OPEN ACCESS Freely available online

PLoS one

The Repertoire of Heterotrimeric G Proteins and RGS Proteins in *Ciona intestinalis*

R. Prasobh, Narayanan Manoj*

Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

Abstract

Background: Heterotrimeric G proteins and regulators of G protein signaling (RGS) proteins are key downstream interacting partners in the G protein coupled receptor (GPCR) signaling pathway. The highly versatile GPCR transmembrane signaling system is a consequence of the coupling of a diverse set of receptors to downstream partners that include multiple subforms of G proteins and regulatory proteins including RGS proteins, among others. While the GPCR repertoire of *Ciona intestinalis*, representing the basal chordate is known, the repertoire of the heterotrimeric G proteins and RGS proteins is unknown.

Methodology/Principal Findings: In the present study, we performed an *in-silico* genome-wide search of *C. intestinalis* for its complement of G proteins and RGS proteins. The identification of several one-to-one orthologs of human G proteins at the levels of families, subfamilies and types and of homologs of the human RGS proteins suggests an evolutionarily conserved structure function relationship of the GPCR signaling mechanism in the chordates.

Conclusions: The *C. intestinalis* genome encodes a highly conserved, albeit, limited repertoire of the heterotrimeric G protein complexes with the size of subunit types comparable with that in lower eukaryotes.

Citation: Prasobh R, Manoj N (2009) The Repertoire of Heterotrimeric G Proteins and RGS Proteins in *Ciona intestinalis*. PLoS ONE 4(10): e7349. doi:10.1371/journal.pone.0007349

Editor: Janet Kelso, Max Planck Institute for Evolutionary Anthropology, Germany

Received July 15, 2009; **Accepted** September 16, 2009; **Published** October 6, 2009

Copyright: © 2009 Prasobh, Manoj. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by a grant (BT/BI/25/001/2006) from the Department of Biotechnology, Government of India. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nmanoj@iitm.ac.in

Introduction

G protein coupled receptors (GPCRs) comprise a large family of diverse transmembrane signaling proteins that receive information from various extracellular stimuli including hormones, neurotransmitters or sensory stimuli. The three-component GPCR signaling system is so named based on its ability to recruit and regulate the activity of intracellular heterotrimeric G proteins (guanine nucleotide-binding protein). Ligand binding at the extracellular recognition sites on the GPCRs is transduced into intracellular signals through the coupling of the receptor and G protein and between G protein and other effector proteins, all of which can be independently regulated by additional proteins or at the transcriptional level.

The heterotrimeric G protein is composed of α, β, and γ subunits. The inactive form of the GDP-bound heterotrimeric state is activated when the agonist activated receptor induces a conformational change in the G protein trimer resulting in the Gα-subunit binding to GTP in exchange for GDP. This exchange leads to the dissociation of Gα-GTP and Gβγ subunits that further interact with downstream target effectors thus activating and regulating a signaling cascade. The turnoff of the cellular response occurs when the Gα subunit hydrolyses GTP to GDP and the Gα-GDP and Gβγ subunits re-associate to form the Gαβγ trimer. The Gβ and Gγ subunits work as an obligate complex, a functional unit that can only be dissociated under denaturing conditions. The structural and functional aspects of G proteins and their receptor-mediated activation have been extensively studied [1–5]. The third component of the GPCR signaling system is the G protein-regulated effector. Several effectors enhance GTPase activity of the Gα subunit thus playing a role in deactivation and modulation of G protein mediated signaling. A recent modification to the standard model of GPCR signaling has come from a family of proteins called "regulators of G protein signaling" or RGS proteins that have been found to accelerate the intrinsic GTPase activity of Gα subunits independent of the Gβγ subunits [6].

The modular architecture of the G protein-mediated transmembrane signaling system is highly versatile and specific and is based on the fact that there are numerous receptors (around 800 in human) and several types of G proteins and effector proteins [7]. Enormous diversity of heterotrimeric complexes can be assembled from a limited repertoire of G protein subunits which are then activated by different receptors [8]. Most receptors also are able to activate more than one type of G protein [9]. Moreover, several Gβγ complexes can interact with the same Gα suggesting that differential expression or subcellular localization are important in the regulation of downstream signaling [10].

Ciona intestinalis (henceforth referred to as *Ciona*) is a protochordate belonging to the ascidian class of chordates that diverged from the vertebrate lineage about 520 million years ago. An out-group to the vertebrates, this ascidian has the smallest genome of any experimentally manipulable chordate. A translucent morphology, availability of developmental mutants, quickly spawning embryos, established transgenic, morpholino-based gene knockdown, *in situ*

hybridization experimental procedures and extensive EST data are some of the many advantages that make *Ciona* an excellent model organism to study developmental and evolutionary biology of the vertebrate-invertebrate split. Moreover, *Ciona* possesses organ systems that are homologous to vertebrate heart, thyroid, blood, digestive and neural complex systems [11–14]. In an earlier study, a genome wide survey of the repertoire of GPCRs in *Ciona* reported the presence of 169 putative receptors. A comparative analysis of the repertoire revealed a high level of orthology with that of human GPCRs (about 40% of *Ciona* receptors) [15]. Here we extend the previous study and report the identification of the repertoire of the heterotrimeric G proteins and the RGS proteins in *Ciona* and present a comparative analysis with that in human. The analyses could provide insights into the origin and evolution of the GPCR signaling system in a protochordate and its further diversification into the vertebrate lineage. Our results thus could serve as a basis for carrying out experimental studies to address functional and regulatory aspects of GPCR mediated signaling.

Materials and Methods

Protein data mining

The Joint Genome Institute (JGI) *C. intestinalis* genome versions 2.0 and 1.0 databases (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html; http://genome.jgi-psf.org/ciona4/ciona4.home.html) were used as the source for obtaining the complete proteome [16–17]. The v2.0 proteome dataset includes 15,852 proteins and sequences in the current version include both automated and to a lesser extent, manually curated annotations. The sequences of the query G proteins and RGS proteins were obtained from the NCBI non-redundant database and the gpDB Database [18]. The gpDB Database is a relational database of GPCRs and GPCR interacting proteins based on interactions reported in the literature. The BLAST program was used for initial identification using a cut off of E-value of 10^{-3} [19]. Additionally, searches were carried out using customized Hidden Markov Models (HMM) built separately for all known families/sub-families listed in the gpDB Database. The HMMs were constructed using HMMBUILD and HMMCALIBRATE programs of HMMER package version 2.3.2 [20]. *Ciona* sequences that returned hits with an E-value better than 0.01 were extracted. Default settings were used in all HMMBUILD and HMMCALIBRATE models constructed. At this stage of the analysis, E-values of low stringency were used for BLAST and HMM search algorithms to enable detection of all true positives. False positives, if any, were sifted out in the next stage with a higher level of stringency where the results of BLAST, best reciprocal BLAST hit and HMM based search were compared and common hits satisfying the cutoff criteria were taken as confirmed hits of *Ciona*. All except two of *Ciona* sequence accession numbers reported in this study, including that of several manually edited fragments, refer to the identifiers assigned in the JGI v2.0 genome database. GenBank accession numbers, where available, are given in the supporting information.

Sequence comparison and phylogenetic methods

Multiple sequence alignment of the *Ciona* sequences with known orthologs picked up from the gpDB database was performed using CLUSTALX 1.8 program [21] using the default parameters for substitution matrices and gap penalties. Manual adjustments of the alignments were made when necessary using the BioEdit program [22]. All phylogenetic analysis was carried out using the PHYLIP package, as implemented in the MOBYLE Portal (http://mobyle.pasteur.fr/) [23]. Neighbor-Joining (NJ) and Maximum-Likelihood (ML) methods as implemented in the NEIGHBOR and PROML programs were employed for tree searching and inference. The statistical reliability of the phylogenetic trees was tested by interior branch analysis with 500 bootstrap replicates used to create a consensus tree. Phylogenetic trees created using the NJ method and the ML method resulted in trees with identical topologies and clusters with significant bootstrap support. Alignments and secondary structures were displayed using ESPript 2.2 (http://espript.ibcp.fr) [24].

