Nucleotide Binding Switches the Information Flow in Ras GTPases

Francesco Raimondi1,2, Guillem Portella3, Modesto Orozco3,4, Francesca Fanelli1,2*

1 Department of Chemistry, University of Modena and Reggio Emilia, Modena, Italy, 2Dulbecco Telethon Institute (DTI), University of Modena and Reggio Emilia, Modena, Italy, 3 Molecular Modeling and Bioinformatics Unit, IRB-BSC Joint Research Program in Computational Biology, Institute for Research in Biomedicine, and Barcelona Supercomputing Center, Barcelona, