2 domain, but only bears one ARH-GAP20 domain.

To analyze the potential reading frame encoded by the gene on these loci, we then submitted all mouse tGAP-like gene loci to the gene prediction program GENSCAN (http://genes.mit.edu/genscan.html). With the exception of BAE21461 and mmtGAP, which were predicted to encode a full length protein, all the other loci rendered short peptides (less than 300 amino acids) with an individual or incomplete ARHGAP20 domain. This suggests that these gene loci could be either pseudo-genes or noncoding transcripts. The latter is exemplified by two RIKEN full-length cDNAs (AK029958 and AK030119), which correspond to the 35.8 and 49.4 M loci, respectively. The cDNAs both contain several frameshifts and premature stop codons that render the mRNA nontranslatable.

The rat homolog of mmtGAP was located on rat chromosome 2 in two copies on 2q34 (205 M) and 2q42 (221 M) (Fig. 1A). Two near-identical mRNAs (XM_001075667 and XM_227675 respective for each copy, with the latter having six more nucleotides) were predicted by the GNOMON program (NCBI). Our primers spanning the first ARHGAP20 domain of these two mRNAs did amplify a DNA fragment in the RT-PCR. However, sequencing of this fragment revealed that 190 bp of the predicted ARHGAP20 domain, representing one-third of the domain, are missing from the mRNA sequences. Furthermore, immediately upstream of the missing part, a stop codon is present (unpublished data). This suggests that the rat homolog of mmtGAP could be an expressed pseudogene or a noncoding transcript.

tGAP Genes in Other Species

For most of the species we have searched, we did not find tGAP-like genes, but were able to identify Arhgap20 and/or T-cell activation RhoGAP genes (Table 2). However, we found the potential presence of tGAP genes in dog (NW_876267 and NW_879505), bovine (NW_001502124, NW_001501897, and NW_001494399), and horse (AAWR010026) genomes. Dog NW_876267 has been predicted to express an 853-amino acid protein with two ARHGAP20 domains (XP_851810), supported by EST evidence. We conclude that tGAP genes, though mainly expressed in rat, are also expressed in mouse and dog and might be expressed in bovine and horse. However, we have not been able to find tGAP genes in primates including Rhesus monkey, chimpanzee, and human.

Protein Sequence Alignments of tGAPs

The most striking feature of the tGAP proteins is their acquisition of multiple ARHGAP20 domains. Yet, the ARHGAP20 domains within one protein are not just copies of one another. In the case of rat tGAP1, the first ARHGAP20 domain (GAP-D1) is substantially different from the other three (GAP-D1 and GAP-D2 only share 35% identities), whereas GAP-D2, -D3, and -D4 are highly similar amongst each other. Indeed, the first ARHGAP20 domains of all tGAP proteins are best conserved (See Table 3).

There is an RA (RalGDS/AF-6) motif with weak similarities in the N-terminus of tGAP1. This motif is also present in all tGAP5 genes, dog tGAP, and partially in tGAP4. Surprisingly we found that although none of tGAP2a, tGAP2b, and tGAP3 genes were predicted by the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/structure/cdd/cdd.shtml) to have an RA motif, all do in fact have a region highly similar to the tGAP1 RA motif (Fig. 3).

Amino acid sequences were gathered from all tGAP proteins, mouse BAE21461 protein and Arhgap20 proteins from human, chimpanzee, mouse, rat, bull, dog and frog. Multiple sequence alignment of the protein sequences showed four conserved regions among tGAP proteins (Fig. 4). Region I consists of the RA domain. As mentioned above, although mmtGAP and tGAP6 do not contain this region, the other tGAP proteins show conservation of both the RA sequences, and the flanking sequences. The most conserved region of tGAP proteins, the GAP-D1, falls into what we have defined as region II. This is also the only conserved region among all tGAP, Arhgap20 and T-cell activation RhoGAPs (Supplement

Fig. 1). The GAP-D1 of tGAP proteins may fold as the tertiary structure of RhoGAP domain, because of the high sequence conservation. We also submitted the GAP-D1 of tGAP proteins to secondary structure prediction programs such as NNPREDICT (http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html) and APSSP (http://imtech.res.in/raghava/apssp/), which both returned a structural pattern highly similar to that of the Cdc42GAP domain (unpublished data). Region III includes the sequence between GAP-D1 and GAP-D2. This region from rat tGAPs is conserved in the mouse BAE21461 protein, but not in mmtGAP and dog tGAP. Interestingly, a serine and a threonine residue in this region are conserved in all tGAP proteins listed (Fig. 4C). These two residues, however, are not potential phosphorylation sites as predicted by NetPhos (www.cbs.dtu.dk/services/netphos/), a program that predicts serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins. Nonetheless, a “GLSSCP” motif between these two residues that is only conserved among rat tGAP proteins is a possible serine phosphorylation site as suggested by NetPhos. The GAP-D2 domain of tGAPs forms the conserved region IV. Although several gaps were introduced in the alignment because of the variations in amino acid sequences, most polar amino acids and aliphatic amino acids are well conserved.

The multiple sequence alignment also revealed that the Arhgap20 proteins, although substantially different from tGAP proteins, share certain similarity with the tGAP proteins from N-terminus to the end of GAP-D1 (Supplement Fig. 1). This implies that Arhgap20 and tGAP proteins might have a common ancestor. Comparing the GAP-D1 sequences with the ARHGAP20 domain from Arhgap20 proteins, we noticed that the “arginine finger,” which is crucial for the catalytic function of RhoGAPs and is well conserved among all functional RhoGAPs, is substituted by Leu, Ile, or Thr, in the GAP-D1 of all rat tGAP proteins and BAE21461 protein. The mouse and dog tGAP proteins, however, retained this crucial arginine residue. We further discovered that in addition to the substituted arginine finger, some residues in and around the shallow pocket formed between the A-A1 loop and the B and F helices that are crucial for stabilization of the GTPase and conserved in the Arhgap20 proteins, are also substituted to other amino acids (Fig. 5). Therefore, this suggests that GAP-D1 is most likely a nonfunctional GAP domain.

