Phylogenetic Analysis of RhoGAP Domain-Containing Proteins

Marcelo M. Brandão[1], Karina L. Silva-Brandão[2], Fernando F. Costa[1], and Sara T.O. Saad[1]*

[1] *Hematology and Hemotherapy Center, State University of Campinas, 13083-970 Campinas, SP, Brazil;*
[2] *Department of Zoology, Institute of Biology, State University of Campinas, 13083-970 Campinas, SP, Brazil.*

Proteins containing an Rho GTPase-activating protein (RhoGAP) domain work as molecular switches involved in the regulation of diverse cellular functions. The ability of these GTPases to regulate a wide number of cellular processes comes from their interactions with multiple effectors and inhibitors, including the RhoGAP family, which stimulates their intrinsic GTPase activity. Here, a phylogenetic approach was applied to study the evolutionary relationship among 59 RhoGAP domain-containing proteins. The sequences were aligned by their RhoGAP domains and the phylogenetic hypotheses were generated using Maximum Parsimony and Bayesian analyses. The character tracing of two traits, GTPase activity and presence of other domains, indicated a significant phylogenetic signal for both of them.

Key words: Bayesian analysis, character tracing, parsimony, phylogenomics, protein domain, RhoGAP

Introduction

The Rho GTPase-activating proteins (RhoGAPs) are defined by the presence of a 150-amino-acid homolog region that is designated as the RhoGAP domain. This domain is necessary and sufficient for GAP activity and shares at least 20% sequence identity among its family members (*1*, *2*). Proteins containing an RhoGAP domain act as molecular switches involved in the regulation of diverse cellular functions, including actin cytoskeleton rearrangements, regulation of gene transcriptions, cell cycle regulation, control of apoptosis, and membrane trafficking (*2–5*). Rho GTPases cycle between active and inactive GTP-bound states. The control of these states is regulated by three main classes of proteins: guanine nucleotide exchange factors, guanine nucleotide dissociation inhibitors, and GAPs. To date, at least 21 Rho GTPases have been defined, among which only three (RhoA, Cdc42, and Rac1) are well characterized. Therefore, most studies have been focusing on these three proteins.

The ability of Rho GTPases to regulate a wide number of cellular processes comes from their interactions with multiple effectors or inhibitors. One class of these inhibitors is the RhoGAP family, which stimulates GTPase activity by enhancing the intrinsic rate of GTP hydrolysis.

In the early analyses of the human genome sequence, 77 different genes containing the RhoGAP domain were found. Further studies have demonstrated that many of these genes are simple gene sequence variations or single nucleotide polymorphisms (*6*, *7*). The structural data available for RhoGAP domain-containing proteins showing their complexity with Rho GTPases (Cdc42 and RhoA) demonstrated the 3D workflow for RhoGAP-mediated GTP-hydrolysis, and highlighted the importance of a well-conserved arginine residue present in the active site that acts as a conformation stabilizer needed for hydrolysis (*8–10*).

Recently, a novel member of the RhoGAP family, ARHGAP21 (Rho GTPase-activating protein 21, alias ARHGAP10), was cloned and characterized in our laboratory (*11*). In addition to the RhoGAP domain, ARHGAP21 presents a PH domain and a P-loop-containing PDZ domain. This gene is widely expressed at high levels in muscle and brain, and is up-regulated during myeloid and erythroid maturation, suggesting a potential role for this RhoGAP in regulating cell differentiation (*11*).

The aim of this study is to infer the evolution of the RhoGAP superfamily using a phylogenetic approach, to determine the roles of other domains and their main GTPase activities in this evolutionary his-

***Corresponding author.**
E-mail: sara@unicamp.br

tory, and to provide a tool that could render some insights regarding subfamily protein functions using RhoGAP-containing proteins as a model.