Expressed Sequence Tag hits and tissue/developmental stage based expression data

The entire UniGene database for the *Ciona* transcript sequences was downloaded from the ftp site (ftp://ftp.ncbi.nih.gov/repository/UniGene/) and locally installed. All *Ciona* G protein and RGS protein sequences were queried against the UniGene EST database using TBLASTN with the E-value set at $1e^{-5}$. Identifiers of the statistically significant EST-hit collection were imported in an excel sheet and categorized based on their derivative developmental stage/tissues.

Results and Discussion

In order to generate an independent data set of the repertoire of G proteins and RGS proteins, a data mining approach was taken using BLAST program and HMM based searches. Sequences from the gpDB Database and the NCBI were collected and classified into different groups and families of G proteins and RGS proteins. The sequences were carefully chosen to represent diverse organismal lineages and groups as reported in the gpDB classification. The search strategy that included using sequences as single queries and also using distantly homologous sequences combined into a profile for a more sensitive HMM based search, resulted in the identification of unambiguous homologs in the *Ciona* proteome. The query sequences used for the BLAST searches and for the construction of HMMs are available as supplemental data (Data S1). Further classification of the *Ciona* sequences into orthologous sets of different classes, families, subfamilies and types were carried out based on phylogenetic analyses (Table 1, Table S1). The complete amino acid sequences of all identified *Ciona* proteins are available as supplemental data (Data S2).

Since EST evidence covers three quarters of the predicted *Ciona* genes, we also identified EST matches for the predicted *Ciona* G proteins and RGS proteins. Except for 2 out of 10 Gα sequences and 5 out of 14 RGS sequences, all other sequences had at least one EST match mapped to the gene sequences. EST matches for these sequences were sorted as being derived from different developmental stages as well as tissues (Table S2).

The Gα subunit

The Gα class of the heterotrimeric G proteins comprises four main families, $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$, each consisting of several members with different expression patterns. Members of these families are structurally conserved and often share functional properties. The gpDB Database lists four other families that include members from nematodes, fruitfly, fungal_plant and one group designated as 'Unclassified Gα". The human repertoire of Gα class as listed in the gpDB database includes a total of 31 sequences with each of the families $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$, containing 9, 13, 5 and 2 members, respectively. In our study, a total of 10 unambiguous orthologs of the Gα class were identified in *Ciona*. Three sequences out of the ten members were fragments, possibly the result of errors in the automated splice site prediction methods used. In these cases, additional searches carried out against the

Table 1. Classification of heterotrimeric G proteins and RGS proteins in *C. intestinalis*.

Class	Family	Subfamily	Type	JGI Accession *C. intestinalis* sequence[a]
Gα	$G_{12/13}$	$G\alpha_{12}$, $G\alpha_{13}$, Unclassified $G_{12/13}$	$G\alpha_{12}$	287420 (Q14344–59%)
	$G_{i/o}$	Gα gust	Gα gust	-
		$G\alpha_i$	$G\alpha_i$, $G\alpha_{i1}$, $G\alpha_{i2}$, $G\alpha_{i3}$	209567 (P63096–83%)
		$G\alpha_o$	$G\alpha_o$, $G\alpha_{io1}$	-
		$G\alpha_t$	$G\alpha_t$, $G\alpha_{t1}$, $G\alpha_{t2}$	-
		$G\alpha_z$	$G\alpha_z$	-
	$G_{q/11}$	$G\alpha_{11}$, $G\alpha_{14}$, $G\alpha_{15/16}$, $G\alpha_q$, Unclassified $G_{q/11}$	$G\alpha_{11}$, $G\alpha_{14}$, $G\alpha_{15/16}$, $G\alpha_q$, Unclassified $G_{q/11}$	272522 (P50148–74%), 390373 (P29992–42%)
	G_s	$G\alpha_{olf}$	$G\alpha_{olf}$	-
		$G\alpha_s$	$G\alpha_s$	208025 (P63092–63%)
	Unclassified Gα	Unclassified Gα	Unclassified Gα	281048 (P63096–33%)
	Other Gα	Other Gα	Other Gα	220350 (P63096–59%) 273922 (P20353–56%) 202119 (P63096–61%) 287065 (P63096–59%)
Gβ	Gβ	Gβ	$G\beta_1$, $G\beta_2$, $G\beta_3$, $G\beta_4$	283613 (P62873–78%)
			$G\beta_5$	297548 (O14775–57%)
	Unclassified Gβ	Unclassified Gβ	Unclassified Gβ	-
Gγ	Gγ	Gγ	Gγ (1–5,7,8, 10–13), $G\gamma_e$, $G\gamma_{t1}$, Unclassified Gγ	225022 (O60262–40%) ci0100144696 (Q9UBI6–35%)
RGS with GGL domain	R7	R7	RGS6,RGS7,RGS9, RGS11	373767 (P49802–55%)

[a]data in parenthesis indicates the UNIPROT accession and values of sequence identity with the top human sequence match.

doi:10.1371/journal.pone.0007349.t001

NCBI non-redundant protein database and the *Ciona* proteome v1.0, yielded two sequences that were full length proteins. The validity of the edited sequences was verified by comparison of these regions against the EST database. In the case of the third fragment (281048), the missing region was identified by the use of translated alignments against the *Ciona* genomic regions where the missing region was expected to be found based on alignment with the human orthologous protein. However, in this case, the EST database contained no hits corresponding to the genomic region. The *Ciona* Gα homologs share a relatively high pair wise sequence identity ranging from 33–83% with the best match in human (Table 1). The sequence identity values within the *Ciona* Gα proteins range from 26–99%.

The classification of the *Ciona* sequences into different families and subfamilies, carried out by phylogenetic analysis, revealed distinct clusters that corresponded to known families and subfamilies in human and other organisms (Figure 1). The phylogenetic tree showed monophyletic clades with good bootstrap support for each of the major families. The clustering shows two orthologs of the $G_{q/11}$ family and one ortholog each of the $G\alpha_{i/o}$ family, the $G\alpha_s$ family and the $G\alpha_{12/13}$ family. The $G\alpha_{i/o}$ ortholog (209567) shows the highest identity to that of human. We however, did not find *Ciona* orthologs of the three non-vertebrate families. Five *Ciona* homologs clustered into three separate groups with no clear one-to-one orthologous relationship to the known families. However, among these, two *Ciona* pairs (220350/273922; 201119/287065) which are located basal to the main $G\alpha_{i/o}$ family grouping, share high sequence identities of about 59% to the human $G\alpha_{i1}$. One of these pairs, (220350/273922) shares an identity of 99% and appears to be a recent gene duplication event. The two genes have the same loci and are positioned in tandem to each other, separated by about 3500 nucleotides. These divergent *Ciona* specific Gα proteins were placed into a group designated 'Other'. One *Ciona* Gα sequence (281048) was found to be an outgroup to all the other clusters. Although this homolog retains all the critical and conserved motifs, it shows slight deviations in two consensus motifs, namely the GAGE motif at the N-terminal and the TCAT motif at the C-terminal regions of the sequence (Figure 2). This homolog also is the most divergent among the *Ciona* Gα proteins (26–34%) and shares an identity of 33% with the top human Gα match. Hence this homolog was placed into a group designated 'Unclassified'. The presence of only a single or two members in *Ciona*, belonging to each of the major Gα subunit families is in contrast to the presence of several members of each family in mammals. Moreover, a lower eukaryote like *Caenorhabditis elegans* has 20 Gα subunit genes out of which four belong to the four major mammalian families, while the remaining 16 are not homologous to the mammalian subunit and are a product of lineage specific gene expansion [25].

