

Research article

Open Access

## The domain architecture of large guanine nucleotide exchange factors for the small GTP-binding protein Arf

Barbara Mouratou<sup>†1</sup>, Valerie Biou<sup>†1</sup>, Alexandra Joubert<sup>1</sup>, Jean Cohen<sup>2</sup>, David J Shields<sup>3</sup>, Niko Geldner<sup>4,5</sup>, Gerd Jürgens<sup>4</sup>, Paul Melançon<sup>3</sup> and Jacqueline Cherfils\*<sup>1</sup>

Address: <sup>1</sup>Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, avenue de la Terrasse, 91198 Gif sur Yvette cedex, France, <sup>2</sup>Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette, France, <sup>3</sup>Department of Cell Biology, University of Alberta, Edmonton, Canada, <sup>4</sup>Center of Plant Molecular Biology, Universitaet Tuebingen, Tuebingen, Germany and <sup>5</sup>Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, USA

Email: Barbara Mouratou - mouratou@lebs.cnrs-gif.fr; Valerie Biou - biou@lebs.cnrs-gif.fr; Alexandra Joubert - joubert@lebs.cnrs-gif.fr; Jean Cohen - cohen@cgm.cnrs-gif.fr; David J Shields - david.shields@ualberta.net; Niko Geldner - geldner@salk.edu; Gerd Jürgens - gerd.juergens@uni-Tuebingen.de; Paul Melançon - paul.melancon@ualberta.ca; Jacqueline Cherfils\* - cherfils@lebs.cnrs-gif.fr

\* Corresponding author †Equal contributors

Published: 17 February 2005

Received: 04 November 2004

BMC Genomics 2005, 6:20 doi:10.1186/1471-2164-6-20

Accepted: 17 February 2005

This article is available from: <http://www.biomedcentral.com/1471-2164/6/20>

© 2005 Mouratou et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Small G proteins, which are essential regulators of multiple cellular functions, are activated by guanine nucleotide exchange factors (GEFs) that stimulate the exchange of the tightly bound GDP nucleotide by GTP. The catalytic domain responsible for nucleotide exchange is in general associated with non-catalytic domains that define the spatio-temporal conditions of activation. In the case of small G proteins of the Arf subfamily, which are major regulators of membrane trafficking, GEFs form a heterogeneous family whose only common characteristic is the well-characterized Sec7 catalytic domain. In contrast, the function of non-catalytic domains and how they regulate/cooperate with the catalytic domain is essentially unknown.

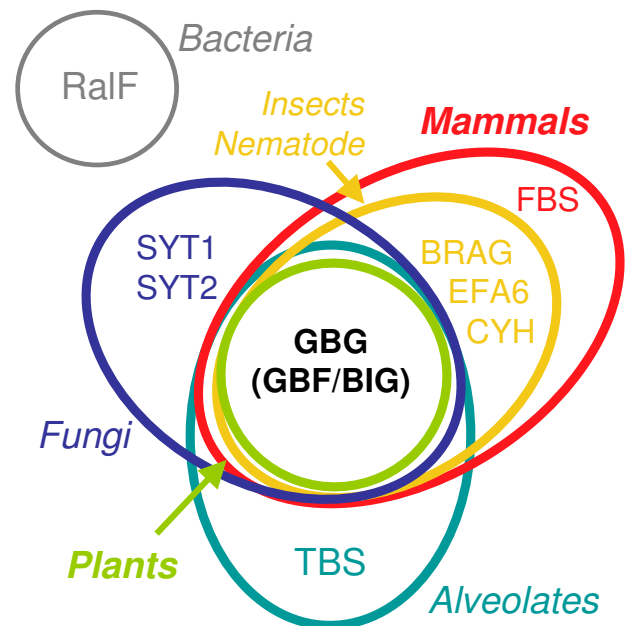
**Results:** Based on Sec7-containing sequences from fully-annotated eukaryotic genomes, including our annotation of these sequences from Paramecium, we have investigated the domain architecture of large ArfGEFs of the BIG and GBF subfamilies, which are involved in Golgi traffic. Multiple sequence alignments combined with the analysis of predicted secondary structures, non-structured regions and splicing patterns, identifies five novel non-catalytic structural domains which are common to both subfamilies, revealing that they share a conserved modular organization. We also report a novel ArfGEF subfamily with a domain organization so far unique to alveolates, which we name TBS (TBC-Sec7).

**Conclusion:** Our analysis unifies the BIG and GBF subfamilies into a higher order subfamily, which, together with their being the only subfamilies common to all eukaryotes, suggests that they descend from a common ancestor from which species-specific ArfGEFs have subsequently evolved. Our identification of a conserved modular architecture provides a background for future functional investigation of non-catalytic domains.

## Background

Guanine Nucleotide Exchange Factors (GEFs) are obligatory components of signaling cascades regulated by small GTP-binding proteins (called small G proteins hereafter). Their biochemical activity is to stimulate the dissociation of the tightly bound GDP nucleotide from the small G protein in response to cellular signals. Thereby, they favor the binding of the more abundant cellular GTP, organizing the active conformation of the small G protein which can recruit its effectors (reviewed in [1]). Each small G protein family features its own ensemble of GEFs characterized by a conserved catalytic domain responsible for nucleotide exchange, which is generally combined with non-catalytic domains that define the spatio-temporal conditions of activation. In the case of small G proteins of the Arf family, which are major regulators in membrane trafficking (reviewed in [2]), the exchange domain is a conserved module of ~200 amino acids called the Sec7 domain [3]. Its biochemical (reviewed in [4]) and structural [5,6] mechanisms have been investigated in detail. Remarkably, the Sec7 domain is the only domain that is conserved in all ArfGEFs (reviewed in [7,8]) and it is to some extent interchangeable between species [9]. In contrast, little is known about the functions of the other domains, which are likely to determine intracellular localization of ArfGEFs and their responsiveness to specific signals.

As for most small G proteins, Arf family members are outnumbered by ArfGEFs in many species. In humans for instance, 5 Arf proteins have been identified, and there are at least 13 proteins carrying a Sec7 domain, of which most have been characterized as *bona fide* ArfGEFs (reviewed in [7,8]). Thus an individual Arf protein may be activated by more than one GEF, emphasizing that essential aspects in building up the Arf responses may be encoded by the modular architecture of their GEFs. Sequence similarity in the non-catalytic regions forms the basis for the classification of ArfGEFs into subfamilies. 8 subfamilies are currently identified in eukaryotes with sizes ranging from small (~40–80 kD including CYH, EFA6 and FBS), to medium (~100–150 kD, including BRAG/LONER, SYT1, SYT2) and large (~170–200 kD) ArfGEFs (reviewed in [7,8]). Large ArfGEFs comprise two subfamilies which we will refer to as the BIG and GBF subfamilies after the name of their human representatives. The GBF subfamily includes human GBF1 [10], Arabidopsis GNOM [11] and Saccharomyces Gea1 and Gea2 [12], the BIG subfamily human BIG1 and BIG2 [13,14] and yeast Sec7p [15]. An additional subfamily called RalF is found in Rickettsia and Legionella bacteria, likely acting on an host Arf pathway [16]. Analysis of the CYH and EFA6 subfamilies, present only in multicellular animals, and that of the large ArfGEFs, found in all eukaryotes, have yielded most of the functional data currently available. CYH and EFA6 are



**Figure 1**  
**Venn diagram of the nine Sec7-containing subfamilies sorted according to the species where each subfamily has been found.** The TBS subfamily was identified in this study. The BIG and GBF subfamilies are merged in a higher order subfamily (GBG), and are the only subfamily common to all eukaryotes.

active on Arf6 at the plasma membrane where they may function in the crosstalk of membrane traffic, cytoskeleton dynamics and signalling in endosomal pathways (reviewed in [17]). Most members of the BIG and GBF subfamilies characterized so far function in vesicular trafficking at the Golgi [12,14,18], except for BIG2, which also localizes on recycling endosomes [19], and GNOM which acts in the endosomal recycling pathway [11].

