2196–2205 *Nucleic Acids Research, 2006, Vol. 34, No. 8*
doi:10.1093/nar/gkl178

Structural stabilization of GTP-binding domains in circularly permuted GTPases: Implications for RNA binding

Baskaran Anand, Sunil Kumar Verma and Balaji Prakash*

Department of Biological Sciences and BioEngineering, Indian Institute of Technology, Kanpur 208 016, India

Received December 21, 2005; Revised and Accepted March 21, 2006

ABSTRACT

GTP hydrolysis by GTPases requires crucial residues embedded in a conserved G-domain as sequence motifs G1–G5. However, in some of the recently identified GTPases, the motif order is circularly permuted. All possible circular permutations were identified after artificially permuting the classical GTPases and subjecting them to profile Hidden Markov Model searches. This revealed G4–G5–G1–G2–G3 as the only possible circular permutation that can exist in nature. It was also possible to recognize a structural rationale for the absence of other permutations, which either destabilize the invariant GTPase fold or disrupt regions that provide critical residues for GTP binding and hydrolysis, such as Switch-I and Switch-II. The circular permutation relocates Switch-II to the C-terminus and leaves it unfastened, thus affecting GTP binding and hydrolysis. Stabilizing this region would require the presence of an additional domain following Switch-II. Circularly permuted GTPases (cpGTPases) conform to such a requirement and always possess an 'anchoring' C-terminal domain. There are four sub-families of cpGTPases, of which three possess an additional domain N-terminal to the G-domain. The biochemical function of these domains, based on available experimental reports and domain recognition analysis carried out here, are suggestive of RNA binding. The features that dictate RNA binding are unique to each subfamily. It is possible that RNA-binding modulates GTP binding or *vice versa*. In addition, phylogenetic analysis indicates a closer evolutionary relationship between cpGTPases and a set of universally conserved bacterial GTPases that bind the ribosome. It appears that cpGTPases are RNA-binding proteins possessing a means to relate GTP binding to RNA binding.

INTRODUCTION

Guanine nucleotide-binding proteins or G proteins are well-known molecular switches controlling several key cellular events (1). Extensive structural and biochemical studies have contributed immensely to our current understanding of their roles in protein synthesis, signaling events leading to cell proliferation and differentiation, protein trafficking, endocytosis, cytoskeletal rearrangement and cell motility (1,2). However, the ongoing genome sequencing projects have led to the identification of potential open reading frames encoding putative GTPases of unknown function. While experimental methods were employed to understand the cellular and biochemical role of these proteins, a broad classification of these putative GTPases encoded by the genomes of various organisms began with a systematic attempt to reconstruct their evolutionary history (3,4). Based on sequence and structural features, Leipe *et al*. (3) divided the vast assemblage of proteins in the GTPase superclass into two major classes. The first class is termed TRAFAC (named after translation factors) and it encompasses enzymes implicated in translation (initiation, elongation and release factors), signal transduction (exemplified by the extended Ras-like family), cell motility and intracellular transport. The second class is referred to as SIMIBI (after signal recognition particle, MinD and BioD) that includes enzymes involved in signal recognition and metabolism.

A structurally invariant core domain that has an ability to bind GTP/GDP bestows a switch mechanism to the G-domains when GTP is hydrolysed to GDP. For this, the G-domains require five sequence motifs G1–G5 that mediate interactions with the nucleotides and effector proteins. Of these, G1 [Gx(4)GKS/T], G3 (DxxG) and G4 (N/TKxD) can easily be recognized in the primary sequences due to their high conservation. In addition to conventional GTPases that contain the motif order G1–G3–G4, the TRAFAC class also includes an atypical family, which seems to present an interesting variation, wherein the order of these sequence motifs is circularly permuted. These atypical circularly permuted GTPases (cpGTPases) were grouped into distinct subfamilies (3)

*To whom correspondence should be addressed. Tel: +91 512 2594013; Fax: +91 512 2594010; Email: bprakash@iitk.ac.in

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

represented by the proteins YlqF (of *Bacillus subtilis*), YqeH (of *B.subtilis*), YjeQ (of *Escherichia coli*) and YawG (of *Schizosaccharomyces pombe*). In the amino acid sequences of these proteins, the occurrence of sequence motifs follows the order G4–G1–G3, as against the usual order, G1–G3–G4, observed in canonical GTPases. Despite such a variation (in motif order) observed at the primary sequence level, which should lead to different topological connections between the secondary structure elements, the three dimensional fold seems to be well preserved, as noted from the structures of YlqF (New york Structural Genomics Research Consortium) and YjeQ (5,6). In addition, the GTP-binding site too, seems to be well preserved, reinforcing the view that the folding and final structure adopted by the G-domain is a defining feature for the switch mechanism.

cpGTPases seem to satisfy important functions for the cell. YlqF and YqeH are reported to be essential for the viability of *B.subtilis* (7). Similarly, YjeQ and its ortholog YloQ have been shown to be important for the growth of *E.coli* and *B.subtilis*, respectively (8). However, unlike for most classical GTPases, little is known about cpGTPases and their functions still need to be elucidated. This work is a first attempt aimed at understanding the structure–function relationships among cpGTPases. Here, we have systematically searched the protein sequences to uncover all possible circular permutations of G1–G5 motifs. Using sequence and structural analysis, coupled with available biochemical data, we address (i) the functional advantages of a circular permutation to GTPases, (ii) the probable evolutionary link to conventional GTPases, and (iii) an RNA-binding function associated to these proteins. Our investigations suggest that cpGTPases share a close evolutionary relationship to a set of bacterial ribosome-binding GTPases and posses a novel means to couple guanine nucleotide binding with an RNA-binding function.

MATERIALS AND METHODS

Identification and classification of cpGTPases

Classical GTPases, wherein the motif order is G1–G2–G3–G4–G5, from 14 different families were identified for this study. In order to contain the volume of sequences within a given family, a redundancy criterion was applied to discard sequences that shared a high sequence identity. For example, a redundancy cutoff of 90% would imply that for sequences that share >90% sequence identity, only one would be retained for the study. Furthermore, to ensure that all families had similar representation in the dataset, the redundancy cut off varied from 60–95%. After this initial screening, a total of 410 GTPases (see a list in Supplementary Data) were aligned using T-Coffee (9) and care was taken to ensure that the G1–G5 motifs align optimally. This alignment (Supplementary Data) was artificially permuted to generate four different alignments that reflected all possible circular permutations of the G1–G5 motifs, i.e. G2–G3–G4–G5–G1, G3–G4–G5–G1–G2, G4–G5–G1–G2–G3 and G5–G1–G2–G3–G4, respectively. These sequence alignments form the initial dataset of permuted G-domains. Profile Hidden Markov Models (HMMs) which represent the multiple sequence alignments (MSAs) based on statistical descriptors were generated for the permuted G-domains using 'hmmbuild' program of HMMER2 package (10). Separate profiles were generated for each alignment representing a given permutation. Profile HMM based searches were initiated for all permutations, separately, using 'hmmsearch' program against Swissprot/TrEMBL (Release 45; October 2004) sequence databases. All the hits were manually inspected for the presence of characteristic sequences motifs G1, G3 and G4, since G2 and G5 are not well defined at the sequence level. Further, the hits that show G1–G3–G4 permutation were removed as they belong to canonical GTPases. The remaining sequences were manually examined for circular permutations and were grouped according to the type of circular permutation. These sequences were further examined based on annotations to include them as GTPases. This analysis revealed that the motif order in GTPases is either G1–G3–G4 or its permutation, G4–G3–G1. Hence, the sequences displaying G4–G3–G1 permutation form the final data set of circularly permuted GTPases for further analysis. Following the classification scheme of Leipe *et al*. (3) *E.coli* YjeQ, *B.subtilis* YqeH and YlqF and *S.pombe* YawG were queried against the data set of circularly permuted GTPases, obtained from our analysis, using PSI-BLAST (11). The hits were manually inspected for highly similar sequences and were grouped into subfamilies. MSA for related sequences of each subfamily was generated using T-Coffee (9) and manually adjusted using Jalview (12) and Seaview (13). Thus these manually curated sequence alignments form the seeds for each of the subfamilies. Profile HMMs were generated for each subfamily seed alignments and were used to search against the dataset of circularly permuted GTPases to ensure that no genuine member is missed out of the analysis.

Domain assignment for cpGTPases

From the MSA of each subfamily, it was noticed that there were extensions at both the N- and C-termini of the circularly permuted G-domain. In order to inquire if these regions could form domains, the following searches were conducted. For each subfamily, MSA of both N- and C-terminal regions was made using T-Coffee (9) and used to initiate a PSI-BLAST (11) search on NR database (National Center for Biotechnology Information, NIH, Bethesda, MD; Dec 2004) along with the representative query sequence with an *E*-value threshold for profile inclusion set at 0.05 (unless specified otherwise). The searches were iterated until convergence or till 15 rounds, whichever occurred earlier. In addition, secondary structure predictions were made using JPRED online server (14) using the MSA as input.

Examination of gene order in bacterial genomes

For all orthologous bacterial proteins in a subfamily, the gene order has been examined using STRING database (15) and VIMSS operon prediction server (16).

Phylogenetic analysis

Representative sequences were chosen for each GTPase families in such a way that they cover all domains of life (Archaea, Bacteria and Eukaryotes) whenever available. For inferring homology, only that part of the full-length protein which forms the GTPase domain was considered. However, in the case of EngA family where there are two tandem GTPase

domains, each domain was considered as a separate sequence. EngA is now represented by its two domains EngA_1 and EngA_2. This circumvents the technical limitation in constructing a MSA. Further, we have permuted the motif order (G4–G5–G1–G2–G3) in cpGTPases back to canonical form (G1–G2–G3–G4–G5); this simplifies the alignment of motifs between cpGTPases and classical GTPases without losing information in the primary sequence, which is essential to decipher the phylogeny.

T-Coffee (9) was used to construct MSA with default parameters. Gaps and the N- and C-terminal extensions with large variations in length (Supplementary Data) were removed from the MSA. Further manual inspection ensured high reliability of the alignment. Trees were constructed using 'protdist' and 'fitch' programs in the PHYLIP 3.62 package (17) with 1000 bootstrap replicates. In addition, trees were also constructed using Treepuzzle 5.2 (18) with BLOSUM62 substitution matrix and 50 000 quartet puzzling steps were employed to get support for the tree branches.

RESULTS

GTPases exhibit only one possible circular permutation

Guided by the secondary structure of Ras, GTPases that exhibit canonical motif order (G1–G2–G3–G4–G5) were artificially (circularly) permuted. This led to sequence alignments representing four other permutations with motif orders G2–G3–G4–G5–G1, G3–G4–G5–G1–G2, G4–G5–G1–G2–G3 and G5–G1–G2–G3–G4, respectively. Profile HMMs were generated for each of these artificially permuted alignments. These profiles were used to search against Swissprot/TrEMBL sequence databases in order to identify GTPases with all possible circular permutations (Materials and Methods). As G2 and G5 regions do not exhibit strong conservation, visual inspection of the hits to identify the highly conserved sequence motifs, G1, G3 and G4 ensured bonafide GTPases. Interestingly, the only circular permutation that these proteins displayed was G4–G1–G3. Classical GTPases such as Ras and heterotrimeric G proteins have now been well studied (2). The recently reported structural and biochemical studies on circularly permuted GTPases such as YjeQ and YloQ (5,6,19–21) and the structure of YlqF (New york Structural Genomics Research Consortium) prompted us to investigate the structural basis for the occurrence of G4–G1–G3 permutation in comparison with the structure of Ras.

The canonical GTPase domain exemplified by Ras consists of six beta strands (β1–β6) forming a central β-sheet, which is flanked on either side by five α-helices (α1–α5). β2 is antiparallel to rest of the beta strands (Figure 1a) and the topological connections exhibited by the secondary structures in the invariant G-domains are ubiquitous (neglecting the insertions and deletions observed in the loops). In principle, circular permutation of the sequence motifs G1–G5 can generate five different permutations (represented P1–P5 in Figure 1a) with different topological connections, of which one, i.e. G1–G2–G3–G4–G5, is observed in the canonical GTPases (P1 in Figure 1a and b). Starting with this topology, other permutations can be generated by joining the existing N- and C-termini (dashed line in Figure 1a) and creating breaks elsewhere between the secondary structures to produce new N- and C-termini (arrows in Figure 1a). For obtaining a permutation such as G2–G3–G4–G5–G1, new N- and C-termini have to be created in the P-loop [Gx(4)GKS/T], which provides the

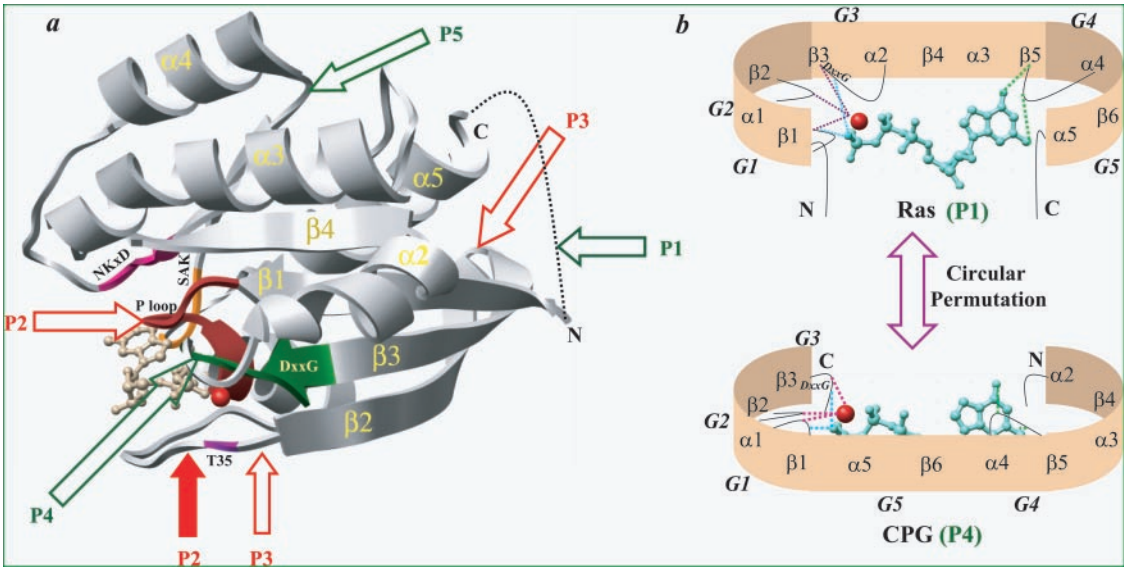

Figure 1. The figure shows all possible circular permutations for GTPases, using the structure of the prototype Ras and compares overall architecture in Ras and cpGTPase. (**a**) The structure of Ras is shown as gray ribbons (PDB code, 5P21). GTP is shown as ball and stick model while Mg^{2+} is shown as sphere (red). The regions harbouring sequence motifs G1–G5 are depicted in different colours along the structure. G1 (^{10}Gx4GKS/T^{17}) is shown in brown, G2 (T^{35}) in blue, G3 (^{57}DxxG^{60}) in green, G4 (^{116}NKxD^{119}) in violet and G5 (^{145}SAK/L^{147}) is colored orange. The arrows indicate ways to create new N- and C-termini. Green arrows represent an allowed permutation that satisfies the features required for binding and hydrolysis. They are P1 (G1–G2–G3–G4–G5), P4 (G4–G5–G1–G2–G3) and P5 (G5–G1–G2–G3–G4). Permutations P2 (G2–G3–G4–G5–G1) and P3 (G3–G4–G5–G1–G2) represented by red arrows do not seem to be viable (see text). Open and filled arrows of P2 indicate two different ways of creating the same permutation P2. (**b**) A cartoon representation of the permutations P1 and P4 observed in classical Ras like GTPases and cpGTPases to indicate that the structural features essential for GTP binding and hydrolysis are preserved in both. GTP and Mg^{2+} are shown as in (a). The hydrogen bond to guanine base is depicted in green, phosphate in cyan and Mg^{2+} in magenta, respectively. The locations of the sequence motifs G1–G4 are indicated.

critical residues for phosphate binding along its entire length (open arrow corresponding to P2 in Figure 1a). Since any break in this loop will impair phosphate binding, this permutation would be deleterious. In addition, the large conformational change associated with Switch-I upon GTP/GDP binding will be enormous for α1 to withstand alone, as it will be the only secondary structure on the N-terminal side of Switch-I. The same permutation can also result when breakage occurs in the G2/Switch-I region (filled arrow corresponding to P2 in Figure 1a), ahead of the crucial Thr (T35 in Ras) residue whose role is to stabilize the Mg^{2+} ion. However, such a breakage would impair the conformational changes required for GTP binding and the Mg^{2+} co-ordination. Further, since Switch-I region is known to interact with effector molecules, its integrity needs to be preserved for efficient switching mechanism. This explains why our hits do not find a G2–G3–G4–G5–G1 permutation.

As the Thr of G2/Switch-I lies just prior to β2, G3–G4–G5–G1–G2 permutation can only result by breaking the reverse turn between β2 and β3 (P3 in Figure 1a). This breakage and hence such a permutation are undesirable as disruption of reverse turns is known to play a decisive role in directing the proper folding of the protein (22). Another possible permutation that can in principle occur without disrupting any important regions of the structure is G5–G1–G2–G3–G4 (P5 in Figure 1a). However, G5, unlike the other motifs, is not very well conserved among all families and it is hence difficult to identify this region at the sequence level. The absence of such a permutation in our searches could also be because this permutation is very similar to the canonical GTPases with motif order G1–G2–G3–G4–G5. This leaves us with only one possible circular permutation i.e. G4–G5–G1–G2–G3 (P4 in Figure 1), concurrent with the results obtained when the databases are searched to identify circularly permuted GTPases. From here on, a circularly permuted G-domain (CPG-domain) implies one carrying the permutation G4–G5–G1–G2–G3 and the corresponding full-length protein comprising such a domain will be referred as 'circularly permuted GTPase' or cpGTPase. Following the classification scheme of Leipe *et al*. (3) and based on sequence similarity, we have categorized cpGTPases identified from database searches into four subfamilies represented by YjeQ (148 members), YlqF (55 members), YqeH (31 members) and YawG (32 members) (Figure 2). Using sequence and structural analysis, we attempted to inquire the structural and functional merits of a circular permutation in G-domains. Several important features of cpGTPases emerge from our analysis and these are presented in the following sections.

A stabilizing domain at the C-terminus is desirable for cpGTPases

Building on an analogy to Ras, a circular permutation such as the one described above should retain the relative positions of G1–G5 in the 3D structure, such that the overall fold is essentially invariant to preserve GTP binding. However, the newly formed N- and C-termini (at the green arrow of P4 in Figure 1a), resulting from a breakage in the region following DxxG motif of Switch-II and prior to helix α2, puts forward an interesting and unavoidable constraint. G3/Switch-II is the region contributing essential binding (DxxG) and catalytic (corresponding to Q^{61} of Ras) residues in canonical GTPases: in the CPG-domain, it lies at the C-terminus (Figure 1b). In canonical GTPases, Switch-II, along with Switch-I, undergoes dramatic conformational changes between GTP and GDP bound forms (23). In CPG-domains, the repositioned Switch-II at the C-terminus is anchored with β3 on one end, leaving the other end unfastened (Figure 3). Therefore, it becomes necessary to stabilize the unfastened end for Switch-II to perform an equally critical role with efficiency similar to that in canonical GTPases. The presence of a super-secondary structure or a domain would satisfy such a requirement and anchor Switch-II in a position amenable for GTP binding and hydrolysis (depicted as cylinders in Figure 3). Indeed, this concept conforms to the presence of 'anchoring' C-terminal domains in all the subfamilies of cpGTPases (Figure 2), as found from our analysis (see below). This is also noticed in the structures of YlqF (GTP bound) and YjeQ (apo form and GDP bound). YlqF subfamily proteins, the smallest of all cpGTPases in domain composition, possess a CPG-domain and a C-terminal domain (Figure 2). YlqF like proteins are thus the prototypes of cpGTPases, which are unlikely to exist as single GTP-binding domains. This suggests that cpGTPases are a special class that requires the presence of an anchoring C-terminal domain, for it to bind GTP and have efficient catalytic activity (see below).

Each subfamily is characterized by unique domain combinations

Of the four subfamilies of cpGTPases (Figure 2), namely YlqF, YqeH, YjeQ and YawG, the 3D structures of *B.subtilis* YlqF (GTP bound) and two members of YjeQ subfamily from *B.subtilis* (apo form) and *Thermatoga maritima* (GDP bound) have been determined (5,6). The structure of YlqF shows the presence of a CPG-domain and a C-terminal α-helical domain, while YjeQ contains three domains in the following order: an RNA-binding OB fold at the N terminus, a CPG and a unique Zn-binding domain at the C-terminus (5,6). All these structures confirm the presence of an additional C-terminal domain following the CPG-domain.

Profile-based searches (Materials and Methods) were initiated on the cpGTPase sequences in an attempt to assign domains in the regions neighboring the CPG-domain. These searches confirmed YlqF subfamily proteins to be the shortest among all in domain composition, with a CPG-domain followed by a C-terminal α-helical domain. YjeQ subfamily (148 members) concur with the domain composition of YloQ and tmYjeQ proteins, where the CPG-domains are sandwiched by an N-terminal RNA-binding OB fold and C-terminal Zn-binding domain.

For the subfamilies YqeH and YawG, where no structural knowledge is yet available, profile based searches suggest interesting domain compositions. All YqeH like proteins seem to contain a potential N-terminal Zn-finger domain, followed by a CPG-domain and a C-terminal domain (Figure 2). Levdikov *et al*. (5) noted a sequence conservation pattern $\underline{C}$x**G**$\underline{C}$**G**$x_n\underline{C}$x**R**$\underline{C}$ (where x denotes any amino acid and *n* denotes a variable length of x intermittent in the conservation pattern) in this Zn-finger domain. Through secondary structure predictions, we identified $\underline{C}$xG$\underline{C}$G part of the motif to lie between two β-strands and $\underline{C}$xR$\underline{C}$ part of

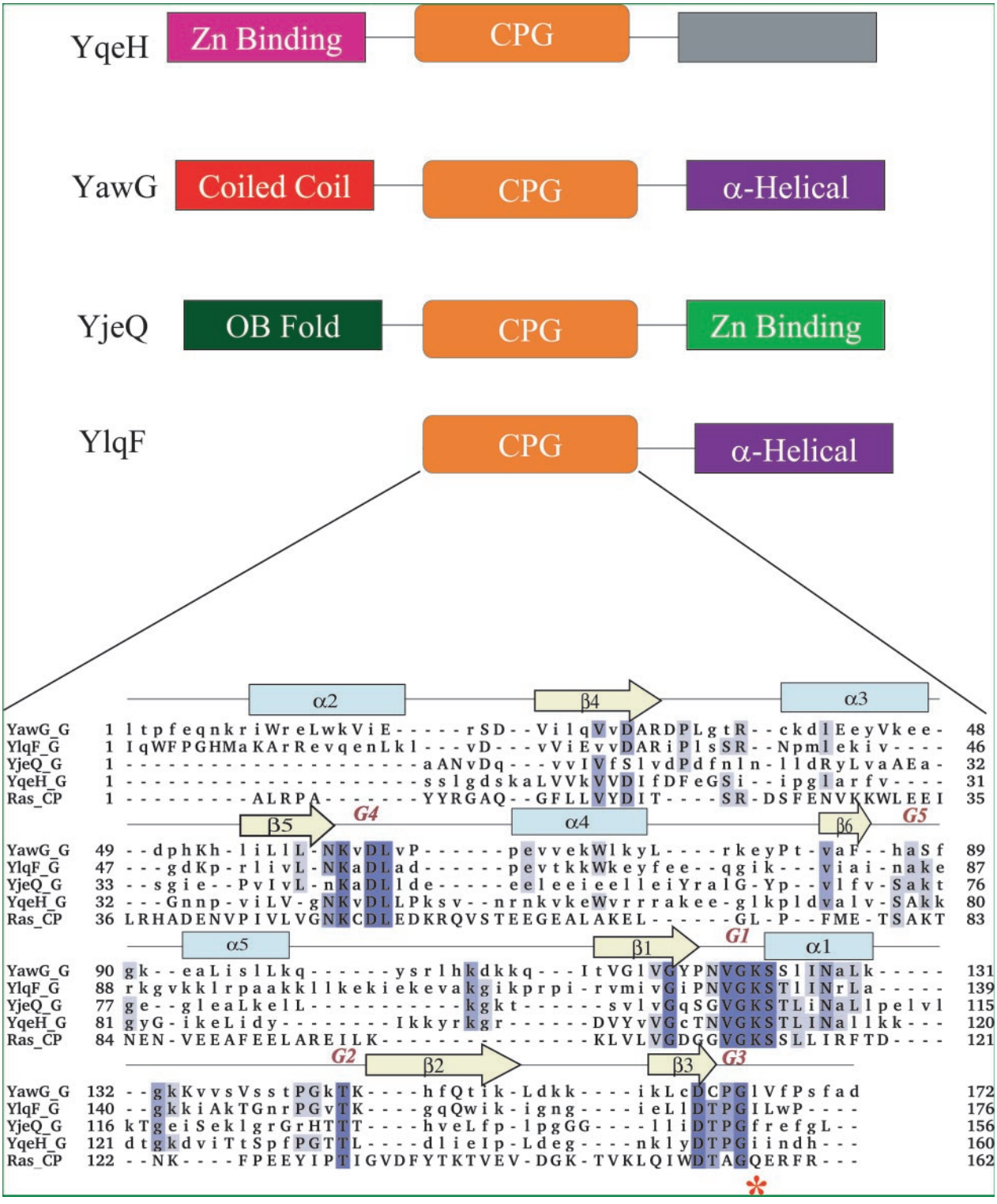

Figure 2. Each subfamily of cpGTPases displays unique domain architecture. Note that the Zinc-binding domains of YjeQ (light green) and YqeH (violet) subfamily are depicted in different colours to show that they belong to different groups of Zn finger motifs. The α-helical domain (blue) of YlqF and YawG subfamily are shown in same colours since they share sequence similarity. It was not possible to assign a domain to the C-terminal region of YqeH (gray) and it appears from the secondary structure prediction that it contains a large proportion of β-strands. The lower panel shows the structure based sequence alignment of consensus sequences representing the CPG-domain of cpGTPase subfamilies along with Ras family, which has been artificially permuted. Alignment position having a probability >0.5 is shown in upper case. The secondary structure labelling is in accordance to Ras (PDB, 5P21). The catalytic residue position corresponding to Gln61 of Ras is indicated by red asterisk. Sequence motifs G1–G5 are marked. Note that there are insertions in G2/Switch-I of cpGTPases in comparison to Ras and the longest insertion is observed in YjeQ subfamily. YlqF and YawG have an additional helix at the N-terminus in contrast to YjeQ and YqeH subfamily. G5 (SAK/L) motif is not well conserved in YlqF and YawG subfamily. Helices are represented as cylinders and beta strands as arrows. Conserved residues are highlighted in blue and the increasing intensity of the blue background reflects the (increasing) extent of conservation at a particular position.