The Gα subunits in *Ciona* share a strong level of sequence conservation with those from other organisms including human, reflecting common structure-function relationships (Figure 2). The Gα subunit contains two domains, a Ras-like domain with a nucleotide-binding pocket and an all α-helical domain, composed of a six-helix bundle that makes up a lid on the nucleotide-binding pocket. Binding to GTP causes conformational changes within three flexible segments of Gα, namely "switch" regions (I–III) leading to a well ordered, GTP bound activated conformation with lowered affinity for the Gβγ complex and increased affinity for downstream effectors. Several crystal structures of Gα-effector complexes have revealed that a common site for effector interactions is provided by a highly conserved hydrophobic cleft

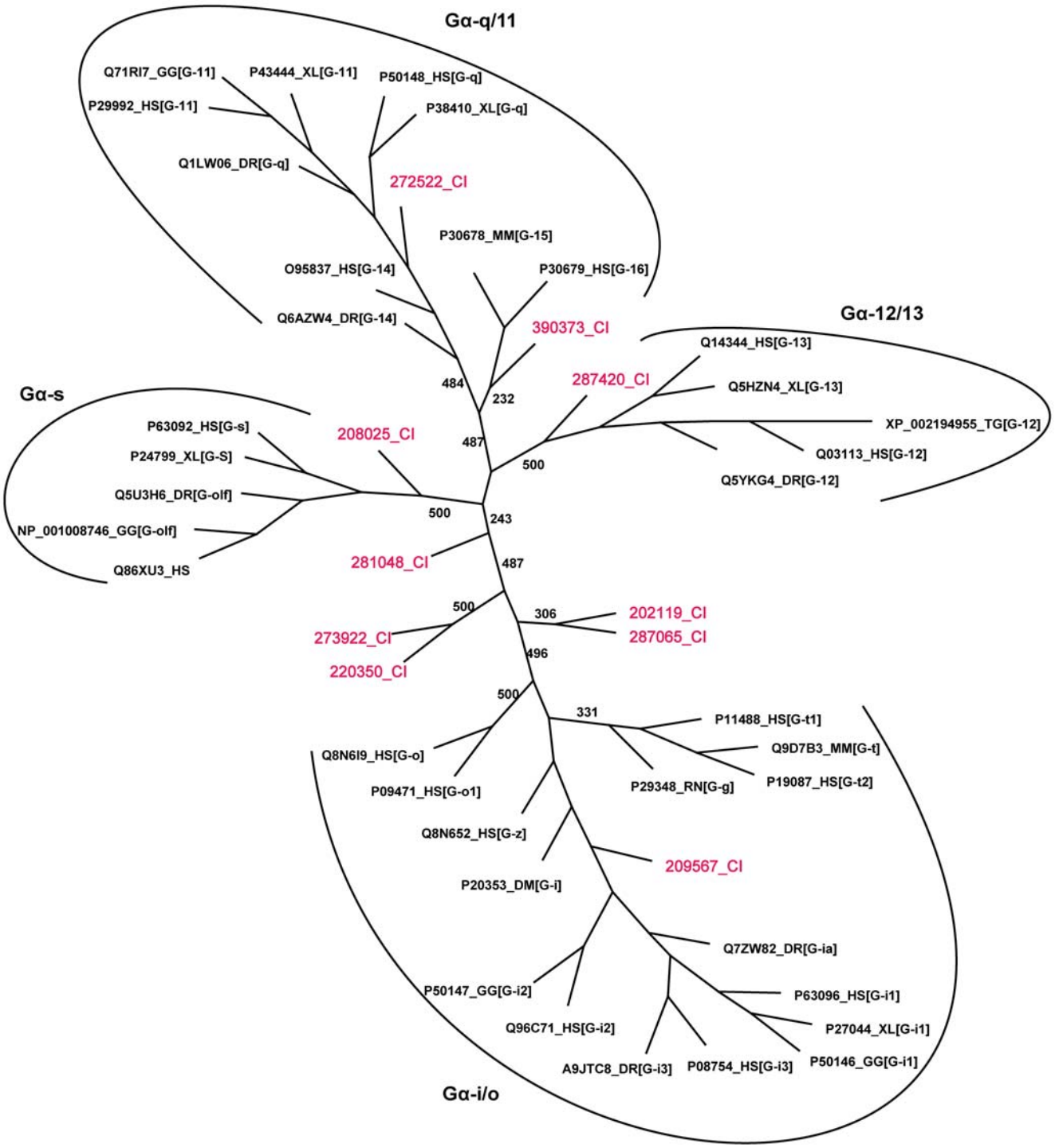

Figure 1. Phylogenetic relationship between Gα subunits in *Ciona* and other genomes. The figure illustrates the unrooted tree of Gα subunits in *Ciona* and representative organisms. Clusters that represent the four main families of Gα proteins are marked. Bootstrap support values (out of 500 data sets) are shown. Abbreviations used in this figure and figures 2,3,4,5a and 5b are HS- human, DM- *Drosophila melanogaster*, MM- mouse, RN- rat, CI- *Ciona*, CE- *C. elegans*, DR- *Danio rerio*, XL- *Xenopus laevis*, TG- *Taeniopygia guttata*, GG- chicken. Family names are given in square brackets.
doi:10.1371/journal.pone.0007349.g001

within active Gα, formed by the α2 and α3 helices. The region of sequences making up these helices are highly conserved between *Ciona* and human sequences, implying similar mechanisms of Gαβγ complex/Gα-effector recognition and binding. Interestingly, one *Ciona* homolog (390373), displays an unusual insertion of about 150 amino acids in the switch III loop that connects the α2 and α3 helices. This functional loop is ordered only in the active state of Gα. The triad of Gα residues that mediates GTP hydrolysis includes Thr181 and Arg178 in switch I that stabilize the γ-phosphate ion and Gln204 in switch II which coordinates