The domain architecture of non-catalytic regions of ArfGEFs, hence their contribution to specific aspects of the build-up of the Arf response, is essentially not established except for those ArfGEFs with domains found in other classes of cellular regulators. The known domains include membrane-interacting PH domains in the CYH (reviewed in [20]), EFA6 [21] and possibly BRAG/LONER[22] subfamilies, and a putative F-box in the FBS subfamily [23], a protein-protein interaction domain that has been involved in the recruitment of substrates to the SCF ubiquitination machinery. Coiled-coil structures have also been predicted in the N-terminus of the CYH subfamily and in the C-terminus of the EFA6 subfamily. In CYH,

**Table 1: BIG and GBF protein sequences used in this study.**

	Species	Protein name <sup>a</sup>	Accession Number	Size in amino acids
Metazoa	Ag	Q7PWN5	EAA14874	1522
		<b>Q7PXQ7</b>	<b>EAA00837</b>	<b>1285</b>
	Ce	Q9XWG5	NP_493386	1628
		<b>Q9XTF0</b>	<b>NP_499522</b>	<b>1820</b>
	Dm	Q9VJW1	AAF53331	1653
		<b>Q9V696</b>	<b>AAF58532</b>	<b>1983</b>
	Hs	BIG1	Q9Y6D6	1849
		BIG2	Q9Y6D5	1785
		<b>GBF1</b>	<b>Q92538</b>	<b>1859</b>
	Rn	BIG1	XP_232614	1987
BIG2		Q7TSU1	1791	
<b>GBF1</b>		<b>XP_347197</b>	<b>1883</b>	
Fungi	Ca	EAL04295	EAL04295	1839
		<b>EAL02873</b>	<b>EAL02873</b>	<b>1015</b>
	Nc	Q7SAX4	EAA33549	1940
		<b>Q7SAL8</b>	<b>EAA33457</b>	<b>1626</b>
	Sc	SEC7	PI11075	2009
		<b>GEA1</b>	<b>P47102</b>	<b>1408</b>
		<b>GEA2</b>	<b>P39993</b>	<b>1459</b>
	Sp	SC71	Q9UT02	1811
		SC72	Q9P7V5	1822
		<b>Q9P7R8</b>	<b>NP_596613</b>	<b>1462</b>
Viridiplantae	At	At1g01960	Q9LPC5	1750
		At3g43300	NP_189916	1728
		At3g60860	Q9LZX8	1793
		At4g35380	O65490	1711
		At4g38200	NP_195533	1698
		<b>GNOM</b>	<b>Q42510, At1g13980</b>	<b>1451</b>
		<b>GNLI</b>	<b>Q9FLY5, At5g39500</b>	<b>1443</b>
	Os	<b>GNL2</b>	<b>NP_197462, At5g19160</b>	<b>1375</b>
		9631.m01366	Q8S565	1789
		9630.m00920	Q9XGN9	1687
		9634.m04029	-	1704
		9635.m03752	-	1680
		<b>9631.m04495</b>	-	<b>1456</b>
		<b>9630.m02122</b>	-	<b>1396</b>
		<b>9632.m00175</b>	<b>Q7XT11</b>	<b>1407</b>
		Alveolata	Pt	GGG1
GGG2	CR533424			1628
GGG3	CR533423			1598
GGG4	CR533422			1599
GGG5	CR533421			1435

<sup>a</sup> Unnamed sequences are designated by their NCBI accession number, AGI (Arabidopsis Genome Initiative) locus numbers for At and TIGR model temporary IDs for Os. BIG and GBF subfamily members are in normal and bold characters respectively, except for Pt members which have not been assigned to either subfamily (see also Figure 8). Species abbreviations are: Ag, *Anopheles gambiae*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Rn, *Rattus norvegicus*; Ca, *Candida albicans*; Nc, *Neurospora crassa*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Pt, *Paramecium tetraurelia*).

they are involved in dimerization [3], recruitment of partners [24] and Golgi targeting [25], and in actin remodeling functions in the case of EFA6 [21]. On the other hand, although the functions of BIG and GBF subfamilies have

been the subject of many investigations, their architecture is barely described, making it difficult to associate biochemical activities with their molecular structure.

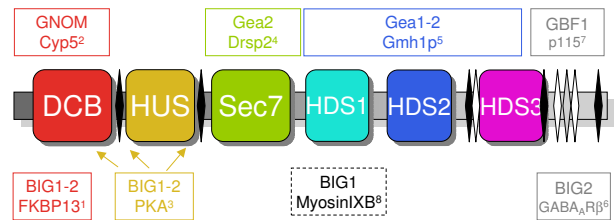
Here we investigate the domain architecture in the BIG and GBF subfamilies, including all sequences from fully annotated eukaryotic genomes and our novel annotation of Sec7-containing proteins from the *Paramecium tetraurelia* alveolate. Sequence comparisons combined with secondary structures and splicing patterns analysis identifies five novel domains that are conserved between BIG and GBF subfamilies, thus unifying them as a higher order subfamily with a probable common ancestor. Our analysis of Sec7-domain containing sequences from *Paramecium* also introduces a novel subfamily of ArfGEFs unique to alveolates, which we call TBS (TBC-Sec7).

## Results and discussion

### A conserved domain architecture in BIG and GBF subfamilies

The BIG and GBF subfamilies are the only ArfGEFs subfamilies common to all eukaryotes [8] and the sole ArfGEFs present in plants [26] (Figure 1). They are therefore possible representatives of ancestral ArfGEF functions and may provide a model to understand the nature and implementation of activities associated with the exchange function carried by the conserved Sec7 domain. However, domain 'hunting' in BIG and GBF subfamilies was complicated by the facts that the Sec7 domain is their only domain that could be identified from known domain repertoires, and that their poorly characterized non-catalytic regions were not found outside these ArfGEF subfamilies. Alternatively, we based our search of candidate structural domains in BIGs and GBFs on the bioinformatics analysis of their own sequences, taking advantage of the growing number of sequences from fully annotated genomes from mammals, insects, plants, nematode, and fungi, to which we included our annotation of Sec7-containing proteins from the newly sequenced genome of *Paramecium*.

Multiple alignments of 42 sequences (listed in Table 1) revealed that the BIG and GBF subfamilies share an unexpected conserved architecture (schematized in Figure 2). Two homology domains are located in N-terminus of the Sec7 domain – the DCB (~150 aa) and HUS (Homology Upstream of Sec7, ~170 aa) domains – and three in its C-terminus – the HDS1 (Homology Downstream of Sec7, ~130 aa), HDS2 (~160 aa) and HDS3 (~120 aa) domains (Figure 3,4,5,6,7). In *Arabidopsis* GNOM, the DCB domain is included in an N-terminal region of ~250 residues involved in dimerization and possibly binding to cyclophilin5 and called the Dimerization/Cyclophilin Binding region [27], after which the new domain was named (Figure 3). All domains are predicted to have a high content of  $\alpha$ -helices that co-align in the multiple sequence alignments, reinforcing the prediction of sequence similarities and suggesting that these domains form folded structural units that may share common functional features. Except for the N-terminal DCB domain