Results

The full dataset contains 267 amino acids, of which 161 are parsimony-informative. Parsimony searches of the equally weighed dataset resulted in 12 equally parsimonious trees with 3,213 steps (CI=0.449; RI=0.525). The strict consensus tree was mostly not resolved at internal nodes (data not shown), but almost all terminal branches showed strong bootstrap support (Figure 1).

A Bayesian tree recovered mostly the same terminal relationships by Maximum Parsimony (MP) analyses with strong values of Posterior Probability (PP). However, internal branches have from low to moderate PP values (Figure 1).

The two characters investigated (domains and GTPase activity) showed a significant phylogenetic signal (P=0.003), which suggests that the distribution of these traits among the proteins can be explained by their phylogenetic relationships (Figure 2). The optimization of the other domains over our phylogenetic hypothesis suggests that the ancestral state of these proteins involves solely the presence of the RhoGAP domain and the activity toward Rac1.

The character tracing of these traits suggests an overall pattern on which proteins sharing equal domains also share equal GTPase activity. The clade joining KIAA0672 + 3BP1 + RICH1 + Nadrin was recovered both by MP and Bayesian analyses with strong support. All proteins in this clade share the presence of the BAR (Bin-Amphiphysin-Rvs) domain and GTPase activity toward Rac1, solely or in addition to other GTPases. The same rule can be applied to the clade joining STARTdom + GT650 + DLC1 + AHRGAP7. All of them share the presence of the START (steroidogenic acute regulatory protein-related lipid transfer) domain and all, but STARTdom, have GTPase activity over RhoA. GTPase activity is unknown for STARTdom. The clade joining GMIP + HA1 + PARG is composed by proteins containing the C1 domain in addition to their GTPase activity toward RhoA.

The clade joining AHRGAP11A + AHRGAP20 + AHRGAP1 + AHRGAP8 has in common the absence of other domains that are not RhoGAP. The GTPase activity of this group is known only for AHRGAP1, which is active toward Cdc42. Considering other terminal clades, we can presume that other AHRGAPs within this group can show the same GTPase activity toward Cdc42.

On the other hand, the clade joining P115 + srGAP3 + srGAP1 + srGAP2 shares the presence of the FCH (FER/CIP4-homology) domain, however, P115 has GTPase activity over Rac1, while the other three proteins have activity toward Cdc42.

Discussion

Phylogenetic reconstruction and bioinformatics analyses that integrate evolutionary considerations are becoming increasingly important tools for applied fields. Numerous gene sequences were generated in the genomics age with little or no accompanying experimental determination of functional information or evolutionary relationships. Previous works from Peck *et al* (*12*) and Moon and Zheng (*2*) also present a phylogenetic approach on the RhoGAP family; however, the authors did not indicate the methodology applied neither did they present any support analyses for their cladograms.

In this work, bioinformatics and phylogenomics tools were used to present a phylogenetic relationship of 59 members of the RhoGAP superfamily. All amino acid alignments and subsequent phylogenetic tree constructions were based on the RhoGAP domain sequence. We demonstrated that these RhoGAP domain-containing proteins, with the conservative argenine residue, form a monophyletic group, that is, all of them share a common protein ancestor in their evolutionary history.

The tracing for GAP activity toward the most studied RhoGTPases (RhoA, Rac1, and Cdc42) (Figure 2) indicates that this trait presents a strong phylogenetic signal (P=0.003), contrasting with previous findings of Peck *et al* (*12*).

The analysis of the resulting phylogenetic tree has suggested that the ancestral state for GTPase activity is the affinity to Rac1. It is still difficult to determine the gap activity by only analyzing the protein sequence; the GAP assay (*13*, *14*) is the most reliable way to determine activity. The phylogenetic approach may give a clue, once it is capable of clustering together different proteins that share common substrates as can be seen on the clades of KIAA0672 + 3BP1 + RICH1 + Nadrin and srGAP3 + srGAP1 + srGAP2. Speculations regarding protein specific functions (only using the GTPase activity character)