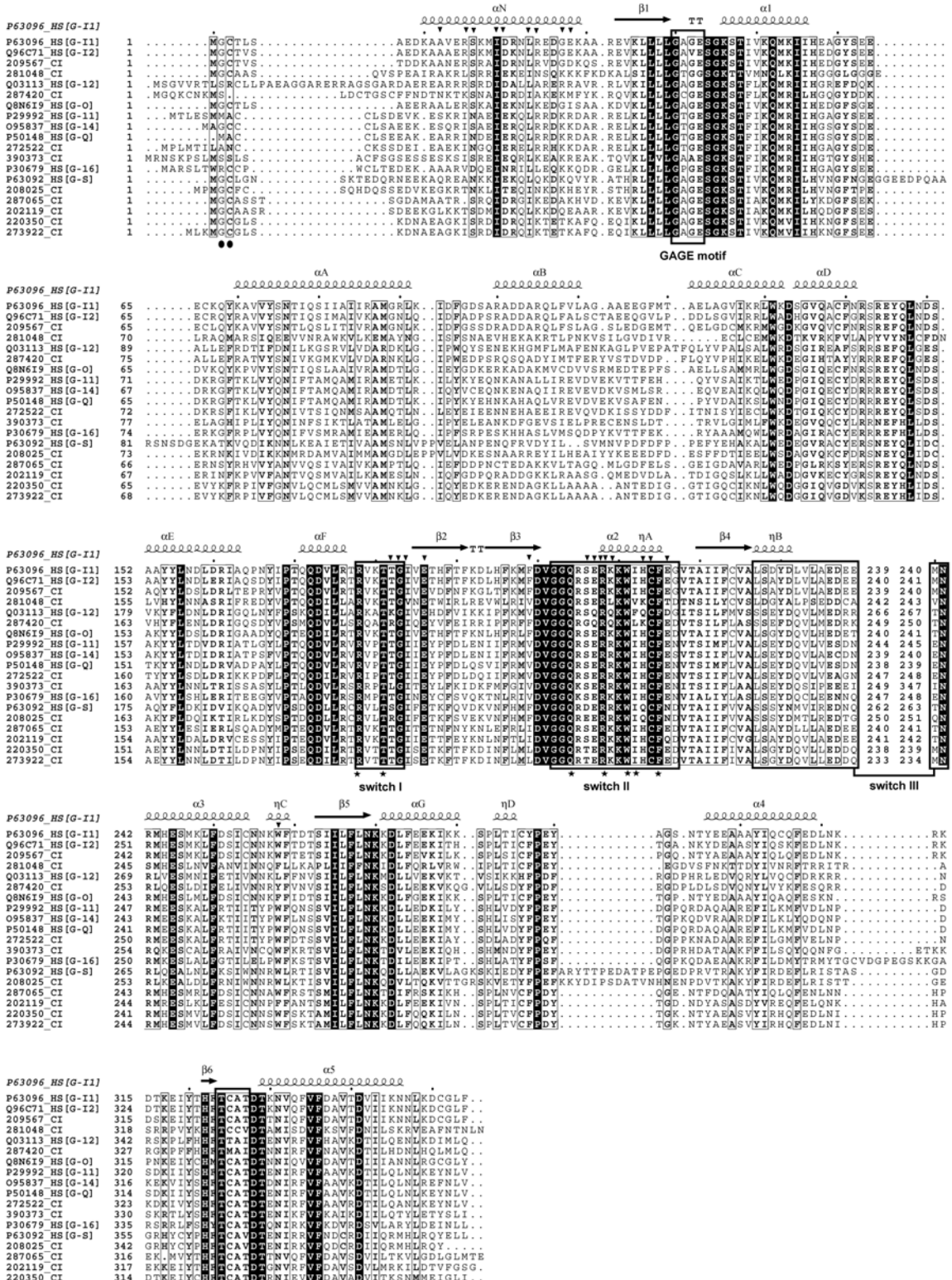

Figure 2. Multiple sequence alignment and secondary structure features of the Gα subunit. Secondary structure assignments are derived from mammalian $G\alpha_{i1}$ (PDB ID: 1GP2). The three conformationally flexible switch regions of Gα (including the entire α2 helix within switch II) are indicated in rectangular boxes; residues in Gα that contact Gβ are marked with inverted triangles above the ruler line. Conserved Gα guanine base and phosphate contact positions are indicated with rectangular boxes and are marked correspondingly (GAGE, DVGGQ motifs and TCAT motif); the conserved Arg, Thr and Glu residues involved in GTP hydrolysis and key switch II residues (Lys209, Trp211, Ile212, and Phe215) that interact with $G\beta_1$ are represented by stars.

doi:10.1371/journal.pone.0007349.g002

the critical nucleophilic water molecule responsible for hydrolysis of the γ-phosphate, is completely conserved (numbered as in human $G\alpha_{i1}$, Figure 2) [26].

The communication between activated GPCR and the Gα subunit is modulated primarily through the C terminus/α5 helix and α4/β6 loop that invoke further structural changes necessary for GDP release [27–29]. The loop connecting β6/α5 contains a completely conserved TCAT motif that stabilizes the binding of GDP. The release of GDP is triggered by the receptor contacts to the C terminus and through the α5 helix that further structurally modulate the TCAT motif. Other highly conserved regions of the Gα subunit involved in transmitting conformational changes of the TCAT motif includes the α3 helix, which connects the α3/β5 loop to switch III and the β6 strand [30–31,5]. The N-terminal region of the *Ciona* Gα subunits also contains highly conserved Cys and Gly residues that are sites of palmitoylation and myristoylation which anchor the subunits to the membrane thus regulating membrane localization and protein-protein interaction [32].

The Gβγ complex

A total of two homologs of the Gβ class were identified in *Ciona* with both homologs exhibiting the characteristic conserved set of WD repeat motifs. Figure 3 denotes the relationships of the repertoire of *Ciona* Gβ homologs. The Gβ class of proteins consists of a single family and subfamily that includes five types of Gβ proteins named Gβ(1–5). In the mammalian genome, the types 1 to 4 are highly conserved sharing greater than 80% sequence identity, but $G\beta_5$ is divergent, sharing only about 50% identity [33–34]. Our phylogenetic analysis indicates that one *Ciona* homolog (283613) exhibits a clear one-to-one orthologous relationship to the Gβ(1–4) group while the other sequence is orthologous to the $G\beta_5$ members (297548) (Figure 3). The two

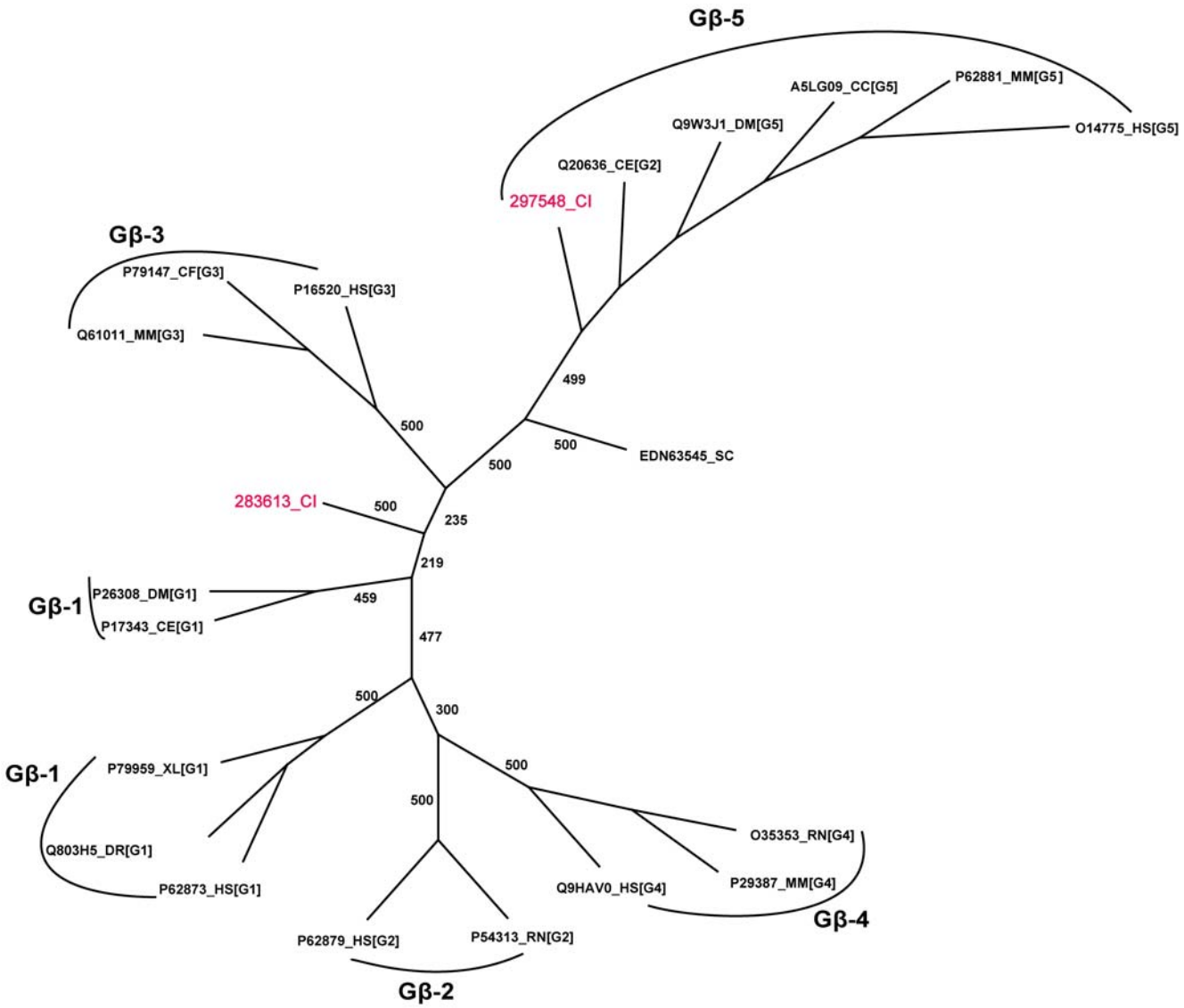

Figure 3. Phylogenetic relationship between Gβ subunits in *Ciona* and other genomes. The tree illustrates the unrooted tree of the Gβ subunits in *Ciona* and representative organisms. The five major clusters represent the five types of Gβ subfamilies. *Ciona* has only two types, in that, one is an ortholog of the $G\beta_5$ type while the other, an ortholog of the Gβ(1–4) type group, is equidistant to each of the four types. Bootstrap support values (out of 500 data sets) are shown. Abbreviations used are CC- *Cyprinus carpio*, CF- *Canis familiaris*, SC- for *Saccharomyces cerevisiae*. Gβ type names are given in square brackets.
doi:10.1371/journal.pone.0007349.g003

Ciona Gβ members share a sequence identity of about 41% between them and are more divergent than that in human. The *Ciona* Gβ(1–4) ortholog shares a high sequence identity of about 77% with that of the human Gβ(1–4) group while sharing about 40% identity with the human $G\beta_5$. The *Ciona* $G\beta_5$ ortholog shares a sequence identity of 51% with the human $G\beta_5$.

The *Ciona* genome holds two homologs of the Gγ class of proteins. It has to be noted here that a search for Gγ homologs in *Ciona* v.2.0 database resulted in the identification of just a single hit. However, parallel searches of the v1.0 database resulted in the identification of yet another homolog of the Gγ protein that has not been modeled and annotated in the v2.0 database. We have therefore annotated this sequence by the accession number corresponding to the *Ciona* v1.0 genome database. In both Gγ homologs, the protein models have been verified as having EST hits. In human, 12 Gγ gene products are known. Furthermore, among all heterotrimeric G protein subunits identified in *Ciona*, the Gγ homologs display the largest divergence to that of human. The *Ciona* Gγ sequences share the highest sequence identities of 40% and 34% to human Gγ type 7 and 12 members. Phylogenetic analysis of the Gγ members revealed that the *Ciona* Gγ subunits show no reliable clustering to any known Gγ types as classified in the gpDB database (Figure 4).

The Gβ subunit has an extended N terminal α-helix followed by a seven-bladed β-propeller fold that is composed of seven WD sequence repeats. The N-terminus is primarily supported by

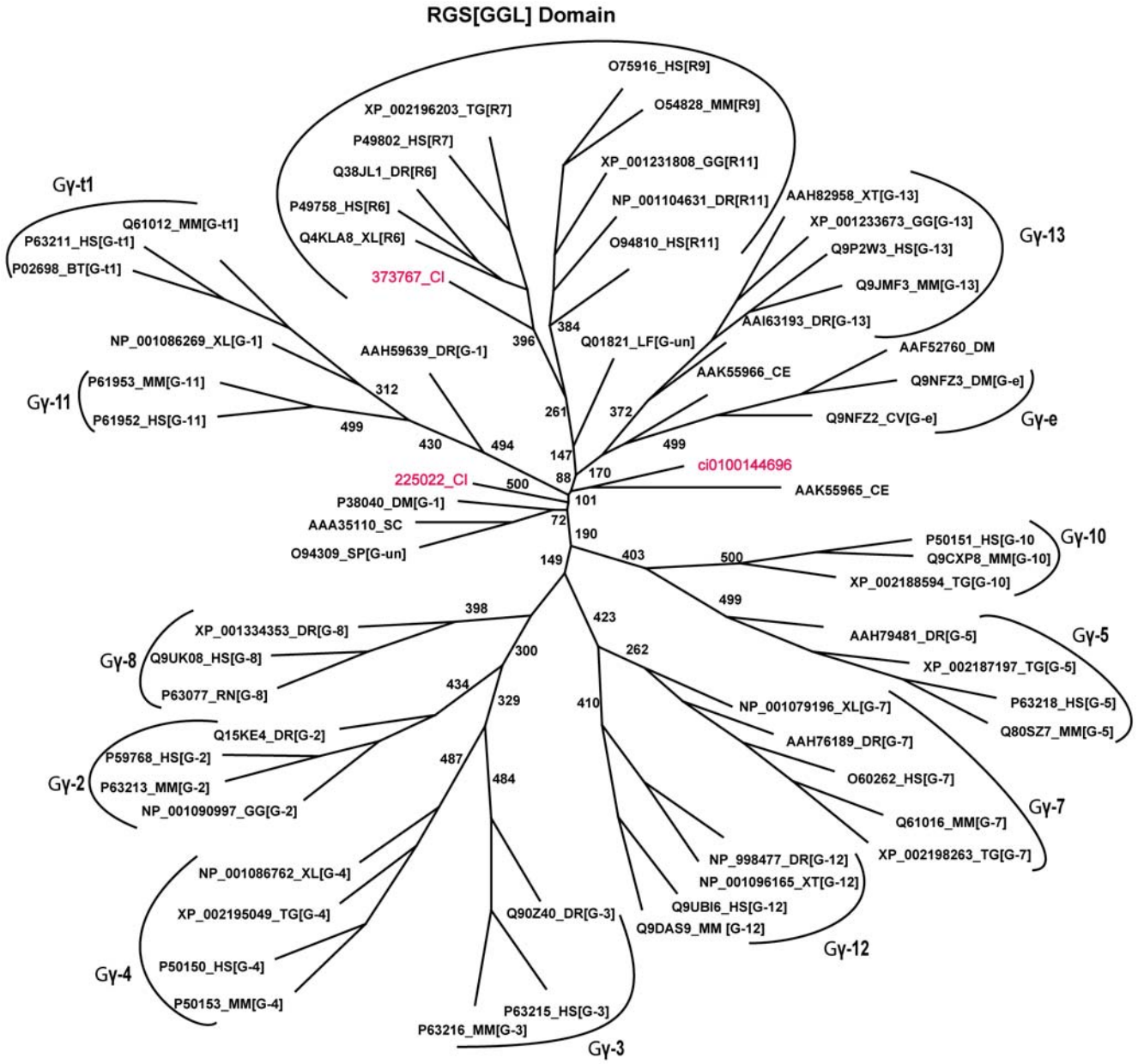

Figure 4. Phylogenetic relationship between Gγ subunit of *Ciona* and other genomes. The figure illustrates an unrooted tree of the Gγ subunits in *Ciona* and representative organisms. Different clusters represent the different types of Gγ subfamilies. The GGL domains form a separate significant cluster. The single Gγ subunit in *Ciona* is an outgroup to all the other members. Abbreviations used are LF- *Loligo forbesi*, SP- *Schizosaccharomyces pombe*, CV- *Calliphora vicina*, XT- *Xenopus tropicalis*. Gγ type names are given in square brackets. Un- refers to 'Unclassified' sequence.
doi:10.1371/journal.pone.0007349.g004