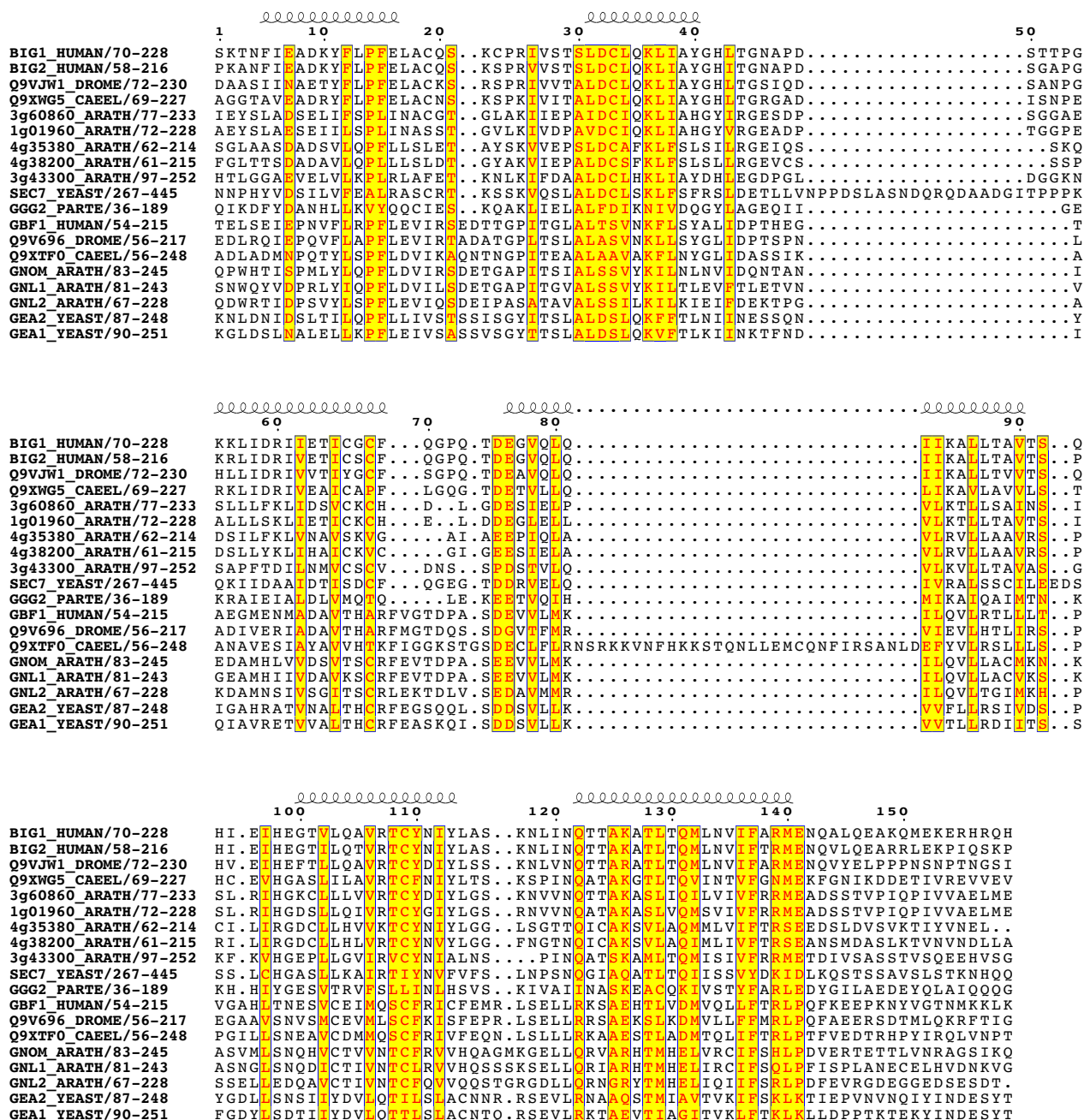


**Figure 2**

**The common domain architecture of the BIG and GBF subfamilies.** From N- to C-terminus : DCB , HUS, Sec7, HDS1, HDS2, HDS3. Linker regions of variable length and sequence are shown in grey, with alternate splicing sites in human GBF1, BIG1 and BIG2 in black, white and grey diamond shapes respectively. Interactions reported in the literature are indicated in boxes of width corresponding to the mapped regions, except for myosin IXb interaction which was studied only with full-length BIG1. Arrows indicate predicted Protein kinase A-anchoring motifs. <sup>1</sup> [45]; <sup>2</sup> [27]; <sup>3</sup> [46]; <sup>4</sup> [47]; <sup>5</sup> [48]; <sup>6</sup> [49]; <sup>7</sup> [50]; <sup>8</sup> [51].

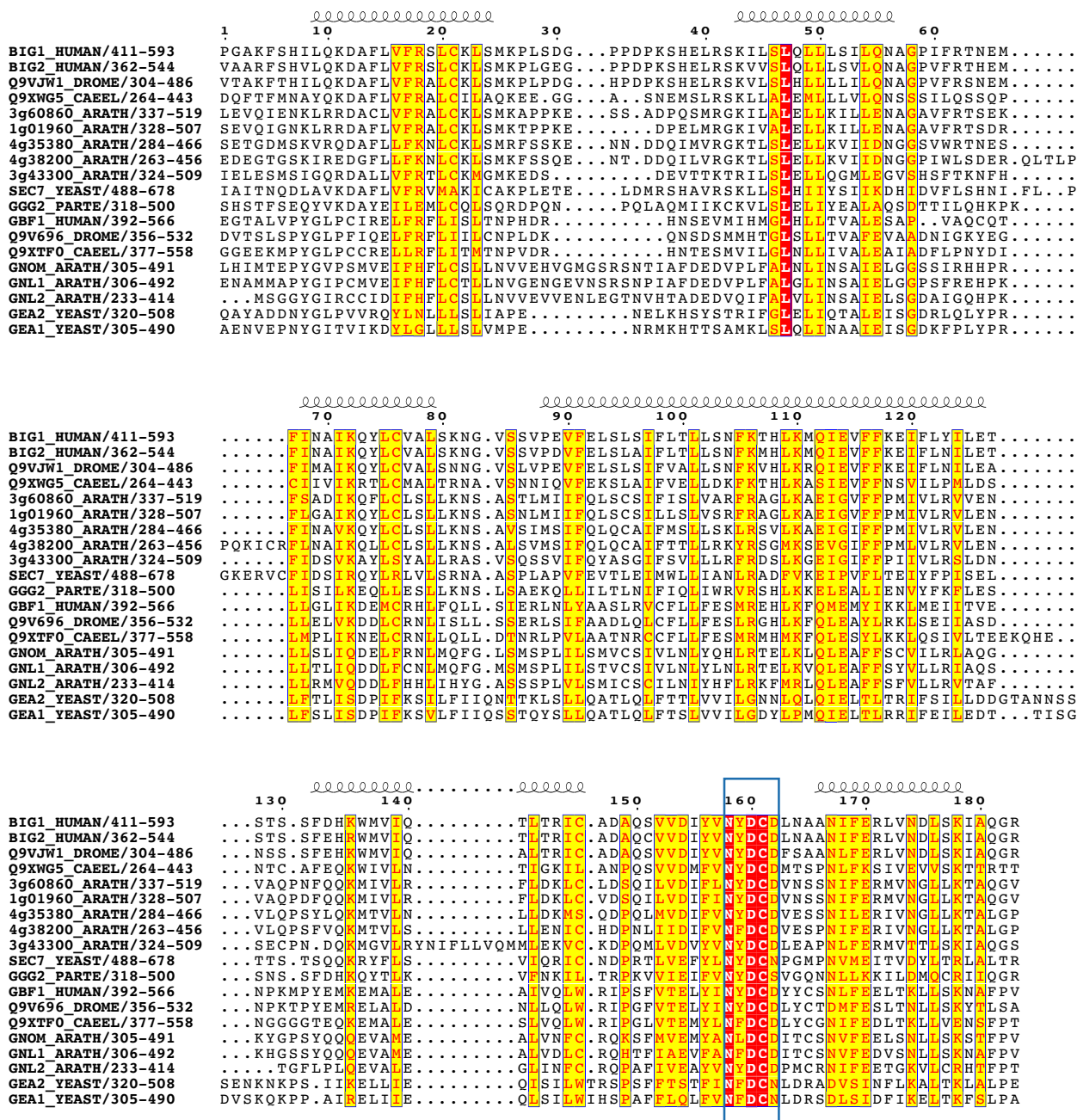
which is also found in the yeast protein Ysl2p [28], all of them are unique to these two ArfGEFs subfamilies within the detection limits of the BLAST search. The HUS domain features a remarkably conserved N(Y/F)DC(D/N) motif, which we call the HUS box, which is predicted to locate in a loop where it may be available for functional interactions (Figure 4). The N- and C-terminal ends of BIGs and GBFs are more variable, including an unusual enrichment in Asp/Glu or Pro residues in some members. A specific feature of BIG members is that their C-terminus is in general less variable than that of GBFs, and is predicted with a significant amount of secondary structures. In contrast to the predicted structural domains, the intervening regions are highly variable in length and do not yield aligned sequences. Analysis of their amino-acid composition reveals a paucity of hydrophobic residues which is predicted to associate with an essentially unfolded conformation, suggesting that they act as linkers to tether the functional domains together.

To further investigate the predicted organization of BIGs and GBFs in 6 conserved helical domains connected by variable linkers, splicing patterns of human BIGs and GBFs were analyzed in the large number of cDNAs and ESTs in the databases that correspond to GBF/BIG transcripts. This revealed the use of alternate splice donor and acceptor sites predicted to yield proteins with insertions and deletions ranging from 1 to 38 residues, and a number of splice variants arising from exon skipping (Table 2). Strikingly, all observed sequence variations occur in regions identified as linkers between conserved



**Figure 3**  
**The conserved domains of the BIG/GBF subfamily: DCB domain.** Multiple sequence alignment of the conserved domains from BIG and GBF representative sequences showing secondary structure predictions that co-align in all sequences. Colour coding is red for invariant residues, yellow for a sequence similarity score threshold of 0.15 using the BLOSUM62 matrix. The gap in helix 4 is due to an insert in the drosophila Q9V696 sequence, and may be resulting from a sequence annotation error.