Phylogenetic Analysis of Rat tGAP Genes

Phylogenetic trees were generated from the amino acid sequences of tGAP genes using the maximum likelihood, distance and parsimony methods with statistical confidence measured by bootstrap analysis (Supplement Fig. 2). In both maximum likelihood and distance trees, the rat tGAP genes formed one large group, which separates them from tGAP genes from mouse and dog. Within the rat tGAP genes, they fall into four distinct groups: tGAP1 and tGAP3, tGAP2 and tGAP4 genes, tGAP5 genes, and tGAP6 gene. The branch structure in the parsimony tree is practically the same as the distance and maximum likelihood trees, although a few genes fall into different groups (e.g., mmtGAP and tGAP2a; Fig. 6). Comparison of the three phylogenetic trees highlights the similarity of tGAP5 genes, which grouped together in all cases, and the slightly more divergent, but nonetheless related, nature of other rat tGAP genes.

DISCUSSION

Very little is known about RhoGAPs during spermatogenesis. Previously, we discovered tGAP1, a multi-GAP domain protein in rat testis (Modarressi et al., 2004). In this report, we have characterized tGAP1 as a member of a gene family that shares the unique feature of multiple ARHGAP20 domains. The 10 tGAP genes present in rat have two, three, or four GAP domains. To date, no other RhoGAPs were shown to possess more than one GAP domain. We also discovered that all tGAP genes are expressed in rat testis (Fig. 1). In addition, some tGAP genes are also expressed in epididymis and skeletal muscle (unpublished data). This distinct expression pattern suggests that the tGAP gene family has a special function during rat male germ cell development.

While the rat genome database is still being completed, quite a few gaps exist within the genomic sequences and we were not able to predict the full sequence for some of the tGAP genes with confidence, including tGAP2b and tGAP4. Although we cannot rule out the possibility that some tGAP genes could be pseudogenes, we deem it unlikely because gene prediction programs, including GENSCAN and TWINSCAN, predict the presence of tGAP mRNAs with long open reading frames in these loci, in contrast to the pseudogene loci, which are all predicted to encode several short open reading frames, often with an incomplete GAP domain. Whether these pseudogene loci are transcribed remains to be determined. We have shown that the rat homolog of mmtGAP at 205 and 221 M can be expressed as a noncoding transcript.

A large number of tGAP gene loci have been discovered in the mouse genome, which show similarity with rat tGAP genes and display a similar clustering pattern on a chromosomal region syntenic to the tGAP loci in rat. However, as described in the previous section, most of them could be pseudogenes or noncoding transcripts. Two noncoding transcripts in this region have been cloned (AK029958 and AK030119). Whether they function as regulators of other RNA species remains to be investigated. In our selected gene loci, we could only identify one significant tGAP gene candidate in mouse. The tGAP genes in both rat and mouse are probably tandem duplications, which occurred in an ancestor of the current species. It has been demonstrated that most newly emerging gene duplicates will degenerate into pseudogenes if they stay at the same location, whereas their translocation to other sites, preferably on a different chromosome, favors their survival (Rodin et al., 2005). In mouse, the tGAP gene loci mainly cluster in a 22 Mb region on chromosome 3 (Fig. 2). It is likely that most of them underwent a conversion to pseudogenes before they could relocate to a new site. However, this theory cannot explain the situation in rat, in which several duplicates survived selection and became newly expressed genes. The rat tGAP genes, which remained on one chromosome, span a larger region (75 Mb) than in mouse. It is also possible that in rat, tGAP gene numbers expanded by a stepwise duplication, that is, the tGAP ancestor was first copied into the seven tGAP loci as single copy and some of these copies were replicated over the time. This is supported by the observation that the tGAP5 genes are quite different from the tGAP genes on other loci, whereas they are only 500 kb apart and highly similar amongst each other. The duplication of the four tGAP5 genes is predicted to be a recent event. A similar case is seen in the region between 154 and 154.2 M, where locate tGAP4

gene and two pseudogenes that are very similar to tGAP4 (Fig. 1). It remains to be examined why tGAP genes are not present in organisms such as primates.

We have identified many ARHGAP20 domain-containing genes by a genomic database search using a consensus GAP-D1 sequence. These genes basically fall into three groups: tGAP, Arhgap20, and T-cell activation RhoGAP. Whereas Arhgap20 and T-cell activation RhoGAP genes are present in many species, tGAP genes are limited to a few mammals. No ARHGAP20 domain-containing genes have been detected in plants such as *Arabidopsis* and in microorganisms such as *Aspergillus*, *Dictyostelium*, and yeast. This suggests that the ARH-GAP20 domain evolved in the animal kingdom. We also failed to find Arhgap20-domain-containing proteins in fruit fly, mosquito, medaka fish, cat, pig, sheep, rabbit, and zebra, but we note that this result is inconclusive because of the early drafting stage of these genome databases.

We were not able to detect any GAP activity toward RhoA, Rac, and Cdc42 for the tGAP1 protein (Modarressi et al., 2004). The multiple sequence alignment of tGAP proteins sheds light on the reason why tGAP1, as well as other rat tGAPs might lack the enzymatic activity. First, the “arginine finger,” which is critical for GTP hydrolysis, is mutated in all GAP-D1 and most of the other GAP domains in rat tGAPs (Fig. 3). The catalytic arginine is conserved among all functional GAP proteins, except for RanGAPs and RapGAPs, which do not use an arginine finger as the catalytic mechanism (Seewald et al., 2002; Chakrabarti et al., 2004). However, the GAP domains of tGAP proteins showed no similarity to either RanGAP or RapGAP. We also failed to detect any interaction between tGAP1 and Ran GTPase in vitro (Supplement Fig. 3). Second, most of the conserved residues involved in interacting and stabilizing the switch domains of GTPases are also substituted (Fig. 5). The latter is also the case for mouse and dog tGAPs, whose arginine finger is nonetheless maintained. This is reminiscent of the p85 regulatory subunit of PI3-kinase, which shares homology with the functional domain of Cdc42GAP but is ineffective as a GAP (Zheng et al., 1994). The p85 subunit contains the catalytic arginine residue but has changes in three residues that hydrogen bond to the residues in the Cdc42 switch domain. We conclude that tGAP proteins do not possess GAP activity and do not function as typical GAP proteins. There have been quite a few RhoGAPs that could act simply as scaffold proteins that mediate crosstalk between Rho GTPases and other signaling pathways, regardless of their GAP activities. (Tcherkezian and Lamarche-Vane, 2007) For example, the p85 subunit of PI3K binds to GTP-bound form of Cdc42 and Rac1 with a nonfunctional GAP domain. Nevertheless, it plays a critical role in recruiting the activated Cdc42 and regulating cytoskeleton change (Zheng et al., 1994). α1-chimaerin, another RhoGAP that lacks GAP activity but retains the ability to bind GTPases, is proposed to co-operate with Rac1 and Cdc42 to promote the formation of lamellipodia and filopodia (Kozma et al., 1996). Provided that the GAP-D1 and GAP-D2 of tGAP proteins are substantially different and may have distinct binding proteins, there is an exciting possibility that tGAP proteins could function as scaffolding proteins and bring different Rho GTPases together to coordinate their cross-talk. Given the fact that tGAP genes are only present in a limited number of species and mainly expressed in testes, their function is likely to be specific to spermatogenesis in such species.

Unlike T-cell activation RhoGAPs, which only resemble tGAP proteins in the GAP-D1 domain (data not shown), Arhgap20 proteins are similar to tGAP proteins in a 300–400 amino acids region spanning the N-terminus to the GAP-D1 domain (Supplement Fig. 1). However, after the GAP-D1 domain, the sequences between tGAP and Arhgap20 proteins diverge completely. We thus hypothesize that tGAP and Arhgap20 genes have a common ancestor.

The parsimony tree shows a few differences when compared to the distance and maximum likelihood trees in our phylogenetic analysis (Fig. 6 and Supplement Fig. 2). We do not believe that this difference is significant because all branches in the parsimony tree that are different from the other two methods have a low statistical score in the bootstrapping analysis (Supplement Fig. 2).

The evolutionary distance of tGAP proteins can be estimated from our phylogenetic analysis. We conclude that tGAP5 proteins are closely related, as are tGAP2 and tGAP4 proteins. tGAP5 proteins, however, are quite distant from other rat tGAP proteins. Interestingly, the mouse BAE21461 protein, which only possesses one ARHGAP20 domain, is more closely related to rat tGAP proteins than to the mouse tGAP protein. This implies that an ancestral tGAP gene in the common ancestor, which led to the tGAP and Arhgap20 genes, might only have possessed one single GAP domain. It was the later acquisition of a second GAP domain that granted tGAP genes new functional features. The second GAP domain was further duplicated within the gene, which led to tGAP genes with three to four GAP domains, such as tGAP1, tGAP2a, tGAP3, and tGAP6. In support of this view, the GAP-D1 domains of the four aforementioned tGAP proteins are all relatively different from the other two or three GAP domains within the protein, whereas the other GAP domains are significantly similar among themselves. In addition, when all GAP domains from rat tGAPs are used to generate phylogenetic trees, all GAP-D1 domains formed one large branched group and other GAP domains formed the other. Within the latter group, however, the GAP-D2, GAP-D3, and GAP-D4 domains are mingled together instead of defined clearly as three subgroups (Supplement Fig. 4).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a CIHR grant to F. A. H. We thank Dr. K. Riabowol, Dr. S. Robbins, and Dr. J. B. Rattner for their comments, which helped to improve the manuscript.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

Chakrabarti PP, Suveyzdis Y, Wittinghofer A, Gerwert K. Fourier transform infrared spectroscopy on the Rap.RapGAP reaction, GTPase activation without an arginine finger. J Biol Chem. 2004; 279:46226–46233. [PubMed: 15292263]

Curry BJ, Su H, Law EG, McLaughlin EA, Nixon B, Aitken RJ. Identification of RARhoGAP, a novel putative RhoGAP gene expressed in male germ cells. Genomics. 2004; 84:406–418. [PubMed: 15234003]

Felsenstein J. PHYLIP (phylogeny inference package). Version 3.2. Cladistics. 1989; 5:164–166.