coiled-coil interactions with Gγ which is an extended stretch of two α-helices joined by a loop with additional interactions to loops that connect the CD and AB strands of Gβ blades 5 and 6 [35–36] (Figures 5a,b). The heterotrimeric Gα and Gβγ complex formation involves two major sites of interactions. Firstly, the extensive burial of the β3/α2 loop and α2 helix (switch II) of Gα within six of the seven WD repeats of Gβ. Numerous interactions involve a hydrophobic core formed by Trp211 of Gα and Trp99 of Gβ (numbering as in human $G\beta_1$) and forms the basis for Gβγ-mediated guanine nucleotide dissociation inhibitor activity and for Gβγ binding between Gα-GDP and Gβγ-effectors [5]. Several key switch II residues such as Lys209, Trp211, Ile212 and Phe215 that interact with $G\beta_1$ are completely conserved in *Ciona*. The other major region of Gα/Gβ interactions involves the WD1-2 of Gβ and the extended N-terminal helix of Gα (Figure 2, Figures 5a,b). There is also considerable overlap of the effector binding site on Gβγ with the region that interacts with the switch II of Gα [37]. Our sequence comparisons reveal that a majority of the functionally important residues of the Gβγ sequences are identical between *Ciona* and other organisms including human, indicating highly conserved modes of complex formation.

The selectivity and diversity of the Gβγ complex of mammalian G proteins arise from the numerous potential dimer combinations that can be assembled from a repertoire of 5 Gβ protein subunits and 12 Gγ subunits. The Gβ(1–4) subunits form a tight complex with Gγ subunits and can only be separated under denaturing conditions [38]. The sequence divergence between the Gβ(1–4) and the $G\beta_5$ types is reflected in the relatively weak interactions of the $G\beta_5$ with the Gγ subunits in contrast to that in the Gβ(1–4)γ complex. Moreover, unlike the other Gβ subunits which are generally found associated with the membrane, $G\beta_5$ exists in a significant amount in the soluble cell fraction [39]. It is now well established that $G\beta_5$ makes an *in-vivo* complex with the R7 subfamily of the RGS proteins in preference to Gγ and that the novel $G\beta_5$-RGS dimer formation is primarily mediated through interactions via a Gγ-like domain (GGL) present N-terminal to the RGS domain [34,40]. The identification of the $G\beta_5$ ortholog in *Ciona* also points to the fact that the *Ciona* genome most likely encodes homologs of GGL domain containing RGS proteins.

RGS proteins

The repertoire of human RGS family includes at least 37 members that are broadly classified into eight subfamilies. All members contain a core Gα-interacting RGS domain while they differ widely in their over all sizes and organization of several functional domains. RGS proteins are multifunctional and have the ability to negatively regulate Gα subunits and also function as effector antagonists. This multiple modular architecture enables regulatory selectivity and specificity in their mediation with diverse signaling partners that include all three components of the GPCR signaling pathway [41,6]. We identified 14 proteins in *Ciona* that possess the RGS domain (Table S1, Data S1). Eight of these are sequences that contain just the RGS domain at the C-terminal end with N-terminal regions having variable lengths ranging from 28 to 349 amino acid residues. It is to be noted here that 9 sequences identified from *Ciona* v2.0 database turned out to be fragments and in all these cases, the complete sequences corresponding to these members were identified in v1.0 database. One other fragment sequence (ci0100153349) was exclusively identified only from the v1.0 database with no corresponding protein model in v2.0. The complete sequence corresponding to this member was identified in the GenBank database. An examination of the RGS protein sequences for additional domains carried out using the Conserved Domain Database (CDD) indicated that six sequences contain additional domains that could confer further functionality [42]. One member has a DEP (Dishevelled, Egl-10 and Pleckstrin) domain N-terminal to the RGS domain, while two sequences contain a C2 domain in addition to the RGS domain. The RGS repertoire also includes one sequence that contains four domains that include an N-terminal PXA (Phox-associated) domain, an RGS domain followed by a PX (Phox) domain and a Nexin_C domain and another sequence with an RGS domain and a C-terminal DIX domain (Figure 6). Among these various domains, the DEP domain has been shown to play an important role in mediating the interaction of an RGS protein to the C-terminal tail of the GPCR, thus placing RGS in close proximity with its substrate Gα subunit [43]. The PX domain is an important phosphoinositide-binding module with varying lipid-binding specificities and implicated in cell signaling.

Our results further revealed that *Ciona* indeed does contain one RGS homolog (373767) that contains a GGL domain located between the DEP and RGS domains analogous to the domain organization that characterizes members of the R7 subfamily (RGS6, RGS7, RGS9, RGS11) of the human RGS family (Figure 6). This finding strongly supports the formation of an RGS complex with the $G\beta_5$ ortholog in *Ciona*. Sequence comparison of the GGL domain with that of the Gγ subunits from *Ciona* as well as other organisms indicates the homologous relationship of the GGL domain to the Gγ subunit (Figure 5b). This is further confirmed by the existence of significant monophyletic clades that separates the Gγ and GGL domains (Figure 4). The GGL domain of *Ciona* shares an identity of 17% and 19% with the Gγ subunit in *Ciona* and human, respectively and shares an identity of 28–49% with the four GGL domains in human. Recent structural studies show that the marked difference in specificities of $G\beta_5$ for the GGL domain versus Gγ arise from the cumulative effect of several minor changes in the corresponding sequences [44]. It is evident from our comparative analysis that the *Ciona* GGL domain containing RGS homolog could indeed interact with the $G\beta_5$ homolog in an analogous mode.

Heterotrimeric G protein complexes

The results of our analysis indicating a repertoire of only one homolog each of the Gβ(1–4) type and the $G\beta_5$ type subunits, two homologs of Gγ subunit and one homolog of the GGL containing RGS protein implies that only a maximum of five potential Gβ heterodimer complex combinations (two Gβ(1–4) γ complexes, two $G\beta_5\gamma$ complexes and one $G\beta_5$-RGS complex) likely exists in *Ciona*. Furthermore, with 10 Gα subunits, the *Ciona* genome potentially displays a limited repertoire of 50 Gαβγ heterotrimeric complexes. It is noteworthy that the diversity in the types of Gβγ complexes in *Ciona* is comparable to those seen in lower eukaryotic model organisms. For example, yeast has one Gβ subunit (Genbank ID: AAA35114) that is equally distant to the mammalian Gβ1-4 and the $G\beta_5$ types. *C. elegans* has two Gβ subunit sequences that are each orthologous to the Gβ1-4 and the $G\beta_5$ types. Furthermore, yeast has a single Gγ subunit (Genbank ID: AAA35110) that shows poor similarity to the human homolog, while *C. elegans* has two Gγ subunits (Genbank ID: AAK55965, AAK55966) that do not fall into any particular mammalian subfamilies (Figure 4).