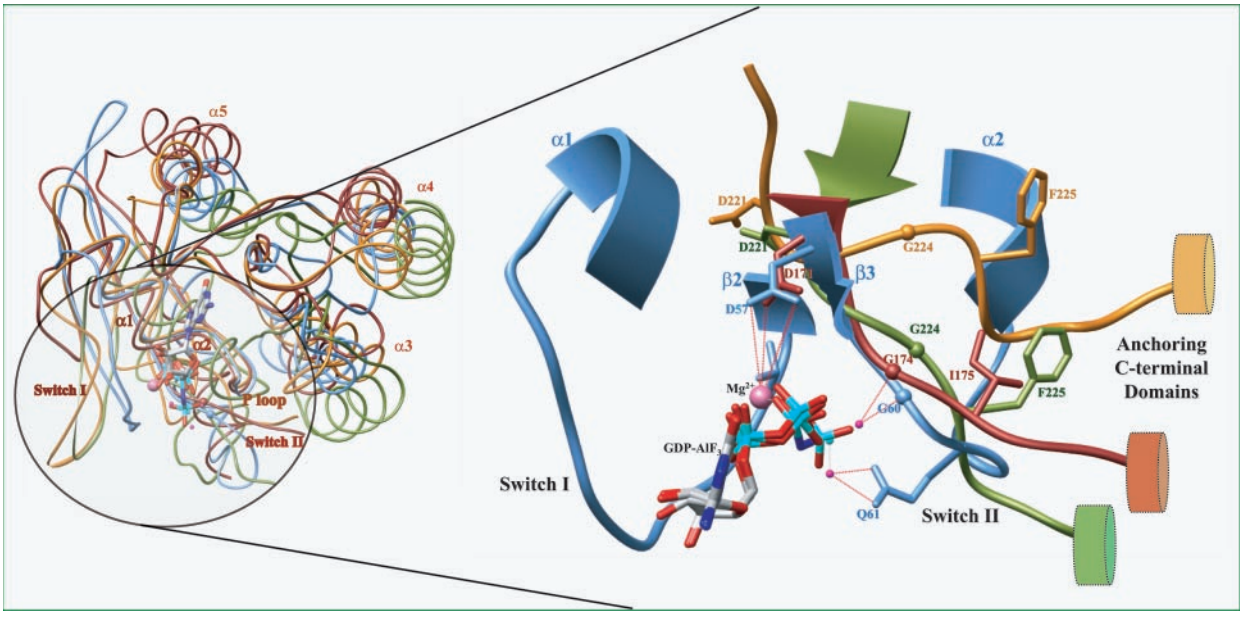

Figure 3. Substrate directed superimposition (wherever applicable) of cpGTPases in apo, GDP and GTP bound forms onto the transition state structure of Ras. The nucleotide-binding region is shown in the inset as a ribbon model. Ras (GDP-AlF_3) is shown in cornflower blue (PDB, 1wq1), YlqF (GTP bound) in brown (PDB, 1puj), YloQ (apo form) in green (PDB, 1t9h) and tmYjeQ (GDP bound) in orange (PDB, 1u0l). Mg^{2+} and the fluoride atoms of AlF_3 (that would correspond to the γ-phosphate oxygens) are shown as small spheres in magenta and pink, respectively. GDP is displayed in sticks with atoms N in blue, C in gray, O in red and P in cyan, respectively. The C-terminal domain is depicted as cylinders. Hydrogen bonds are displayed as red dotted lines. The figure shows that the Switch-II in cpGTPases is repositioned towards the C-terminus in comparison to Ras and it requires the presence of a domain for structural stabilization (see text). The conformation of Asp57 and Gly60 in the Switch-II of Ras and its equivalent residues in YlqF, YloQ and tmYjeQ are indicated. The different positions of Switch-II in the apo form of YloQ, GDP bound form of tmYjeQ and the GTP bound form of YlqF indicate that the conformational changes associated with Switch-II upon GTP/GDP binding (through Asp 57 and Gly 60) may be propagated to the C-terminal anchoring domain or vice versa. The hydrophobic residues I175, F225 and F225 of YlqF, YloQ, tmYjeQ, respectively, correspond to Gln61 of Ras. The figure was prepared using Chimera (45).

the motif at the N-terminus of an α-helix. This structural motif is a characteristic feature of Treble-clef family of Zn finger domains according to the recent classification scheme (24,25). In this family, the structural motif is composed of a Zn knuckle followed by a loop, a β-hairpin and an α-helix. The Zinc knuckle and the N-terminus of the α-helix typically donate two Zn^{2+} ligands each. Based on conserved cysteines, signature motifs and secondary structure predictions, we find that the N-terminal domain of YqeH like proteins can be placed under Treble-clef family Zn finger domains. This fold group consists of proteins with diverse functions such as the ribosomal proteins S14 and L24, RING finger of RAG2, ARF-GAP domain of pyk2-associated protein β, retenoid X receptor α DNA-binding domain, recombination endonuclease VII, to name a few. A MSA of N-terminal region from representative YqeH proteins and the ribosomal proteins S14 and L24 displays these conserved sequence motifs (Supplementary Figure S1). While it was possible to assign such a domain to the N-terminal region, no clear assignment was possible for the C-terminal region of YqeH. Secondary structure predictions revealed a high proportion of β-strands in this 140 amino acid stretch that is likely to form a separate domain.

Like YqeH, YawG subfamily proteins too, contain domains N- and C-terminal to the CPG-domain (Figure 2). Based on secondary structure predictions, it was found that the N-terminal domain of YawG possesses high α-helical content. In order to identify a protein/domain homologous to this region, we initiated a PSI-BLAST search on NR database, which revealed significant homology (Supplementary Figure S2) to the C-terminal region of RNase E from *Vibrio cholerae* (gi:15642032; *E*-value 0.48, 18% sequence identity, 34% sequence similarity with an alignment length coverage of 163 amino acids). A recent study on the C-terminal region of *E.coli* RNase E has shown the presence of a coiled coil structure and this region has been proposed to bind RNA (26). This observation led us to investigate the likelihood for the N-terminal region of YawG to form a coiled coil structure and indeed, the COILS program (27) predicts coiled coil alpha helical structure at the N-terminus of YawG with a high degree of confidence covering ~25 residues. Furthermore, the predicted coiled coil regions show a high degree of sequence similarity and are devoid of gaps in the MSA of YawG family members (Supplementary Figure S2). The similarity to C-terminal RNA-binding region of RNase E and the presence of coiled coil structure brings up an interesting possibility that the N-terminal domain of YawG may be involved in RNA binding.

Secondary structure predictions indicate the presence of α-helical structures in the C-terminal domain of YawG proteins. This is further confirmed by PSI-BLAST searches, which suggest a high sequence homology to the C-terminal α-helical domain of YlqF (average sequence identity of 20% and similarity of 40%). This observation suggests that both YawG and YlqF subfamilies share similar α-helical regions at the C-terminus. Apart from this, no homology seems to prevail among the neighboring domains of the four subfamilies of cpGTPases.

cpGTPases carry a hydrophobic amino acid in place of a catalytic glutamine

We recently studied a group of GTPases that have hydrophobic amino acids substituted in lieu of a catalytic glutamine (Gln^{cat}) of classical GTPases, like Ras (28). These GTPases are termed 'hydrophobic amino acid substituted for catalytic glutamine GTPases' or HAS-GTPases. cpGTPases too, exhibit a similar hydrophobic substitution (highlighted by an asterisk in Figure 2, I175, F225 and F225 in Figure 3). The importance of Gln^{cat} in GTP hydrolysis is now well documented (23) and its mutation to other hydrophobic amino acids is reported to be oncogenic. However, this substitution seems to be a paradox, since HAS-GTPases bind and hydrolyze GTP efficiently (29–31). In HAS-GTPases, the substituted hydrophobic residue retracts away from the nucleotide-binding pocket compared with the Gln^{cat} of Ras and such a retracted conformation is further stabilized by a neighboring hydrophobic environment (28). A similar arrangement is observed in the structures of YjeQ and YlqF and thus, cpGTPases should also be classified as a subclass of HAS-GTPases that are likely to use newer ways to achieve GTP hydrolysis, in comparison to their classical counterparts.

Interestingly, CPG-domains show significant sequence homology to the G-domain of Era family GTPases. This homology becomes appreciable only when the query sequences (e.g. CPG-domains) are manually permuted back to reflect the motif order (G1–G2–G3–G4–G5) as in canonical GTPases. Era family belongs to the universally conserved bacterial GTPases (32,33) that are known to associate with the ribosome. These GTPases fall in the class of HAS-GTPases as they lack the catalytic glutamine (28). In order to establish an evolutionary link between cpGTPases and the Era family, we attempted to construct a phylogenetic tree. Since it is believed that Ras superfamily GTPases evolved recently, as compared with the universally conserved bacterial GTPases (3,32,33), they were included into the phylogenetic analysis together with Era family and cpGTPases. The significant sequence homology between Era family and cpGTPases implies that they are more closely related in evolution. If this presumption is true, then, in the phylogenetic tree, cpGTPases should form a cluster along with Era family rather than with the Ras superfamily. Indeed, an unrooted phylogenetic tree, shown in Figure 4, depicts clearly the branching of cpGTPases with the Era family. This close resemblance suggests that cpGTPases are more closely related in evolution to the universally conserved bacterial GTPases and perhaps indicates a commonality in their function.