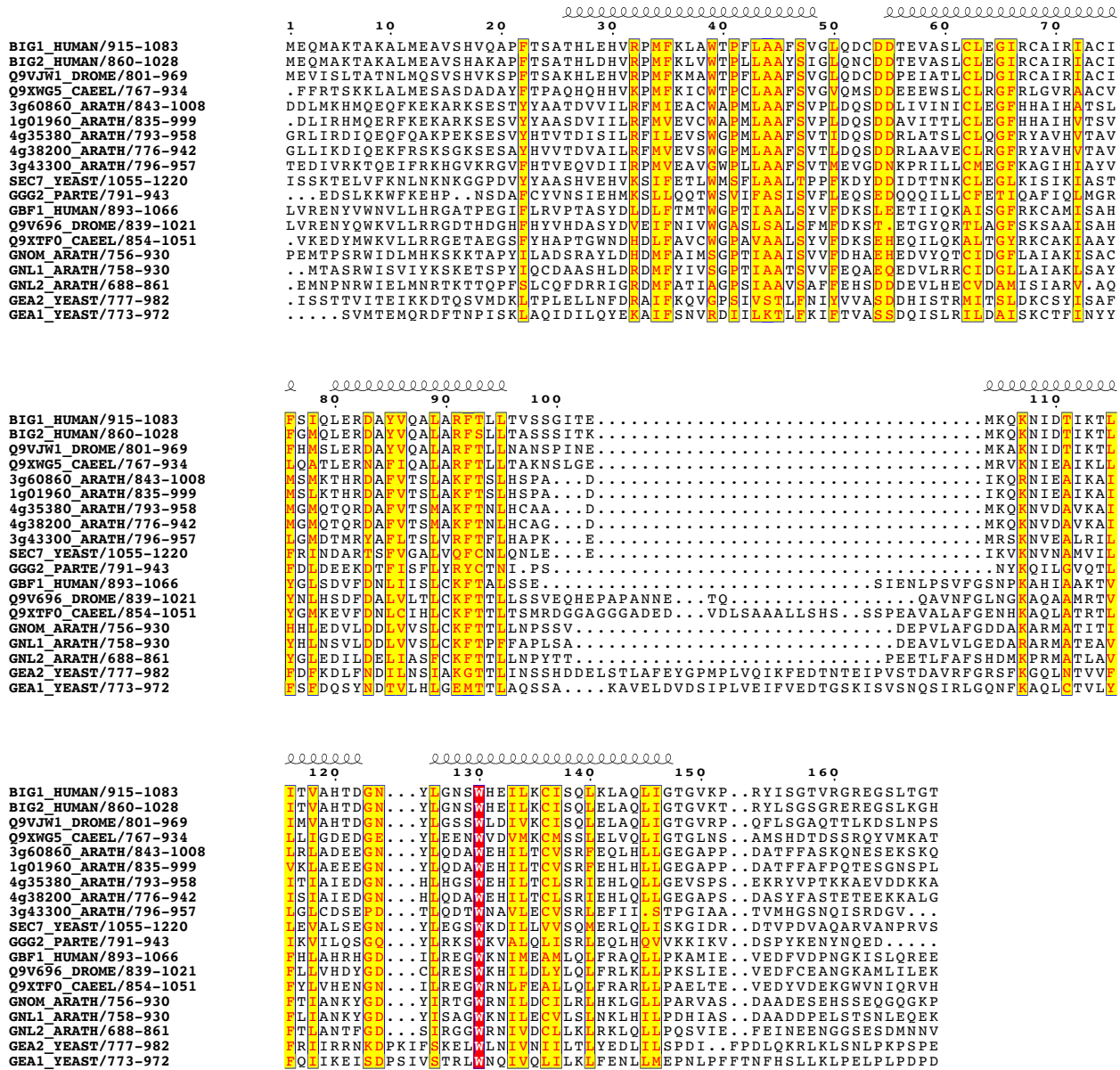




**Figure 4**  
**The conserved domains of the BIG/GBF subfamily: HUS domain.** See Figure 3 legend for alignment details. The highly conserved HUS motif is boxed in blue. The gap in helix 5 domain is due to an insert in the Arabidopsis 3g43300 sequence, and may be resulting from a sequence annotation error.

domains (Figure 2). Together with our domain analysis, this suggests that splicing at non-canonical exon/intron boundaries is only tolerated in regions of the protein

where the impact upon folding of domains with essential function would be minimal.

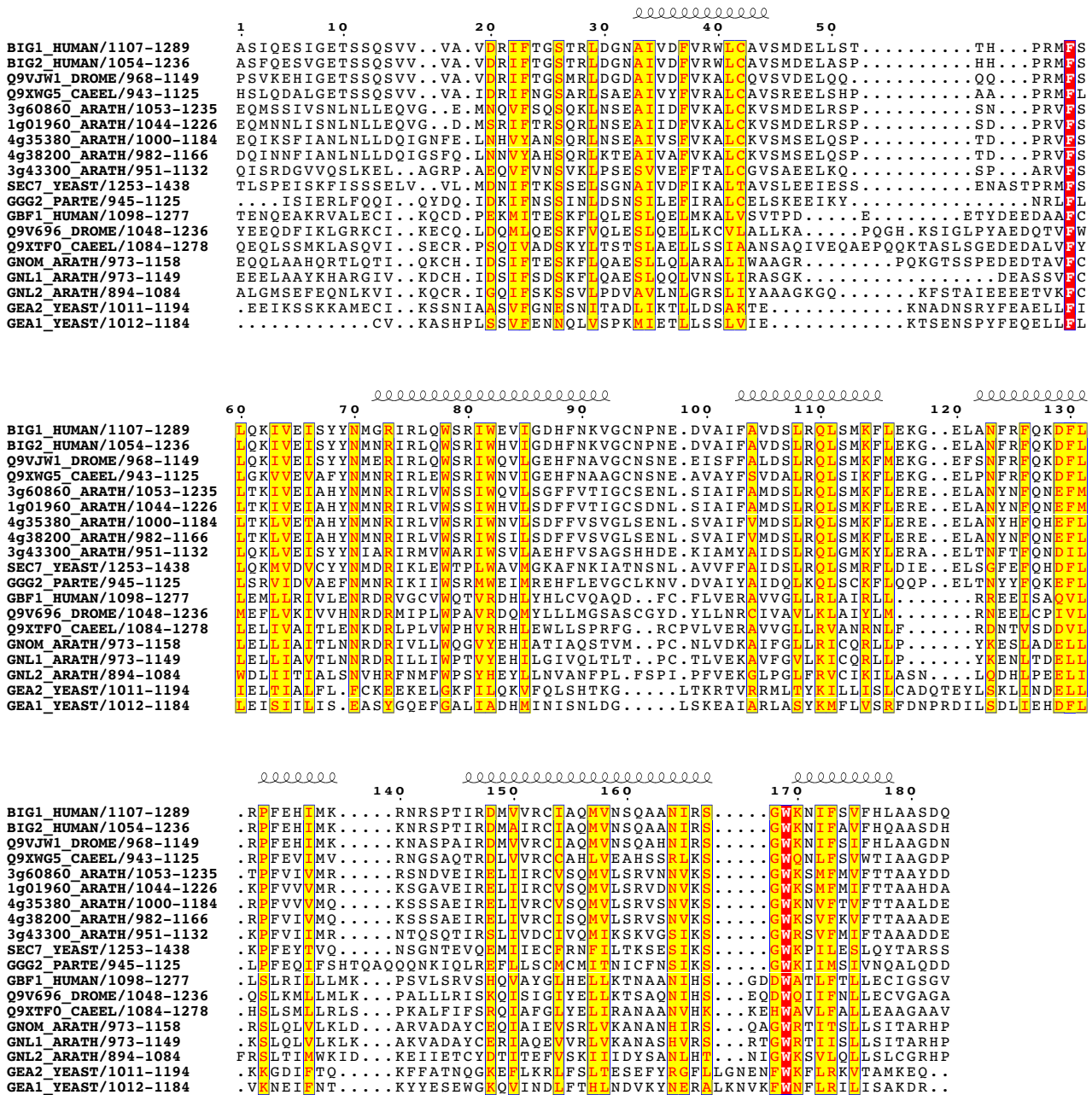


**Figure 5**  
**The conserved domains of the BIG/GBF subfamily: HDS1 domain.** See Figure 3 legend for alignment details.