Conclusions

Ciona appears to have a highly compact set of the heterotrimeric G protein complexes and these results complement the earlier findings of a compact set of GPCRs in the *Ciona* genome. The high levels of sequence similarity to the human orthologs indicate that they serve functions that are general to vertebrate signaling biology and that *Ciona* can indeed be used as a model organism to study

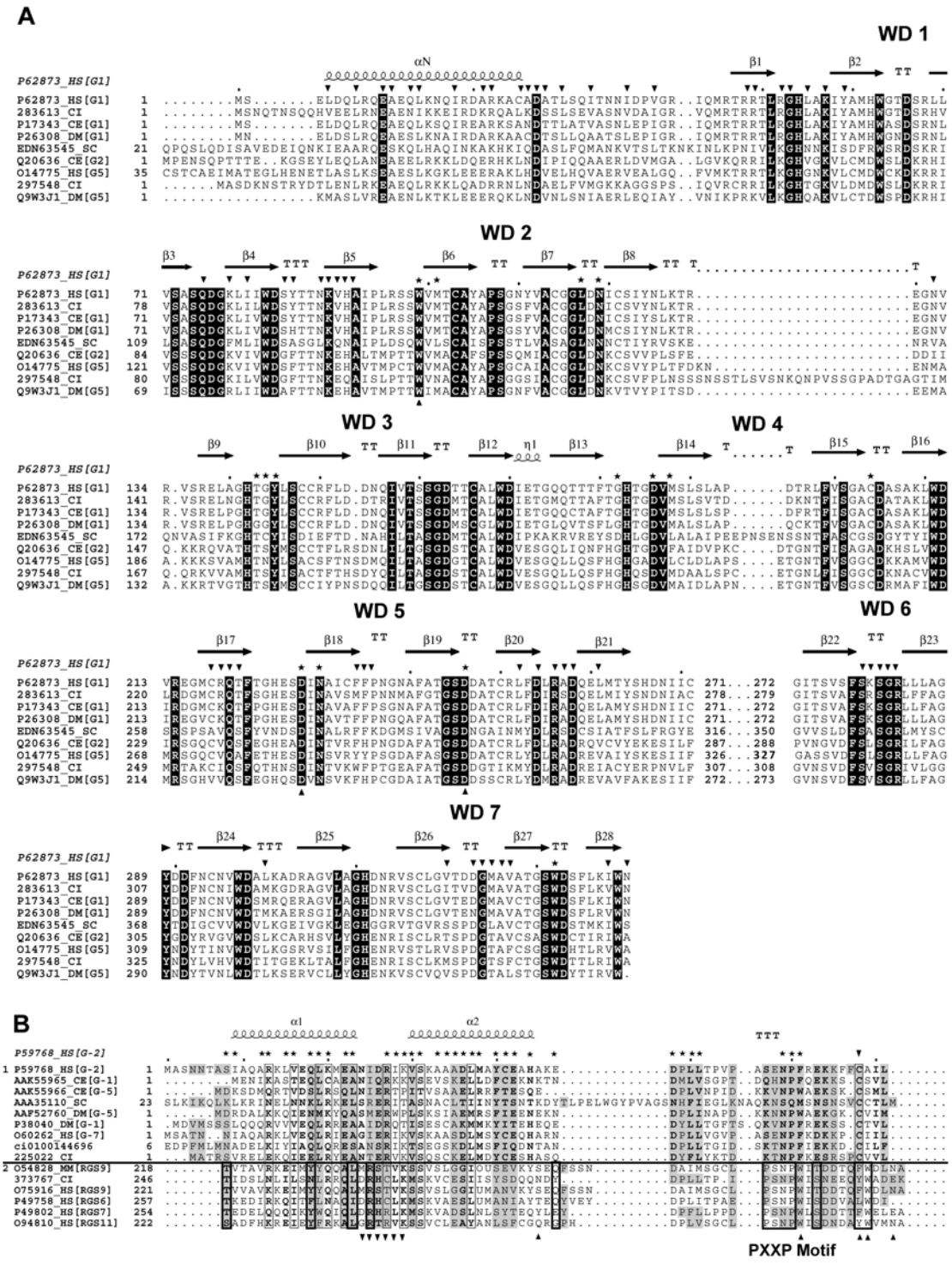

Figure 5. Multiple sequence alignments and secondary structure features of Gβ subunits, Gγ subunits and GGL domains. A. Alignment of the Gβ sequences from representative organisms. Secondary structure assignments (α-helices, β-strands) are derived from mammalian $G\beta_1$ (PDB ID: 1GP2). The four β-strands that comprise each of the seven WD repeat segments within Gβ subunits are marked with horizontal arrows to match the tertiary structure of $G\beta_1$. Residues in Gβ that contact Gγ are marked with inverted triangles and the contacts with the switch regions in Gα are indicated with stars. Key $G\beta_1$ residues (Trp99, Asp228, and Asp246) interacting with the $G\alpha_{i1}$ switch II helix are represented with upright triangles. B. Alignments of the Gγ subunits and GGL domains from representative organisms. Secondary structure assignments are derived from mammalian $G\gamma_2$ (PDB ID: 1GP2). Gγ subunit–like (GGL) domains and Gγ subunits are structurally equivalent. Residues in Gγ that contact Gβ are marked with stars. Bold letters indicate highly conserved residues. The inverted triangle in the C-terminal of Gγ sequences represents prenylation site. The horizontal line in the alignment separates the two blocks of Gγ and the RGS sequences (GGL domain region). The black vertical boxes in the lower block show conserved residues specific to the GGL sequences. The inverted triangles corresponds to residues in the RGS9 of human and mouse that make contacts with Gα subunit, while the upright triangles corresponds to the specific contacts with the $G\beta_5$ subunit.
doi:10.1371/journal.pone.0007349.g005

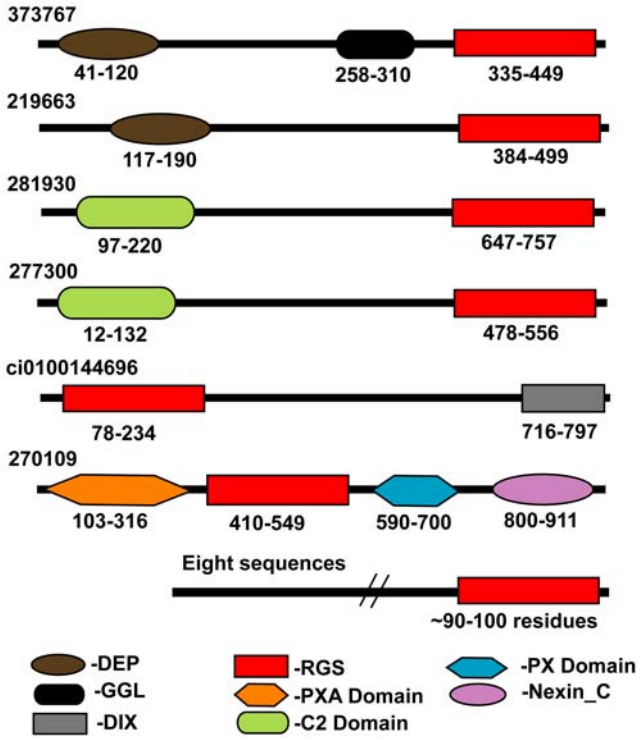

Figure 6. Domain organization of the repertoire of RGS proteins in *Ciona*. Eight *Ciona* sequences have just the RGS domain with varying lengths of the N-terminal region, while six sequences have other types of domains as well. DEP (Dishevelled, Egl-10, and Pleckstrin domain); RGS (Regulator of G protein signaling domain); GGL (Gγ like domain), Nexin_C (Sorting nexin C terminal domain); PX (Phox domain).
doi:10.1371/journal.pone.0007349.g006

vertebrate GPCR signaling events. Extensive experimental characterization of the developmental regulatory network in *Ciona* points to the possibilities of similar studies for GPCR signaling pathways [45]. The clear one-to-one orthology between human and *Ciona* G protein repertoires that extends at the levels of protein family, subfamily and type, indicates a strong conservation in the evolution of the mechanistic basis of specificity and diversity of the chordate GPCR signaling pathways. Thus, the limited repertoire of heterotrimeric G protein complexes in *Ciona* provides an unique opportunity to extend the predictions to guide experimental studies to explore the basis of receptor activation, specificity and selectivity, phenotypic effects of GPCR signaling events, reinforce functional importance of existing receptor G protein/RGS protein interactions and for identification of new partners and downstream targets.