DISCUSSION

Probable functional advantage of a circular permutation

We have analyzed the GTPases that exhibit a circular permutation in their G-domain and tried to inquire whether such a permutation renders any functional advantage over the well-studied classical GTPases. Although, in principle several

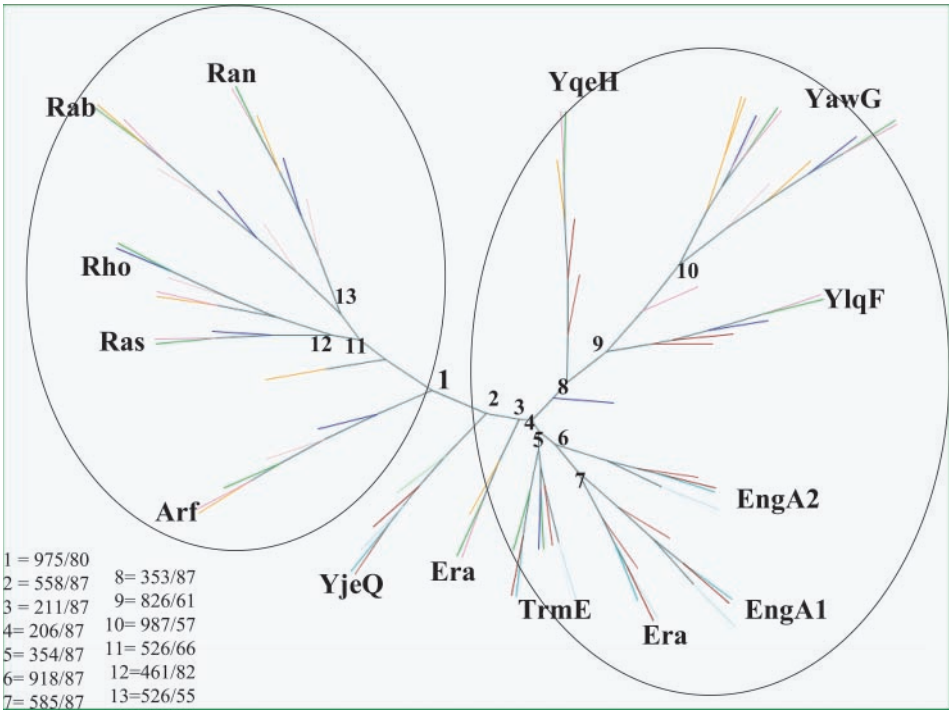

Figure 4. Unrooted tree diagram depicting the evolutionary relationship between the GTPases represented by Ras superfamily, Era family and cpGTPase sub-families. Note that there is a clear clustering of cpGTPases with Era family which is shown to bind RNA. The unrooted tree was generated based on phylogenetic analysis carried out using PHYLIP. Major nodes in the tree are numbered and two corresponding values for each node are provided in the bottom left corner. Of these, the first value provides a support for major branches based on 1000 bootstrap probabilities as implemented in 'protdist/fitch' distance analysis programs of PHYLIP and the second value indicates a reliability value computed using Treepuzzle 5.2. The leaves of the tree depict representative species from each family. The colour code for species is as follows: *Aquifex aeolicus* (dark cyan), *Arabidopsis thaliana* (pink), *B.subtilis* (dark red), *Borrelia burgdorferi* (greenish red), *E.coli* (light red), *Homo sapiens* (green), *Methanococcus jannaschii* (magenta), *Pseudomonas aeruginosa* (light cyan), *S.cerevisiae* (blue), *T.maritima* (brown).

permutations are possible, we found that apart from the canonical G1–G2–G3–G4–G5 (P1 in Figure 1), only the permutation G4–G5–G1–G2–G3 (P4 in Figure 1) is preferred by nature. In the context of evolution, introducing a circular permutation in GTPases seems futile unless imparted with additional functional advantages. In addition to circular permutation, cpGTPases display multi-domain architecture which tempts to raise the following questions. What is the function of such additional domains and what could be the basis for the co-evolution of such domains with the CPG-domain? Examination of the available sequence and structural information on cpGTPases has provided clues to address these questions.

An inevitable consequence of the only permutation observed in nature, is the creation of a new C-terminus following DxxG motif in the Switch-II region of the G-domain (Figures 1 and 3). Switch-I and Switch-II are known to undergo dramatic conformational changes between GTP and GDP bound forms in the canonical GTPases (23). Since stabilizing GTP/GDP in its nucleotide-binding pocket and hydrolysis of GTP requires that Switch-II be held rigidly, it becomes necessary to have an additional domain or a super secondary structure at the C-terminus that can hold and position Switch-II as it may be required (cylinders in Figure 3). Such a structural rearrangement confers two advantages: first, the Switch-II is precisely positioned in an orientation that favours guanine nucleotide binding and thus reduces the high conformational flexibility associated with the loop. Second, the coupling allows the propagation of conformational changes, associated with Switch-II, to the C-terminal domain. Consequently, the 'GTP' or 'GDP' bound form of CPG-domain can directly regulate the biochemical function associated with the C-terminal domain by maneuvering the relative 3D position of Switch-II. As shown in Figure 3, where the structures of YloQ (apo form), tmYjeQ (GDP bound) and YlqF (GTP bound) are superimposed onto the transition state structure of Ras, the positions of Switch-II and the C-terminal domains indicate such a possibility. The nucleotide occupancy at the binding site can propagate conformational changes to the C-terminal domain (depicted in cylinders in Figure 3) through the Switch-II region. Alternatively, the biochemical function of the C-terminal domain may also influence the position of Switch-II and hence nucleotide binding/hydrolysis. On the other hand, the domain N-terminus to the CPG-domain may not sense the conformational changes involving Switch-I and Switch-II merely due to the position it occupies with respect to the CPG-domain. Hence, its biochemical function need not be coupled to nucleotide binding at the CPG-domain and instead be an independent event. These features could in part, confer biochemical merits for the evolution of cpGTPases.

cpGTPases exhibit interesting variations in their domain combinations. In tune with the above arguments, all cpGTPases at least contain a C-terminal domain (Figure 2). In three of the four subfamilies, the domain architecture is as follows: an N-terminal domain followed by CPG-domain and a C-terminal domain. Although in these subfamilies, in principle, CPG-domains can be positioned N-terminus to the protein (i.e CPG-domain at the N-terminus followed by the other two domains), they are always sandwiched between an N- and C-terminal domain (Figure 2). The advantage of such a domain-architecture remains elusive with the limited biochemical information available currently. However, based on the structures of YjeQ (5,6), it seems highly likely that the binding of a cognate molecule to the N-terminal domain, would in turn result in efficient binding and hydrolysis of guanine nucleotides by the CPG-domain, since its Switch-I and the N-terminal domain are positioned close to each other. In such circumstances, a sandwiched CPG-domain seems to present a biochemical advantage to the cpGTPase.

cpGTPases possess RNA-binding domains

Driven by the hypothesis that the nucleotide bound state of CPG-domain would regulate the biochemical function of the C-terminal domain, we attempted to further probe into the possible function of members of the four cpGTPase subfamilies YlqF, YjeQ, YqeH and YawG.