**Evolution of BIGs and GBFs from a common ancestor**

Combined, our analysis reveals that the BIG and GBF subfamilies share the same overall domain organization, and are likely to descend from a common ancestor gene that duplicated first to form the BIG and GBF groups, and again within these groups to yield species-specific BIG and GBF members. These two subfamilies can therefore be

unified as a higher order ArfGEF subfamily (called below GBG for GBF/BIG GEFs), from which unrooted phylogenetic trees can be built (Figure 8). Unlike previous phylogenetic analysis which compared ArfGEFs based on their Sec7 domains after diverging non-catalytic regions have been trimmed [8], our trees were established from the simultaneous alignment of all 6 conserved domains

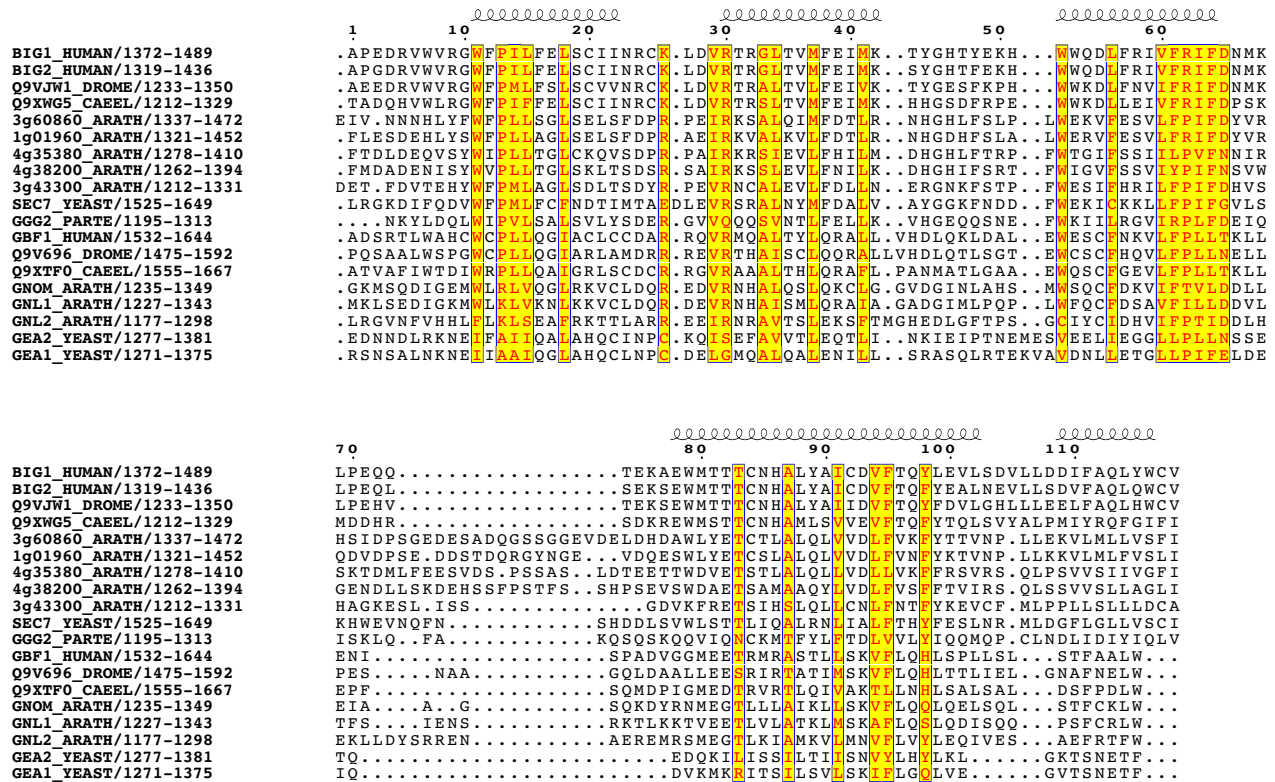


**Figure 6**  
The conserved domains of the BIG/GBF subfamily: HDS2 domain. See Figure 3 legend for alignment details.

(DCB, HUS, Sec7, HDS1, HDS2 and HDS3), excluding variable linkers. The same tree topology was obtained with both neighbor-joining and maximum likelihood methods, and was retained using any one of the new conserved domains alone (data not shown). Bootstrap analy-

sis strongly supported this topology for most branches. Only a few small branches located at the base of the groups were found in less than 60% of the trials in one of the two methods, but this never occurred with both methods simultaneously.





**Figure 7**  
**The conserved domains of the BIG/GBF subfamily: HDS3 domain.** See Figure 3 legend for alignment details.

The tree topology strongly suggests that in most organisms, GBG members sort in separate branches, corresponding to their classification in BIGs and GBFs. Remarkably, our annotation of Sec7-containing proteins in the genome of Paramecium reveals the first departure from this distribution, as all GBGs in this species are located in a single branch, which is closer to the BIGs. This unexpected tree topology may indicate that alveolates diverged from animals/fungi and plants before the duplication of an ancestral GBG into the BIG and GBF, and that GBGs in that organism are representative of this ancestral gene. Alternatively, duplication may have been followed by loss of GBF genes. Current knowledge of the phylogenetic branching of alveolates relative to the plants and animal/fungi branches does not permit resolution between those two possibilities.

**GBGs in plants: refining the functional evolution of BIGs and GBFs**

In fungi and mammals, BIGs and GBFs are represented by only one or two members, whose functions in vesicular

trafficking at the Golgi within each group appear largely overlapping [12,14]. In contrast, plants encode a large number of GBGs in both the BIG and GBF branches but lack other ArfGEFs (Figures 1, 8). In Arabidopsis, none of the GBGs map to duplicated chromosomes where identical functions may be encoded [26,29]. In addition, comparative analysis with the rice genome nearing completion identifies at least five branches each represented by one rice and one or two Arabidopsis homologs (Figure 8). This correspondence between two highly divergent plant species indicates that GBGs diversified early during plant evolution, probably reflecting functional specialization along with the establishment of plant multicellularity. While GNOM has a plant-specific function in recycling plasma-membrane proteins needed for cell-cell communication and cell polarity establishment [11], possibly closer to the function of EFA6 or CYH subfamilies in metazoans, other plant GBGs are expected to fulfill the presumed ancient function of regulating Golgi trafficking exemplified by mammalian and yeast GBGs. Comparison of orthologous pairs in plants further reveals that they

**Table 2: Alternate splice variants of human GBF1, BIG1 and BIG2<sup>a,b</sup>**

	Change in protein	Apparent cause of variation in transcript
GBF1	Extra Q at 337, 55 residues upstream of HUS domain	Insertion of 3 nucleotides (nt) resulting from use of alternate 3' acceptor site within intron during splicing of exons 10 and 11
	New Ser and loss of 14 residues at 613, between HUS and Sec7 domains	Loss of 36 nt resulting from use of alternate 5' donor site within exon 15 during splicing with exon 16
	Loss of VSQD at 1494, 38 residues upstream of HDS3	Loss of 12 nt resulting from use of alternate 5' donor site within exon 33 during splicing with exon 34
	Frame-shift at 1625 causing loss of last 19 residues of HDS3	Intron retention between exons 36 and 37 leading to frame shift and premature termination
	Loss of 38 residues starting at 1784, near C-terminus	Loss of 114 nt resulting from use of novel cryptic splice donor and acceptor sites within exon 40.
BIG1	Frame-shift at 1340, 32 residues upstream of HDS3	Loss of 59 nt resulting from use of alternate 5' donor site within exon 28 during splicing with exon 29
	Loss of VSEKPL at 1557, 68 residues downstream of HDS3	Loss of 18 nt resulting from use of alternate 5' donor site within exon 33 during splicing to exon 34
	New T and loss of 33 residues at 1607, 118 residues downstream of HDS3	Loss of 96 nt resulting from use of alternate 3' acceptor site within exon 35 during splicing with exon 34
BIG2	Frame-shift at 1542, 106 residues downstream of HDS3	Loss of exon 35 resulting from splicing of 5'donor site of exon 34 with 3' acceptor site of exon 36

<sup>a</sup> All changes were expressed relative to the reference sequence stored under accession number NM\_004193 (hGBF1), NM\_006421 (hBIG1) and NM\_006420.1 (hBIG2).