Jaffe AB, Hall A. Rho GTPases: Biochemistry and biology. Annu Rev Cell Dev Biol. 2005; 21:247–269. [PubMed: 16212495]

Katoh M, Katoh M. Identification and characterization of human KIAA1391 and mouse Kiaa1391 genes encoding novel RhoGAP family proteins with RA domain and ANXL repeats. Int J Oncol. 2003; 23:1471–1476. [PubMed: 14532992]

Kierszenbaum AL, Gil M, Rivkin E, Tres LL. Ran, a GTP-binding protein involved in nucleocytoplasmic transport and microtubule nucleation, relocates from the manchette to the centrosome region during rat spermiogenesis. Mol Reprod Dev. 2002; 63:131–140. [PubMed: 12211070]

Kozma R, Ahmed S, Best A, Lim L. The GTPase-activating protein n-chimaerin cooperates with Rac1 and Cdc42Hs to induce the formation of lamellipodia and filopodia. Mol Cell Biol. 1996; 16:5069–5080. [PubMed: 8756665]

Kusano A, Staber C, Ganetzky B. Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in Drosophila. Dev Cell. 2001; 1:351–361. [PubMed: 11702947]

Kusano A, Staber C, Ganetzky B. Segregation distortion induced by wild-type RanGAP in Drosophila. Proc Natl Acad Sci USA. 2002; 99:6866–6870. [PubMed: 11997467]

Modarressi MH, Cheng M, Tarnasky HA, Lamarche-Vane N, de Rooij DG, Ruan Y, van der Hoorn FA. A novel testicular RhoGAP-domain protein induces apoptosis. Biol Reprod. 2004; 71:1980–1990. [PubMed: 15306557]

Moon SY, Zheng Y. Rho GTPase-activating proteins in cell regulation. Trends Cell Biol. 2003; 13:13–22. [PubMed: 12480336]

Naud N, Toure A, Liu J, Pineau C, Morin L, Dorseuil O, Escalier D, Chardin P, Gacon G. Rho family GTPase Rnd2 interacts and co-localizes with MgcRacGAP in male germ cells. Biochem J. 2003; 372:105–112. [PubMed: 12590651]

Nicholas KB, Nicholas HB, Deerfield DW II. GeneDoc: Analysis and visualization of genetic variation. EMBNEW News. 1997; 4:14.

Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000; 302:205–217. [PubMed: 10964570]

Peck J, Douglas G, Wu CH, Burbelo PD. Human RhoGAP domain-containing proteins: Structure, function and evolutionary relationships. FEBS Lett. 2002; 528:27–34. [PubMed: 12297274]

Quimby BB, Dasso M. The small GTPase Ran: Interpreting the signs. Curr Opin Cell Biol. 2003; 15:338–344. [PubMed: 12787777]

Rittinger K, Walker PA, Eccleston JF, Smerdon SJ, Gamblin SJ. Structure at 1.65 A of RhoA and its GTPase-activating protein in complex with a transition-state analogue. Nature. 1997; 389:758–762. [PubMed: 9338791]

Rodin SN, Parkhomchuk DV, Rodin AS, Holmquist GP, Riggs AD. Repositioning-dependent fate of duplicate genes. DNA Cell Biol. 2005; 24:529–542. [PubMed: 16153154]

Sambrook, J., Russell, D. Molecular cloning: A laboratory manual. 3. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001. c2001

Seewald MJ, Korner C, Wittinghofer A, Vetter IR. RanGAP mediates GTP hydrolysis without an arginine finger. Nature. 2002; 415:662–666. [PubMed: 11832950]

Steggerda SM, Paschal BM. Regulation of nuclear import and export by the GTPase Ran. Int Rev Cytol. 2002; 217:41–91. [PubMed: 12019565]

Tcherkezian J, Lamarche-Vane N. Current knowledge of the large RhoGAP family of proteins. Biol Cell. 2007; 99:67–86. [PubMed: 17222083]

Toure A, Morin L, Pineau C, Becq F, Dorseuil O, Gacon G. Tat1, a novel sulfate transporter specifically expressed in human male germ cells and potentially linked to rhogtpase signaling. J Biol Chem. 2001; 276:20309–20315. [PubMed: 11278976]

Yamada T, Sakisaka T, Hisata S, Baba T, Takai Y. RA-RhoGAP, Rap-activated Rho GTPase-activating protein implicated in neurite outgrowth through Rho. J Biol Chem. 2005; 280:33026–33034. [PubMed: 16014623]

Zheng Y, Bagrodia S, Cerione RA. Activation of phosphoinositide 3-kinase activity by Cdc42Hs binding to p85. J Biol Chem. 1994; 269:18727–18730. [PubMed: 8034624]

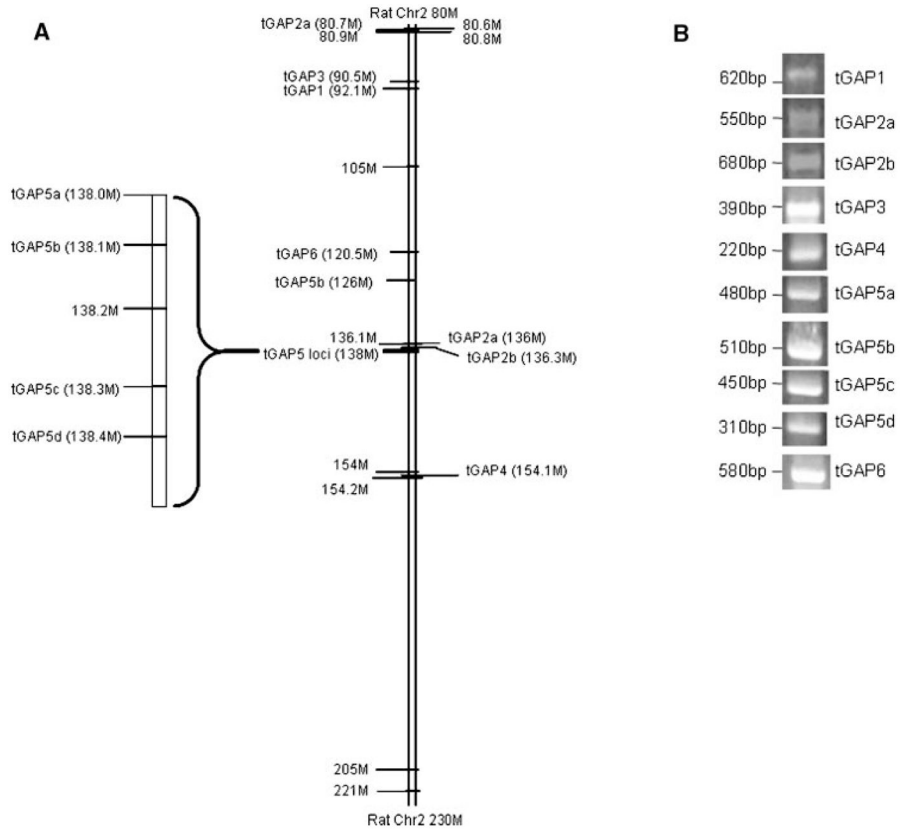

Fig. 1.