Our analysis of the regions adjacent to CPG-domains of all YjeQ proteins shows that they have an N-terminal OB fold and a C-terminal Zn-binding domain. This is in conformity with the three-dimensional structures of YjeQ (5,6). The OB fold is known to bind RNA (34), thus suggesting a role for YjeQ subfamily in RNA binding. Recently, it was observed that YjeQ interacts with the 30S subunit of ribosome via its N-terminal OB fold (19). Another study hinted that the A site on 30S subunit of the ribosome would be the likely binding site for YjeQ (20). The association of YjeQ with ribosome is also confirmed *in vivo* through chemical genetics study (21). We observe from the structure that although Switch-I is not well defined, the OB fold that is positioned close to it could mediate favourable interactions for nucleotide binding. The importance of suitably positioning a domain N-terminus to CPG-domain could be to provide favorable interactions for nucleotide binding when it interacts with its cognate effector molecule. A further support to this conjecture is the report that GTP hydrolysis rates are enhanced by orders of magnitude when OB fold binds the 30S subunit of ribosome (19–21). These observations point towards an additional advantage of optimally positioning N-terminal domains and Switch-I regions, and it is possible that similar features will be seen in other subfamilies such as YqeH and YawG which have a domain N-terminus to CPG-domain.

Analysis of the sequences of YlqF and YawG subfamilies suggests that they possess a highly similar CPG-domain and C-terminal α-helical domain (Figure 2). In fact, since YlqF proteins lack an N-terminal domain, they may be considered a special subset of YawG subfamily. The difference between YawG and YlqF, apart from the domain organization, seems to be in their localization in eukaryotes. YawG subfamily consists of nuclear/nucleolar and cytoplasmic proteins that are found only in eukaryotes, while YlqF subfamily, which is largely represented in prokaryotes, is also found localized in mitochondria of eukaryotes. Furthermore, the view that YlqF and YawG are closely related is strengthened by the reports that both the subfamilies are implicated in the process leading to ribosome biogenesis (35–37). The N-terminal domain of YawG is predicted to contain a potential coiled coil region. This coiled coil region shows significant sequence homology with the RNA-binding coiled coil region of RNase E (Supplementary Figure S2), thus suggesting that the N-terminal domain could be involved in RNA binding. Indeed, recent studies (36,37) show that YawG homologues, NOG2 and LSG1, associate with 60S pre-ribosomal particle. Though

it is not clear how they bind rRNA, the predominance of basic and aromatic amino acids at the N-terminal domain suggests that like YjeQ subfamily, YawG could use the N-terminal domain to bind RNA. However, no such function could be proposed for the C-terminal domain of YawG subfamily. Given the close relationship to YawG subfamily, one may expect YlqF like proteins too would bind RNA. However, it lacks a similar N-terminal domain and only shares the C-terminal domain with YawG. Furthermore, unlike YawG, no such evidence for the direct association of YlqF with RNA exists. However, a systematic study conducted on MTG1, YlqF ortholog in *S.cerevisiae,* suggests its likely involvement in the assembly of the large subunit of mitochondrial ribosome (35). To seek further evidence, we examined the gene order in the genomes of all YlqF orthologs in bacteria. Conservation of gene order within the gene cluster in prokaryotes can be considered as a selection pressure conferring merits on regulation of gene expression. We found that YlqF shares a neighborhood with RNase H in 19 bacterial genomes (out of 38 microbial genomes). RNase H is an endonuclease that cleaves RNA of the DNA-RNA hybrid. Hence, from the conserved gene order, a plausible physical or functional association between YlqF and RNase H and a tentative role for YlqF in RNA binding cannot be ruled out.

The domain assignment and secondary structure prediction analysis shows that YqeH subfamily proteins are characterized by the presence of N-terminal Treble clef Zn finger motif followed by CPG-domain and a C-terminal domain (Figure 2), which is predominantly composed of β-strands. In conjunction with YjeQ and YawG subfamilies, which are likely to have RNA-binding N-terminal domains, we expected a similar link for YqeH subfamily as well. Treble clef Zn-finger motifs are present in a variety of proteins with diverse functions, of which ribosomal proteins such as S14, L24 use them for RNA binding (Supplementary Figure S1) (38,39). It is possible that the N-terminal domain in YqeH is similarly designed for RNA binding. Interestingly, a high throughput affinity precipitation on yeast proteome revealed the association between FMP38, a YqeH ortholog in yeast, localized in mitochondria (40,41), and the small subunit S5 of mitochondrial ribosomal protein (42). Although the functional role of YqeH proteins needs to be elucidated, our analysis for the first time raises the possibility of the likely involvement of the N-terminal Zn-finger motif in RNA binding.

From the sequence and structural analysis presented here, it appears that cpGTPases, though diverse in their domain architectures, converge to a common theme i.e RNA binding. It seems that YjeQ, YqeH and YawG subfamilies use their N-terminal domain to bind RNA and whether the C-terminal domain also participates in RNA binding remains to be established. However, YlqF subfamily which is the minimal prototype of cpGTPases lacks an N-terminal domain and it is highly likely that it uses the C-terminal domain to bind RNA. In conclusion, RNA binding, or perhaps more precisely, ribosome binding seems to be an emerging theme for all cpGTPases.

cpGTPases are related to ribosome-binding bacterial GTPases

From an evolutionary perspective, cpGTPases would have evolved due to the duplication of a classical GTPase domain followed by deletion of both N- and C-terminal regions (3,43). However, in all the four subfamilies analyzed, the CPG-domain is accompanied by at least one additional domain and does not exist as a single domain. This suggests the likelihood of a domain fusion event following the elimination of N- and C-terminal regions of the duplicated GTPase domain (44). It appears that the co-evolution of neighboring domains along with CPG-domain (Figure 3) confers functional advantages synergistically, in line with the arguments presented here. In cpGTPases, the switching event that is perhaps modulated by guanine nucleotide binding can be propagated to the neighbouring domain through Switch-II (Figure 3). This mode of regulation is novel in contrast to those classical GTPases involved in RNA binding. While it remains to be seen whether RNA binding regulates GTP binding or vice versa, there are interesting evolutionary implications as cpGTPases also possess a hydrophobic substitution like the HAS-GTPases (28). Most HAS-GTPases are also universally conserved bacterial GTPases that are believed to be progenitor GTPases (32,33). These bacterial GTPases are important regulators of ribosome function and nucleic acid binding. Interestingly, we also observe a significant sequence homology between the CPG-domains and the G-domains of the bacterial GTPases and this is further reinforced in the phylogenetic analysis (Figure 4) wherein, cpGTPases are clustered with the Era family GTPases as against the Ras superfamily GTPases. This suggests that cpGTPases could have been a part of the ancient ribosome-binding bacterial GTPases regulating important cellular processes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

B.A. acknowledges AICTE (All India Council for Technical Education) for the National Doctoral Fellowship. SKV acknowledges UGC (University Grants Commission) for the Senior Research Fellowship. This work is supported by Wellcome Trust, UK (Grant No- GR-073616) in the form of an 'International Senior Research Fellowship in Biomedical Science' awarded to B.P. and grants from DST (Department of Science and Technology) and MHRD (Ministry of Human Resource and Development), India. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust, UK.

NOTE ADDED IN PROOF

While this work was under review, Matsuo et. al. show that YlqF associates with the 50S Ribosomal Subunit (J. Biol. Chem., Vol. 281, 8110-8117, 2006). This conclusion is in agreement with our present work in which we raise the possibility that YlqF like the other cpGTPases should bind RNA (see the Discussion section).