<sup>b</sup> All variants are supported by one or more cDNA/ESTs as detailed in the Aceview for each gene that can be obtained at [38].

have different sensitivities to Brefeldin A (a widely used fungal inhibitor of Golgi traffic) as predicted from the sequences of the binding site of the drug carried by the Sec7 domain [6]. This observation clearly illustrates that differences in outcome following BrefeldinA treatment may not reflect differences in underlying molecular mechanisms, but instead simply reflect neutral sequence differences at the Sec7 domains between species. In particular, not all BIGs may be BFA-sensitive or GBFs BFA-resistant, unlike suggested by their original nomenclature.

#### **A novel ArfGEF subfamily in alveolates**

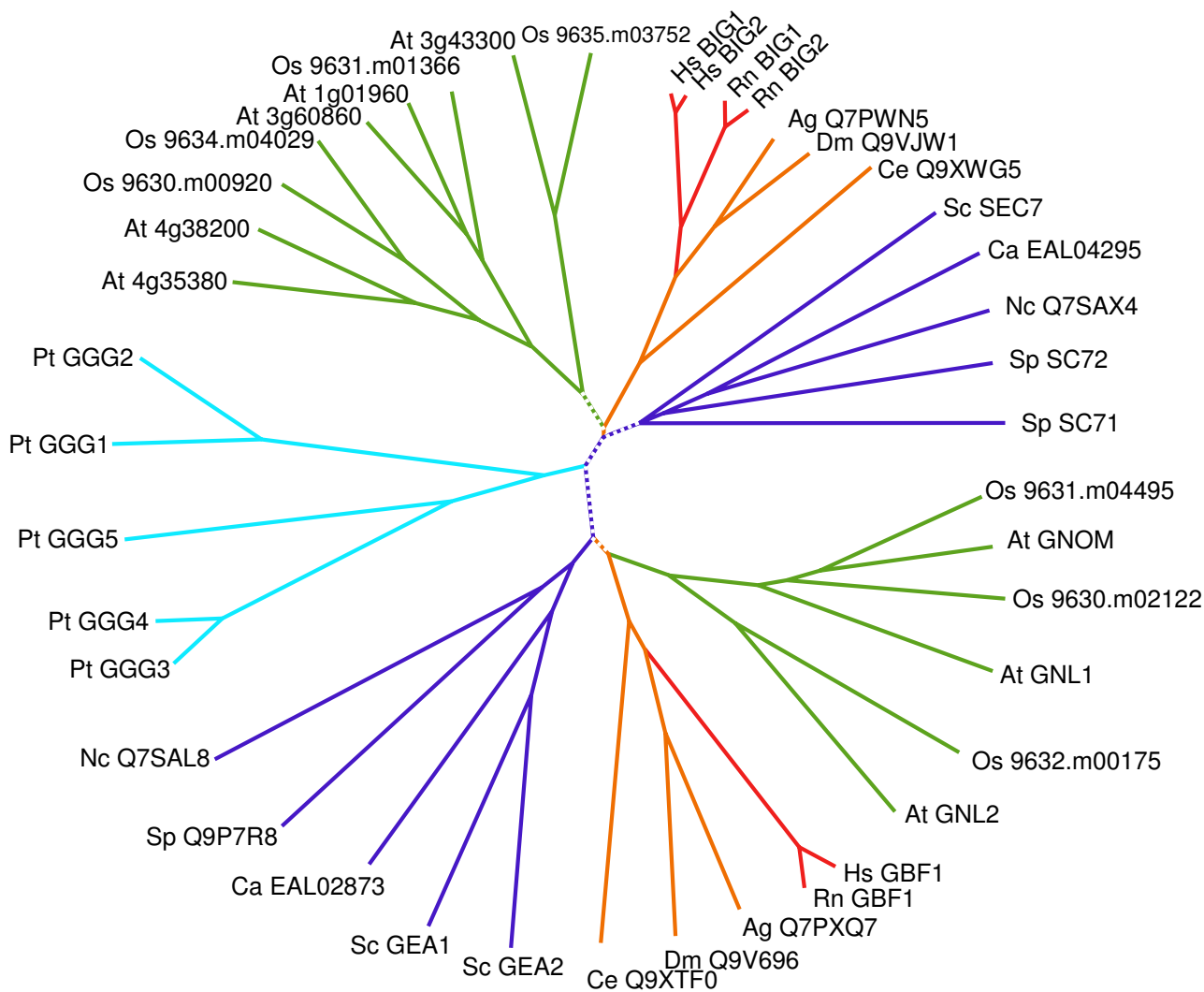
A remarkable evolutionary feature of ArfGEFs is that while GBGs seem to be ubiquitous to all eukaryotes, fungi and animals kingdoms evolved their own ArfGEFs subfamilies unrelated to those of the other kingdoms. We thus addressed the question of whether *Paramecium*, which has a large number of GBGs (at least five, of which four are present as pairs as the result of recent duplications) but appears to lack the specialization into the BIG and GBF subgroups, has the same ArfGEF distribution as plants or features a second ArfGEF subfamily. We thus searched the newly sequenced genome from *Paramecium tetraurelia* and the available alveolate genomes from *Cryptosporidium parvum* and *Tetrahymena thermophila* for additional Sec7-containing proteins. This identified a novel putative ArfGEF subfamily characterized by the association of the Sec7 domain with a TBC (Tre/Bub2/Cdc16) domain (Fig-

ure 9), which was found only in the protists kingdom. The TBC domain is predicted to carry a GAP (GTPase activating protein) activity towards small G proteins of the Rab family [30], suggesting a potential crosstalk between Rab and Arf pathways. Such a relationship between these two small G proteins families, which are major regulators of membrane traffic, would not be unprecedented, as for example the SYT1 ArfGEF gene was identified in yeast by its genetic interactions with Rab proteins in the exocytic pathway [31]. Interestingly, alveolates have specialized exocytic pathways based on a membrane organelle lying beneath the plasma membrane, the trichocyst, where this unique ArfGEF family may potentially function.

#### **Conclusion**

##### **A conserved scenario for the activation of Arf proteins by their GEFs?**

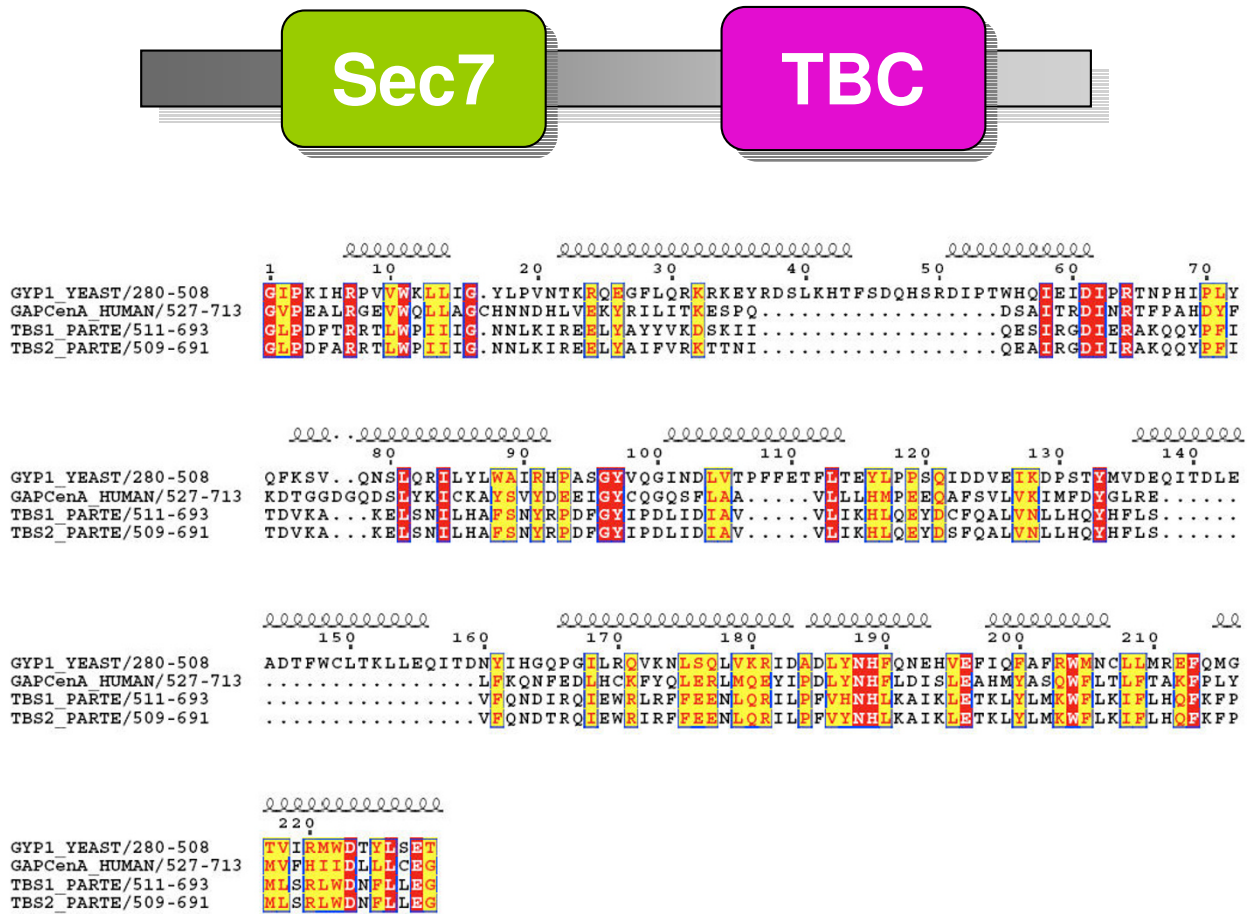
The identification of a conserved modular architecture in all GBG subfamily members suggests that the mechanistic basis for their activation of Arf is likely to follow a similar scenario. Candidate functions for the conserved domains include oligomerization, the collection of input signals, membrane localization, regulation of the exchange activity, scaffolding of Arf proteins to their downstream effectors, not excluding signaling to partners outside the Arf pathways. Dimerization has been reported in the BIG subgroup for BIG1, which forms heterodimers with the highly homologous BIG2 ArfGEF [14], and in the GBF



**Figure 8**  
**Unrooted neighbour-joining phylogenetic tree of the BIG/GBF subfamily.** Colour coding for the main groups is green for plants, marine blue for fungi, orange and red for animals, cyan for protists. Branches found in less than 60% bootstrap trials by either the neighbor-joining or the maximum likelihood method are in dotted lines. Species abbreviations are as in Table I.

subgroup for GNOM, which forms homodimers [27]. The conservation of the DCB domain in GBGs, which is responsible for the dimerization of GNOM, suggests that such a dimerization function may be general to this domain in GBGs. Another unresolved issue is the conservation of the cellular partners effecting the functions associated with the conserved domains. Our identification of

an almost invariant motif in the HUS domain argues in favor of this domain interacting with a conserved partner. However, the ancient divergence into the BIG and GBF groups and their subsequent divergence into species-specific members suggest that specialized requirements are likely to have evolved in most organisms, possibly yielding less conserved partners outside the Sec7 and HUS



**Figure 9**  
**TBS: a novel ArfGEF subfamily in alveolates.** Top: Domain structure of the TBS subfamily. Below: Sequences of the TBC domain from *Paramecium* TBS aligned with TBC domains from known RabGAPs. Secondary structures are from the crystal structure of yeast GYP1 [30].

domains. Finally, whereas in plants all ArfGEFs are predicted to function according to the scheme defined by the conserved domains, other species have additional ArfGEF subfamilies with a modular architecture unrelated to that of the GBG subfamily. It is not known to what extent the GBG's scenario for Arf activation will also apply to non-GBGs ArfGEFs, acting alone or in association with protein partners. In the case of the GBGs, our definition of the structural homology domains as reported here should now provide a robust background for future investigations of their interactions and functions.

**Methods**

Protein sequence databases were searched with amino acid sequences from human BIG1, human GBF1 and Arabidopsis GNOM using the BLAST algorithm [32]. *Paramecium tetraurelia* genes were identified with the BLAST algorithm using genome sequence data from Genoscope [33] and manually annotated using Artemis [34]. Tetrahymena sequences were retrieved from the *Tetrahymena thermophila* genome sequencing project server [35]. Arabidopsis sequences were retrieved from the Arabidopsis Genome Initiative database [36], rice sequences from the TIGR Rice annotation project [37].



Splice variants for hGBF1, hBIG1 and hBIG2 were identified from information provided under Aceview in the December (03) release for their respective listings at the NCBI [38]. Multiple sequence alignments were performed using ClustalW [39] with default alignment parameters or T-coffee [40,41]. Reliability of the alignments was evaluated according to the T-coffee score, and ranged from average to good for all predicted domains. Average sequence identities were respectively 24 % (DCB domain), 26 % (HUS domain), 44% (Sec7 domain), 26% (HDS1 domain), 28% (HDS2 domain) and 21% (HDS3 domain). Aligned sequences were displayed with ESPrnt [42] using a similarity global score of 0.15 calculated using the BLOSUM62 matrix. Unrooted phylogenetic trees were generated using the neighbor-joining algorithm of ClustalW excluding gapped regions, and with a maximum likelihood method using the PHYML package [43]. Phylogenetic trees for individual domains was performed on the subset of sequences used in Figure 3. The reliability of the trees was assessed by a bootstrap analysis (1000 replicates). Trees were drawn with TreeView version 1.6.6. Secondary structure predictions on aligned sequences were carried out with the PHD program along with the ClustalW multiple alignment [39]. Non-structured linkers poor in hydrophobic residues were predicted with the PONDR algorithm [44].

### Abbreviations

GEF: Guanine nucleotide exchange factor. CYH: cytohesins/ARNO; EFA: Exchange Factor for Arf6; FBS: F-Box/Sec7; TBS: TBC/Sec7; GBF: Golgi-associated BFA-resistant guanine nucleotide exchange factor; BIG: BFA-Inhibited Guanine nucleotide exchange factor; GBC: GBF/BIG Gefs; SYT1: Suppressor of *ypf*. DCB: Dimerization/Cyclophilin Binding; HUS: Homology Upstream of Sec7; HDS: Homology Downstream of Sec7; TBC: Tre/Bub2/Cdc16; SCF: Skp1/Cull1/F box.

### Authors' contributions

B.M. and V.B. carried out sequence and phylogenetic analysis. A.J. participated in the domain analysis. J.Co. annotated Paramecium sequences. D.S. and P.M. performed splicing pattern analysis. N.K. and G.J. analyzed the distribution of large ArfGEFs in plants. J.Ch. conceived and coordinated the study and wrote the manuscript.

### Acknowledgements

This work was supported by a Human Frontiers in Science Program grant to G.J., P.M. and J.Ch. We thank Genoscope for access to Paramecium whole genome shotgun primary data and Linda Sperling (CNRS, Gif-sur-Yvette) for help with annotations.

### References

1. Cherfils J, Chardin P: **GEFs: structural basis for their activation of small GTP-binding proteins.** *Trends Biochem Sci* 1999, **24**:306-311.