tGAP genes in rat. **A**: Chromosomal localization of rat tGAP genes. BLAST searches were performed on NCBI rat genomic database with tGAP1 mRNA or amino acid sequences. Most significant hits on potential tGAP-like genes cluster on a 150 Mbp region (denoted in figure) on chromosome 2. Ten tGAP genes were found in this region on six loci: tGAP1 (92.1 M); tGAP2 (136 M); tGAP3 (90.5 M); tGAP4 (154 M); tGAP5 (138 M); tGAP6 (120.5 M). tGAP2a and tGAP5b genes have two copies on 80.7 and 126 M, respectively. Several tGAP-like pseudogenes (or noncoding transcripts) were identified in the same region. **B**: Expression of tGAP genes was investigated in rat testis using RT-PCR with primers specific to each tGAP gene.

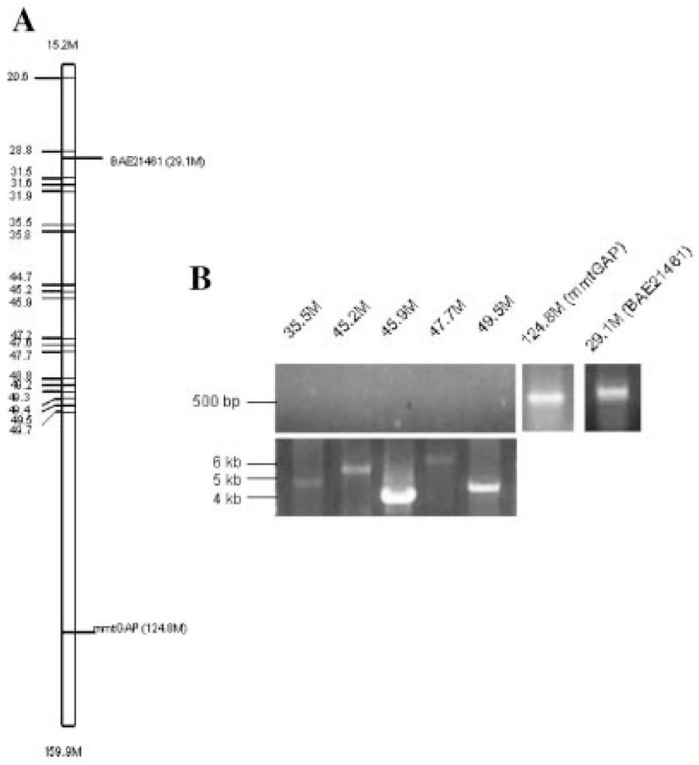

Fig. 2.

tGAP genes in mouse. **A**: Chromosomal localization of the mouse tGAP gene and related pseudogenes. NCBI mouse genomic database was searched using tblastn with the rat consensus GAP-D1 amino acid sequence. Most significant hits on potential tGAP-like genes cluster on a region on chromosome 3 that is syntenic with the rat chromosome region depicted in Figure 1. In contrast to rat tGAP genes, we could only identify one tGAP gene (mmtGAP) on 124.8 Mbp. The other loci are possibly pseudogenes. BAE21461 is highly similar to rat tGAP genes and is not a pseudogene in our analysis. However, it has only one GAP domain, thus does not meet our definition of tGAP genes. **B**: Expression of mmtGAP gene in mouse testis. In the upper panel, primers designed for several potential mouse tGAP gene loci were used in RT-PCR analysis with mouse testis cDNA. The band size of mmtGAP and BAE21461 is 520 and 570 bp, respectively. In the lower panel, the same primers were used to amplify the genomic DNA fragments. The size of band for each locus is: 35.8 M, 4.9 kb; 45.2 M, 5.8 kb; 45.9 M, 3.9 kb; 47.7 M, 6.5 kb; 49.5 M, 4.4 kb.