Conflict of interest statement. None declared.

REFERENCES

1. Bourne,H.R., Sanders,D.A. and McCormick,F. (1990) The GTPase superfamily: a conserved Switch for diverse cell functions. *Nature*, **348**, 125–132.

2. Sprang,S.R. (1997) G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.*, **66**, 639–678.
3. Leipe,D.D., Wolf,Y.I., Koonin,E.V. and Aravind,L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.*, **317**, 41–72.
4. Pandit,S.B. and Srinivasan,N. (2003) Survey for G-proteins in the prokaryotic genomes: prediction of functional roles based on classification. *Proteins*, **52**, 585–597.
5. Levdikov,V.M., Blagova,E.V., Brannigan,J.A., Cladiere,L., Antson,A.A., Isupov,M.N., Seror,S.J. and Wilkinson,A.J. (2004) The crystal structure of YloQ, a circularly permuted GTPase essential for *Bacillus subtilis* viability. *J. Mol. Biol.*, **340**, 767–782.
6. Shin,D.H., Lou,Y., Jancarik,J., Yokota,H., Kim,R. and Kim,S.H. (2004) Crystal structure of YjeQ from *Thermotoga maritima* contains a circularly permuted GTPase domain. *Proc. Natl Acad. Sci. USA*, **101**, 13198–13203.
7. Morimoto,T., Loh,P.C., Hirai,T., Asai,K., Kobayashi,K., Moriya,S. and Ogasawara,N. (2002) Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. *Microbiology*, **148**, 3539–3552.
8. Arigoni,F., Talabot,F., Peitsch,M., Edgerton,M.D., Meldrum,E., Allet,E., Fish,R., Jamotte,T., Curchod,M.L. and Loferer,H. (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.*, **16**, 851–856.
9. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.
10. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
11. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
12. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
13. Galtier,N., Gouy,M. and Gautier,C. (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, **12**, 543–548.
14. Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M. and Barton,G.J. (1998) Jpred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–893.
15. Von Mering,C., Jensen,L.J., Snel,B., Hooper,S.D., Krupp,M., Foglierini,M., Jouffre,N., Huynen,M.A. and Bork,P. (2005) STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res.*, **33**, D433–D437.
16. Price,M.N., Huang,K.H., Alm,E.J. and Arkin,A.P. (2005) A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res.*, **33**, 880–892.
17. Felsenstein,J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.*, **266**, 418–427.
18. Strimmer,K. and von Haeseler,A. (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**, 964–969.
19. Daigle,D.M. and Brown,E.D. (2004) Studies of the interaction of *Escherichia coli* YjeQ with the ribosome *in vitro*. *J. Bacteriol.*, **186**, 1381–1387.
20. Himeno,H., Hanawa-Suetsugu,K., Kimura,T., Takagi,K., Sugiyama,W., Shirata,S., Mikami,T., Odagiri,F., Osanai,Y., Watanabe,D. *et al.* (2004) A novel GTPase activated by the small subunit of ribosome. *Nucleic Acids Res.*, **32**, 5303–5309.
21. Campbell,T.L., Daigle,M. and Brown,E.D. (2005) Characterization of the *Bacillus subtilis* GTPase YloQ and its role in ribosome function. *Biochem J.*, **389**, 843–852.
22. Richardson,J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.*, **34**, 167–339.
23. Vetter,I.R. and Wittinghofer,A. (2001) The guanine nucleotide-binding Switch-In three dimensions. *Science*, **294**, 1299–1304.
24. Grishin,N.V. (2001) Treble clef finger–a functionally diverse zinc-binding structural motif. *Nucleic Acids Res.*, **29**, 1703–1714.
25. Krishna,S.S., Majumdar,I. and Grishin,N.V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.*, **31**, 532–550.
26. Callaghan,A.J., Aurikko,J.P., Ilag,L.L., Gunter Grossmann,J., Chandran,V., Kuhnel,K., Poljak,L., Carpousis,A.J., Robinson,C.V., Symmons,M.F. *et al.* (2004) Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *J. Mol. Biol.*, **340**, 965–979.
27. Lupas,A., Van Dyke,M. and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
28. Mishra,R., Gara,S.K., Mishra,S. and Prakash,B. (2005) Analysis of GTPases carrying hydrophobic amino acid substitutions in lieu of the catalytic glutamine: implications for GTP hydrolysis. *Proteins*, **59**, 332–338.
29. Hwang,J. and Inouye,M. (2001) An essential GTPase, Der, containing double GTP binding domain from *Escherichia coli* and *Thermotoga maritima*. *J. Biol. Chem.*, **276**, 31415–31421.
30. Sullivan,S.M., Mishra,R., Neubig,R.R. and Maddock,J.R. (2000) Analysis of guanine nucleotide binding and exchange kinetics of *Escherichia coli* GTPase Era. *J. Bacteriol.*, **182**, 3460–3466.
31. Yamanaka,K., Hwang,J. and Inouye,M. (2000) Characterization of GTPase activity of TrmE, a member of novel GTPases superfamily, from *Thermotoga maritima*. *J. Bacteriol.*, **182**, 7078–7082.
32. Caldon,C.E., Yoong,P. and March,P.E. (2001) Evolution of a molecular Switch: universal bacterial GTPases regulate ribosome function. *Mol. Microbiol.*, **41**, 289–297.
33. Caldon,C.E. and March,P.E. (2003) Function of the universally conserved bacterial GTPases. *Curr. Opin. Microbiol.*, **6**, 135–139.
34. Draper,D.E. and Reynaldo,L.P. (1999) RNA binding strategies of ribosomal proteins. *Nucleic Acids Res.*, **27**, 381–388.
35. Barrientos,A., Korr,D., Barwell,K.J., Sjulsen,C., Gajewski,C.D., Manfredi,G., Ackerman,S. and Tzagoloff,A. (2003) MTG1 codes for a conserved protein required for mitochondrial translation. *Mol. Biol. Cell*, **14**, 2292–2302.
36. Saveanu,C., Bienvenu,D., Namane,A., Gleizes,P.E., Gas,N., Jacquier,A. and Fromont-Racine,M. (2001) Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. *EMBO J.*, **20**, 6475–6484.
37. Hedges,J., West,M. and Johnson,A.W. (2005) Release of the export adapter, Nmd3p, from the 60S ribosomal subunit requires Rpl10p and the cytoplasmic GTPase Lsg1p. *EMBO J.*, **24**, 567–579.
38. Wimberly,B.T., Brodersen,D.E., Clemons,W.M.,Jr, Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
39. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
40. Sickmann,A., Reinders,J., Wagner,Y., Joppich,C., Zahedi,R., Meyer,H.E., Schonfisch,B., Perschil,I., Chacinska,A., Guiard,B. *et al.* (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl Acad. Sci. USA*, **100**, 13207–13212.
41. Huh,W.K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
42. Gavin,A.C., Bosche,M., Krause,R., Grandi,P., Marzioch,M., Bauer,A., Schultz,J., Rick,J.M., Michon,A.M., Cruciat,C.M. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.
43. Ponting,C.P. and Russell,R.B. (1995) Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem. Sci.*, **20**, 179–180.
44. Yanai,I., Wolf,Y.I. and Koonin,E.V. (2002) Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol.*, **3**, research0024.1–0024.13.
45. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.