Supporting Information

Data S1 Amino acid sequences of heterotrimeric G proteins and RGS proteins used as queries in BLAST and HMM search programs (FASTA format)
Found at: doi:10.1371/journal.pone.0007349.s001 (0.09 MB TXT)

Data S2 Amino acid sequences of heterotrimeric G proteins and RGS proteins in *Ciona* (FASTA format)
Found at: doi:10.1371/journal.pone.0007349.s002 (0.01 MB TXT)

Table S1 Accession numbers and classification of *Ciona* G protein and RGS protein sequences and human orthologs.
Found at: doi:10.1371/journal.pone.0007349.s003 (0.05 MB DOC)

Table S2 Accession numbers for EST-hits of the *Ciona* heterotrimeric G proteins and RGS proteins in the Unigene database
Found at: doi:10.1371/journal.pone.0007349.s004 (0.03 MB XLS)

Acknowledgments

The authors wish to acknowledge support from the Bioinformatics Infrastructure Facility, IIT Madras.

Author Contributions

Conceived and designed the experiments: NM. Performed the experiments: RP. Analyzed the data: RP NM. Wrote the paper: NM.

References

1. Cabrera-Vera TM, Vanhauwe J, Thomas TO, Medkova M, Preininger A, et al. (2003) Insights into G protein structure, function, and regulation. Endocr Rev 24: 765–781.
2. Lambright DG (2004) Heterotrimeric G protein signaling at atomic resolution. In: Bradshaw RA, Dennis EA, eds. Hand book of Cell Signaling. Boston: MA Academic. pp 575–579.
3. Preininger AM, Hamm HE (2004) G protein signaling: insights from new structures. Sci STKE 218: re3.
4. Birnbaumer L (2007) Expansion of signal transduction by G proteins. The second 15 years or so: from 3 to 16 α subunits plus βγ dimmers. Biochim Biophys Acta 1768: 772–793.
5. Johnston AC, Siderovski PD (2007) Receptor-Mediated Activation of Heterotrimeric G-Proteins: Current Structural Insights. Mol Pharmacol 72: 219–230.
6. Burchett SA (2000) Regulators of G protein signaling: a bestiary of modular protein binding domains. J Neurochem 75: 1335–1351.
7. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63: 1256–1272.
8. Clapham DE, Neer EJ (1997) G protein βγ subunits. Annu Rev Pharmacol Toxicol 37: 167–203.
9. Wettschureck N, Offermanns S (2005) Mammalian G Proteins and Their Cell Type Specific Functions. Physiol Rev 85: 1159–1204.
10. Graf R, Mattera R, Codina J, Evans T, Ho YK, et al. (1992) Studies on the interaction of alpha subunits of GTP-binding proteins with beta gamma dimers. Eur J Biochem 210: 609–619.
11. Corbo CJ, Di Gregorio A, Levine M (2001) The ascidian as a model organism in developmental and evolutionary biology. Cell 106: 535–538.
12. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298: 2157–2167.
13. Satoh N, Satou Y, Davidson B, Levine M (2003) *Ciona intestinalis*: an emerging model for whole-genome analyses. Trends Genet 7: 376–381.
14. Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439: 965–968.
15. Kamesh N, Aradhyam GK, Manoj N (2008) The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. BMC Evo Biol 8: 129.
16. *Ciona intestinalis* v2.0 Sequencing Project database. (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html).
17. *Ciona intestinalis* v1.0 Sequencing Project database.(http://genome.jgi-psf.org/ciona4/ciona4.home.html).
18. Theodoropoulou MC, Bagos PG, Spyropoulos IC, Hamodrakas SJ (2008) gpDB: a database of GPCRs, G-proteins, effectors and their interactions. Bioinformatics 24: 1471–1472.
19. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
20. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755–763.
21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24: 4876–4882.
22. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/N. Nucleic Acids Symp 41: 95–98.
23. Felsenstein J (1989) PHYLIP, Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.

24. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15: 305–308.
25. Jansen G, Thijssen KL, Werner P, van der Horst M, Hazendonk E, et al. (1999) The complete family of genes encoding G proteins of *Caenorhabditis elegans*. Nat Genet 21: 414–419.
26. Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, et al. (1994) Structures of active conformations of Gi alpha1 and the mechanism of GTP hydrolysis. Science 265: 1405–1412.
27. Onrust R, Herzmark P, Chi P, Garcia PD, Lichtarge O, et al. (1997) Receptor and $\beta\gamma$ binding sites in the α subunit of the retinal G protein transducin. Science 275: 381–384.
28. Sprang SR (1997) G protein mechanisms: insights from structural analysis. Annu Rev Bioche 66: 639–678.
29. Marin EP, Krishna AG, Sakmar TP (2002) Disruption of the alpha5 helix of transducin impairs Rhodopsin-catalyzed nucleotide exchange. Biochemistry 41: 6988–6994.
30. Marsh SR, Grishina G, Wilson PT, Berlot CH (1998) Receptor-mediated activation of $G_{s\alpha}$: evidence for intramolecular signal transduction. Mol Pharmacol 53: 981–990.
31. Grishina G, Berlot CH (2000) A surface-exposed region of $G_{s\alpha}$ in which substitutions decrease Receptor-mediated activation and increase receptor affinity. Mol Pharmacol 57: 1081–1092.
32. Resh MD (1996) Regulation of cellular signalling by fatty acid acylation and prenylation of signal transduction proteins. Cell Signal 8: 403–412.
33. Watson AJ, Aragay AM, Slepak VZ, Simon MI (1996) A novel form of the G protein beta subunit Gbeta5 is specifically expressed in the vertebrate retina. J Biol Chem 271: 28154–28160.
34. Jones MB, Siderovski DP, Hooks SB (2004) The G betagamma dimer as a novel source of selectivity in G-protein signaling: GGL-ing at convention. Mol Interv 4: 200–214.
35. Wall AM, Coleman ED, Lee E, Iniguez-Lluhi AJ, Posner AB, et al. (1995) The Structure of the G Protein Heterotrimer $G_{i\alpha 1\beta 1\gamma 2}$,. Cell 83: 1047–1058.
36. Sondek J, Siderovski DP (2001) Gγ-like (GGL) domains: new frontiers in G-protein Signaling and β-propeller scaffolding. Biochem Pharmacol 61: 1329–1337.
37. Davis TL, Bonacci TM, Sprang SR, Smrcka AV (2005) Structural and molecular characterization of a preferred protein interaction surface on G protein beta gamma subunits. Biochemistry 44: 10593–10604.
38. Higgins JB, Casey PJ (1994) *In vitro* processing of recombinant G protein gamma subunits. Requirements for assembly of an active beta gamma complex. J Biol Chem 269: 9067–9073.
39. Liang JJ, Cockett M, Khawaja XZ (1998) Immunohistochemical localization of G protein β1, β2, β3, β4, β5, and γ3 subunits in the adult rat brain. J Neurochem 71: 345–355.
40. Yost AE, Mervine MS, Sabo LJ, Hynes RT, Berlot HC (2007) Live cell analysis of G protein β5 complex formation, function and targeting. Mol Pharmacol 72: 812–825.
41. Snow BE, Krumins AM, Brothers GM, Lee SF, Wall MA, et al. (1998) A G protein γ-subunit-like domain shared between RGS11 and other RGS proteins specifies binding to $G\beta_5$ subunits. Proc Natl Acad Sci U S A 95: 13307–13312.
42. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. (2009) CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 37(Database issue): D205–10.
43. Chen S, Hamm HE (2006) DEP domains: More than just membrane anchors. Dev Cell 11: 436–438.
44. Cheever ML, Snyder JT, Gershburg S, Siderovski DP, Harden TK, et al. (2008) Crystal structure of the multifunctional Gbeta5-RGS9 complex. Nat Struct Mol Biol 15: 155–162.
45. Imai KS, Levine M, Satoh N, Satou Y (2006) Regulatory blueprint for a chordate embryo. Science 5777: 1183–1187.