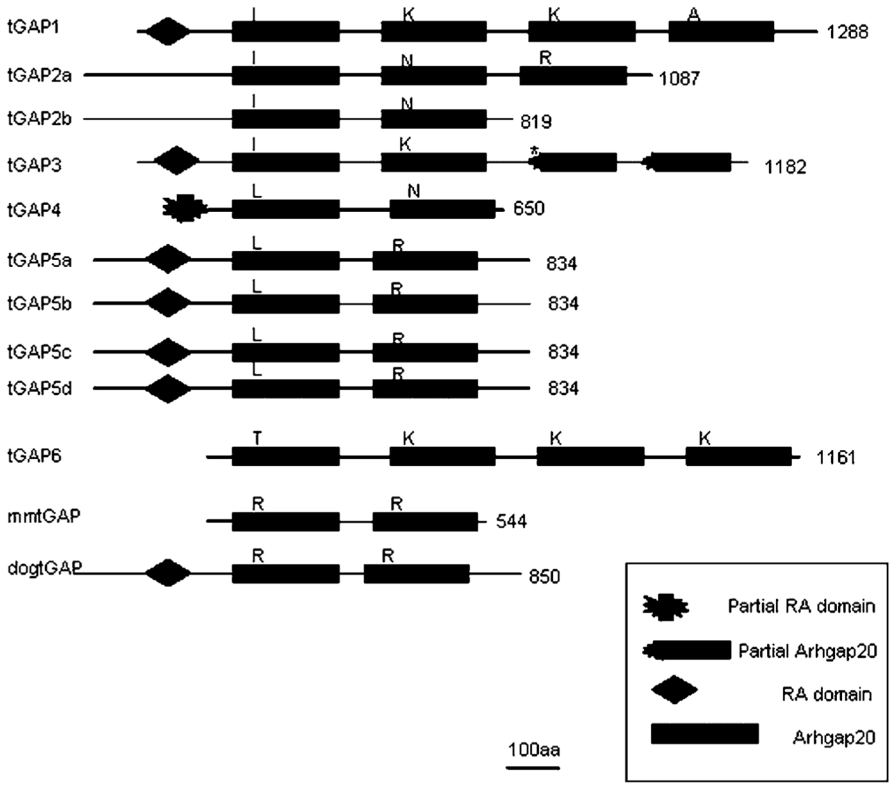

Fig. 3.

Domain structure of tGAP proteins. The letters on Arhgap20 domains indicate the amino acid at the position of the arginine finger. The numbers indicate the length of the protein. * The last two GAP domains of tGAP3 are predicted to be incomplete (missing the first one-third).

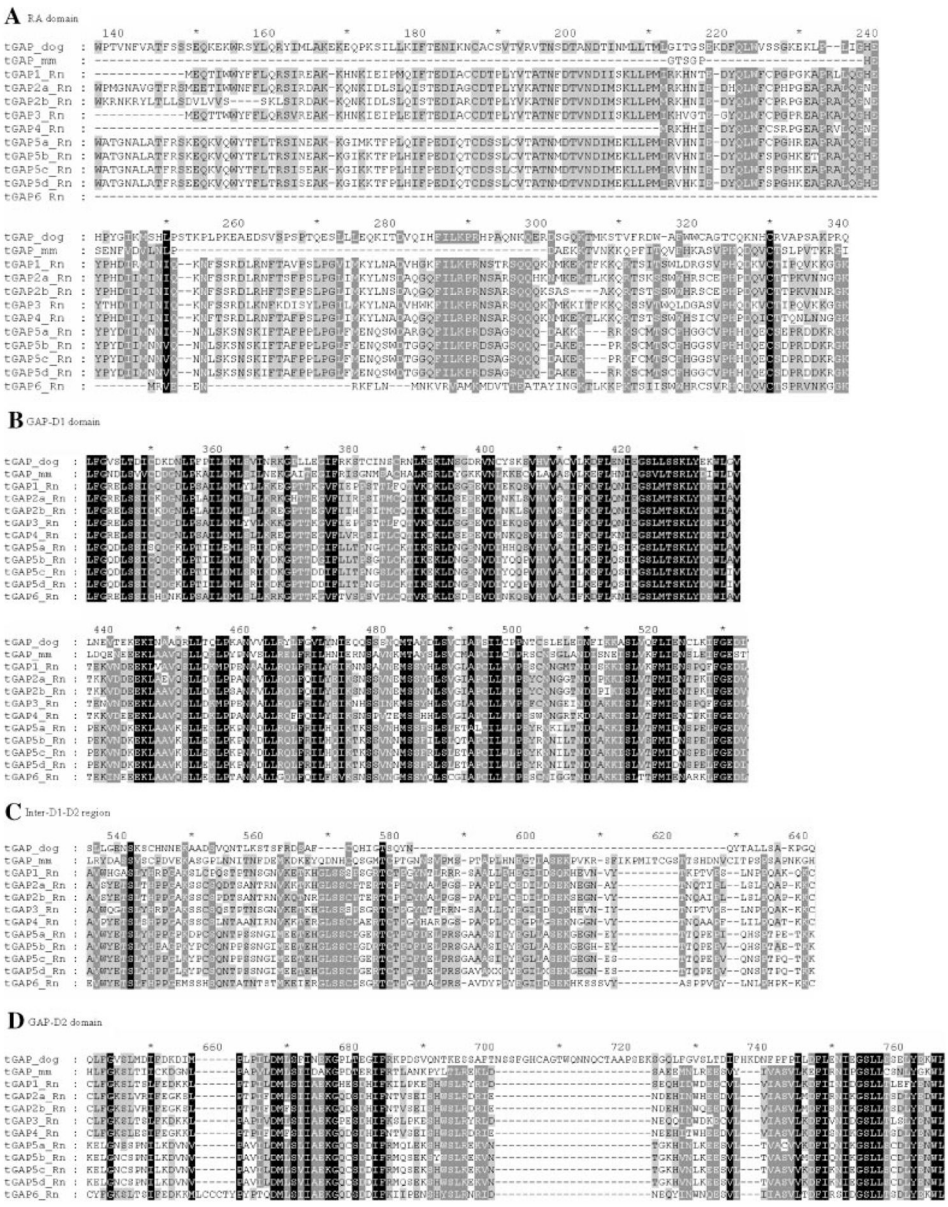

Fig. 4.

The four conserved regions in tGAP proteins. The multiple sequence alignment was done using T-COFFEE. The alignment was then visualized by the multiple sequence alignment editor, GENEDOC. Shading is based on conservation, with the darkest shading representing 100% conservation and the lightest representing 60%. The four conserved region indicated are: (**A**) the RA domain; (**B**) GAP-D1 domain; (**C**) sequences between GAP-D1 and GAP-D2; (**D**) GAP-D2 domain.