2. Chavrier P, Goud B: **The role of ARF and Rab GTPases in membrane transport.** *Curr Opin Cell Biol* 1999, **11**:466-475.
3. Chardin P, Paris S, Antony B, Robineau S, Beraud-Dufour S, Jackson CL, Chabre M: **A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains.** *Nature* 1996, **384**:481-484.
4. Pasqualato S, Renault L, Cherfils J: **The GDP/GTP cycle of Arf proteins. Structural and biochemical aspects.** *The ARF Book Richard A Kahn, editor Kluwer Academic Publishers* 2004:23-48.
5. Goldberg J: **Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching.** *Cell* 1998, **95**:237-248.
6. Renault L, Guibert B, Cherfils J: **Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor.** *Nature* 2003, **426**:525-530.
7. Jackson CL, Casanova JE: **Turning on ARF: the Sec7 family of guanine-nucleotide-exchange factors.** *Trends Cell Biol* 2000, **10**:60-67.
8. Cox R, Mason-Gamer RJ, Jackson CL, Segev N: **Phylogenetic analysis of Sec7-domain-containing Arf nucleotide exchangers.** *Mol Biol Cell* 2004, **15**:1487-1505.
9. Peyroche A, Paris S, Jackson CL: **Nucleotide exchange on ARF mediated by yeast Gea1 protein.** *Nature* 1996, **384**:479-481.
10. Claude A, Zhao BP, Kuziemytsky CE, Dahan S, Berger SJ, Yan JP, Arnold AD, Sullivan EM, Melançon P: **GBF1: A novel Golgi-associated BFA-resistant guanine nucleotide exchange factor that displays specificity for ADP-ribosylation factor 5.** *J Cell Biol* 1999, **146**:71-84.
11. Geldner N, Anders N, Wolters H, Keicher J, Kornberger W, Muller P, Delbarre A, Ueda T, Nakano A, Jürgens G: **The Arabidopsis GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth.** *Cell* 2003, **112**:219-230.
12. Peyroche A, Courbeyrette R, Rambourg A, Jackson CL: **The ARF exchange factors Gea1p and Gea2p regulate Golgi structure and function in yeast.** *J Cell Sci* 2001, **114**:2241-2253.
13. Mansour SJ, Skaug J, Zhao XH, Giordano J, Scherer SW, Melançon P: **p200 ARF-GEF1: a Golgi-localized guanine nucleotide exchange protein whose Sec7 domain is targeted by the drug brefeldin A.** *Proc Natl Acad Sci U S A* 1999, **96**:7968-7973.
14. Yamaji R, Adamik R, Takeda K, Togawa A, Pacheco-Rodriguez G, Ferrans VJ, Moss J, Vaughan M: **Identification and localization of two brefeldin A-inhibited guanine nucleotide-exchange proteins for ADP-ribosylation factors in a macromolecular complex.** *Proc Natl Acad Sci U S A* 2000, **97**:2567-2572.
15. Achstetter T, Franzusoff A, Field C, Schekman R: **SEC7 encodes an unusual high molecular weight protein required for membrane traffic from the yeast Golgi apparatus.** *J Biol Chem* 1988, **263**:11711-11717.
16. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR: **A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes.** *Science* 2002, **295**:679-682.
17. Donaldson JG: **Multiple roles for Arf6: sorting, structuring, and signaling at the plasma membrane.** *J Biol Chem* 2003, **278**:41573-41576.
18. Zhao X, Lasell TK, Melançon P: **Localization of large ADP-ribosylation factor-guanine nucleotide exchange factors to different Golgi compartments: evidence for distinct functions in protein traffic.** *Mol Biol Cell* 2002, **13**:119-133.
19. Shin HW, Morinaga N, Noda M, Nakayama K: **BIG2, a guanine nucleotide exchange factor for ADP-ribosylation factors: its localization to recycling endosomes and implication in the endosome integrity.** *Mol Biol Cell* 2004, **15**:5283-5294.
20. Cullen PJ, Chardin P: **Membrane targeting: what a difference a G makes.** *Curr Biol* 2000, **10**:R876-8.
21. Derrien V, Couillault C, Franco M, Martineau S, Montcourrier P, Houlgatte R, Chavrier P: **A conserved C-terminal domain of EFA6-family ARF6-guanine nucleotide exchange factors induces lengthening of microvilli-like membrane protrusions.** *J Cell Sci* 2002, **115**:2867-2879.
22. Chen EH, Pryce BA, Tzeng JA, Gonzalez GA, Olson EN: **Control of myoblast fusion by a guanine nucleotide exchange factor, loner, and its effector ARF6.** *Cell* 2003, **114**:751-762.
23. Ilyin GP, Rialland M, Pigeon C, Guguen-Guillouzo C: **cDNA cloning and expression analysis of new members of the mammalian F-box protein family.** *Genomics* 2000, **67**:40-47.

24. Mansour M, Lee SY, Pohajdak B: **The N-terminal coiled coil domain of the cytohesin/ARNO family of guanine nucleotide exchange factors interacts with the scaffolding protein CASP.** *J Biol Chem* 2002, **277**:32302-32309.
25. Lee SY, Pohajdak B: **N-terminal targeting of guanine nucleotide exchange factors (GEF) for ADP ribosylation factors (ARF) to the Golgi.** *J Cell Sci* 2000, **113** ( Pt 11):1883-1889.
26. Jürgens G, Geldner N: **Protein secretion in plants: from the trans-Golgi network to the outer space.** *Traffic* 2002, **3**:605-613.
27. Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Salchert K, Koncz C, Jürgens G: **A conserved domain of the arabidopsis GNOM protein mediates subunit interaction and cyclophilin 5 binding.** *Plant Cell* 2000, **12**:343-356.
28. Jochum A, Jackson D, Schwarz H, Pipkorn R, Singer-Kruger B: **Yeast Ysl2p, homologous to Sec7 domain guanine nucleotide exchange factors, functions in endocytosis and maintenance of vacuole integrity and interacts with the Arf-Like small GTPase Arl1p.** *Mol Cell Biol* 2002, **22**:4914-4928.
29. Initiative AG: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2000, **408**:796-815.
30. Rak A, Fedorov R, Alexandrov K, Albert S, Goody RS, Gallwitz D, Scheidig AJ: **Crystal structure of the GAP domain of Gyp1p: first insights into interaction with Ypt/Rab proteins.** *Embo J* 2000, **19**:5105-5113.
31. Jones S, Jedd G, Kahn RA, Franzusoff A, Bartolini F, Segev N: **Genetic interactions in yeast between Ypt GTPases and Arf guanine nucleotide exchangers.** *Genetics* 1999, **152**:1543-1556.
32. **Bork Group's WU-BLAST2 Search Service at EMBL** [<http://dove.embl-heidelberg.de/Blast2/>]
33. **Paramecium Genomics** [<http://paramecium.cgm.cnrs-gif.fr/>]
34. **The Sanger Institute: Informatics Software: Artemis** [<http://www.sanger.ac.uk/Software/Artemis/>]
35. **The TIGR Tetrahymena thermophila Genome Project** [<http://www.tigr.org/tdb/e2k1/ttg/>]
36. **TAIR BLAST** [<http://www.arabidopsis.org/Blast/>]
37. **TIGR Rice Genome Annotation** [<http://www.tigr.org/tdb/e2k1/osa1/>]
38. **AceView** [<http://www.ncbi.nlm.nih.gov/EB/Research/AceView/>]
39. **Pole Bioinformatique Lyonnais** [<http://pbil.univ-lyon1.fr/>]
40. 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.
41. **Tcoffee** [<http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee.cgi/index.cgi>]
42. **Easy execution of ESPrpt 2.x / ENDscript 1.x** [<http://ribosome.toulouse.inra.fr/ESPrpt/cgi-bin/ESPrpt.cgi>]
43. **PHYML: fast, accurate estimation of large PHYlogenies by Maximum Likelihood** [<http://atgc.lirmm.fr/phyml/>]
44. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z: **Intrinsic disorder and protein function.** *Biochemistry* 2002, **41**:6573-6582.
45. Padilla PI, Chang MJ, Pacheco-Rodriguez G, Adamik R, Moss J, Vaughan M: **Interaction of FK506-binding protein 13 with brefeldin A-inhibited guanine nucleotide-exchange protein 1 (BIG1): effects of FK506.** *Proc Natl Acad Sci U S A* 2003, **100**:2322-2327.
46. Li H, Adamik R, Pacheco-Rodriguez G, Moss J, Vaughan M: **Protein kinase A-anchoring (AKAP) domains in brefeldin A-inhibited guanine nucleotide-exchange protein 2 (BIG2).** *Proc Natl Acad Sci U S A* 2003, **100**:1627-1632.
47. Chantalat S, Park SK, Hua Z, Liu K, Gobin R, Peyroche A, Rambourg A, Graham TR, Jackson CL: **The Arf activator Gea2p and the P-type ATPase Drs2p interact at the Golgi in Saccharomyces cerevisiae.** *J Cell Sci* 2004, **117**:711-722.
48. Chantalat S, Courbeyrette R, Senic-Matuglia F, Jackson CL, Goud B, Peyroche A: **A novel Golgi membrane protein is a partner of the ARF exchange factors Gea1p and Gea2p.** *Mol Biol Cell* 2003, **14**:2357-2371.
49. Charych EI, Yu W, Miralles CP, Serwanski DR, Li X, Rubio M, De Blas AL: **The brefeldin A-inhibited GDP/GTP exchange factor 2, a protein involved in vesicular trafficking, interacts with the beta subunits of the GABA receptors.** *J Neurochem* 2004, **90**:173-189.
50. Garcia-Mata R, Sztul E: **The membrane-tethering protein p115 interacts with GBF1, an ARF guanine-nucleotide-exchange factor.** *EMBO Rep* 2003, **4**:320-325.
51. Saeki N, Tokuo H, Ikebe M: **BIG1 is a binding partner of myosin IXB and regulates its Rho gap activity.** *J Biol Chem* 2005.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