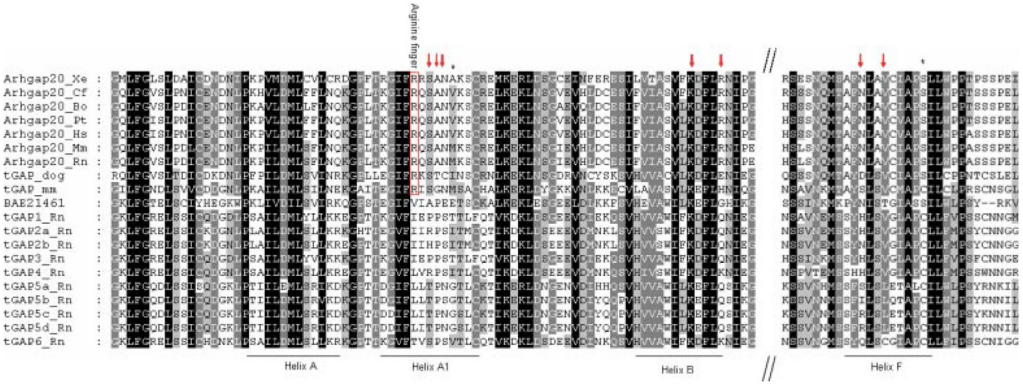

Fig. 5.

The arginine finger and flanking important regulatory residues are substituted in tGAPs. The arrows indicate residues conserved in both canonical RhoGAPs (such as Cdc42GAP) and Arhgap20 proteins but not conserved in tGAP proteins. The asterisks indicate residues conserved in canonical RhoGAPs but not in Arhgap20 and tGAP proteins. [See color version online at www.interscience.wiley.com.]

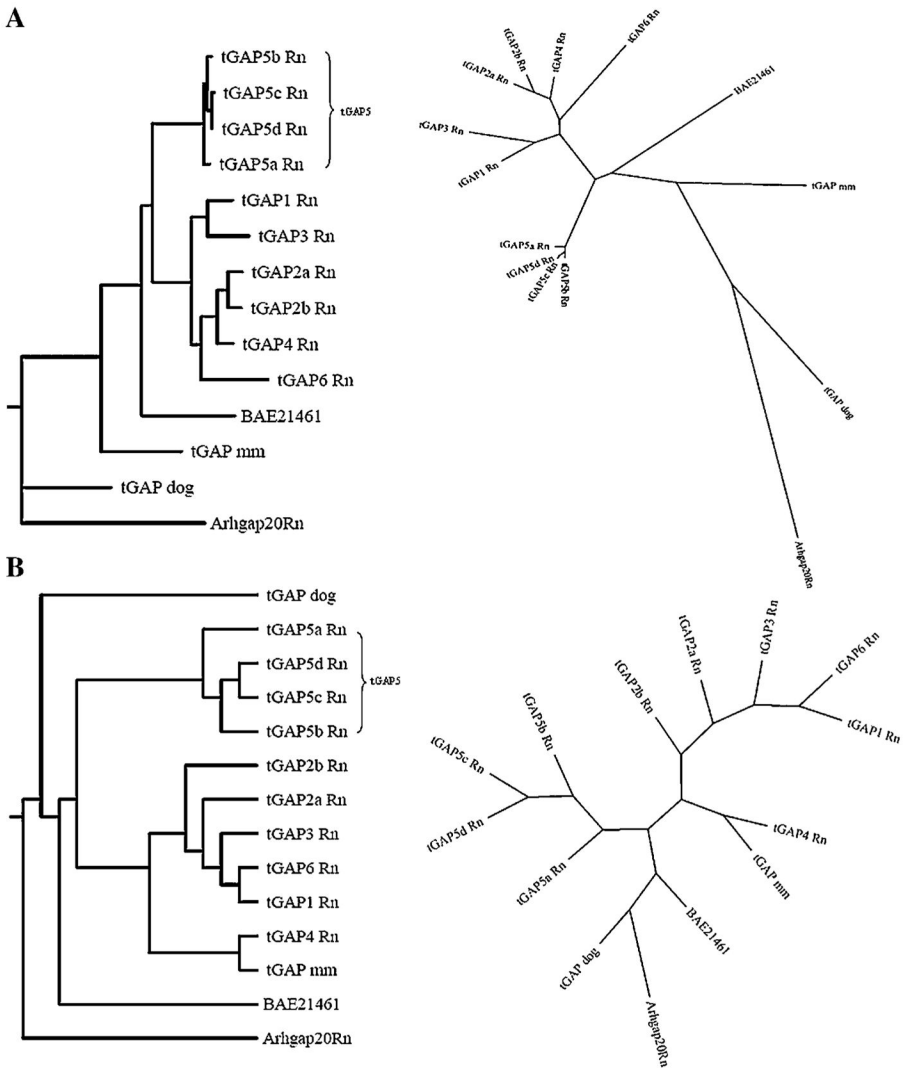

Fig. 6.

Phylogenetic trees of tGAP proteins. The tree in 6A is derived by maximum-likelihood analysis, whereas the tree in 6B is derived by parsimony analysis.

TABLE 1

The Primers Used in the PCR Analysis

Species	Corresponding genes	Primer sequences
Rattus norvegicus	tGAP1	Fwd: 5′-GGAACGCGTGCTTGTAATAGCAG-3′ Rev: 5′-TGATGGTGTTGGATTAGAA-3′
	tGAP2a	Fwd: 5′-CCAGAAACGTGAAGAAAACCAA-3′ Rev: 5′-TTCGGGCATCTGTCGTGTAAGT-3′
	tGAP2b	Fwd: 5′-CCAGAAACGTGAAGCAAACCAA-3′ Rev: 5′-AGTCCAACCCCATAAGACATGT-3′
	tGAP3	Fwd: 5′-CTTACTAACATTGTCTTTAT-3′ Rev: 5′-AAATTTATTGACGCTTTACG-3′
	tGAP4	Fwd: 5′-CCAAGCAGCACCTTTTCTTATTC-3′ Rev: 5′-TGCCAGGTTATGTGTTCTTCAT-3′
	tGAP5a	Fwd: 5′-TCTCCCTCCCAGGCTTCG-3′ Rev: 5′-GGTTTGGATATCCTCAGGGAATATC-3′
	tGAP5b	Fwd: 5′-GAGTGCATAACATTGAAGATTACCAG-3′ Rev: 5′-CTTTATTGTTTTGCATAATGTTCCAC-3′
	tGAP5c	Fwd: 5′-CAATGGTATACTTTCCTCACCAGG-3′ Rev: 5′-ACATGATGTCATGCAAAATTTTCTAG-3′
	tGAP5d	Fwd: 5′-TGTGGGTGGTGGATGATGA-3′ Rev: 5′-TGAACCAAAGTTGGTAATCTTCAATT-3′
	tGAP6	Fwd: 5′-CGAGAAGTGCAGTTGATTATCC-3′ Rev: 5′-TTGTCAATCATGATTTTTATAAGAGATG-3′
Mus musculus	mmtGAP	Fwd: 5′-TGCACATCCCTGCCAGTTAC-3′ Rev: 5′-AGGAACGAGGCAGACATAGGA-3′
	BAE21461	Fwd: 5′-CTCTTCGGAACAGAACTGAGTTGCATC-3′ Rev: 5′-CATGTCAATTCCAAAAATTTCAGGAGAGTT-3′
	35.5 M	Fwd: 5′-CAAGAATGCCCTGCCCCAACA-3′ Rev: 5′-GCATCATGACCAGGCATGCGT-3′
	45.2 M	Fwd: 5′-TCAAGCAAATGCTGCACTTCTCAGAAC-3′ Rev: 5′-CAGGGACCAGTGTGATTCAATGGGAAT-3′
	45.9 M	Fwd: 5′-GGGAAGAGGCAGATATAAAAAAGCAC-3′ Rev: 5′-GTGGATGAAATGGTGATGTTTCATAAC-3′
	47.7 M	Fwd: 5′-TGAATCCAATATTGTTCCAAACACTAAGG-3′ Rev: 5′-TGGATAAATTTTTATCACAAATTTTACCAGAG-3′
	49.5 M	Fwd: 5′-ATGTGCACTGCCCCAACAGTTAATAGAG-3′ Rev: 5′-CGAGGAAGCTGCTGAGTTCCTGT-3′

TABLE 2

Summary of the Results From a Search of Available Genomic Databases Using a Consensus GAP-D1 Amino Acid Sequence From Rat tGAP Proteins

Species	Arhgap20[a]	T-cell activation RhoGAP[b]	tGAP like[c]	Notes for tGAP-like genes
Human	BAA92629[d]	NP_473455		
Bovine	XP_593534		NW_001502124[e] NW_001501897 NW_0014943993	Predicted, no EST evidence
Chimpanzee	XP_522177			
Beetle (*Tribolium castaneum*)	XP_974029			
Chicken	XP_417154	XP_419701		
Dog	XP_854212	XP_541176	NW_876267 NW_879505	Predicted, the first one has EST evidence
Xenopus tropicalis	AAI25688	AAI24045		
Horse			AAWR010026	Incomplete preliminary sequences
C. elegans		NP_493035		
Rhesus monkey	XP_001098860	XP_001095675		
Sea urchin	XP_001185876			
Zebrafish	XP_001336763	AAH81595		
Fugu fish	CAG08249	CAG01527		
Rat	NP_998794	XP_001067756	6 loci, 10 genes	Presence of tGAP pseudogenes
Mouse	NP_780744	Q8K2L9	XP_995294	Presence of many tGAP pseudogenes

[a] Arhgap20 genes which are also named KIAA1391 or RARhoGAP genes.

[b] T-cell activation RhoGAP genes that contain one ARHGAP20 domain but no other recognizable conserved domain.

[c] tGAP-like genes that contain at least two ARHGAP20 domains.

[d] The accession numbers presented in the table for Arhgap20 and T-cell activation RhoGAP are protein sequences.

[e] The accession number for tGAP-like genes in bovine, dog and horse are genomic sequences.

TABLE 3

Percentage Identity Among the Amino Acid Sequences of the First ARHGAP20 Domain (GAP-D1) of All tGAP Proteins

	tGAP2a	tGAP2b	tGAP3	tGAP4	tGAP5a	tGAP5b	tGAP5c	tGAP5d	tGAP6	mmtGAP	dogtGAP
tGAP1	80	79	93	80	63	66	64	64	78	52	45
tGAP2a		95	77	88	62	64	63	63	79	52	50
tGAP2b			76	87	62	64	62	62	79	52	49
tGAP3				76	63	66	64	64	77	51	46
tGAP4					61	63	61	61	80	50	48
tGAP5a						92	91	92	63	53	45
tGAP5b							94	94	65	53	47
tGAP5c								99	63	54	47
tGAP5d									63	54	47
tGAP6										51	47
mmtGAP											54

The 192-amino-acid sequences of GAP-D1 from all tGAP genes were compared using NCBI pairwise BLAST. The numbers in the table indicate the percentage identity between the GAP-D1 of two tGAP genes. For example, 80 at the cross of tGAP1 and tGAP2a means these two proteins have 80% identity in the amino acids of their GAP-D1.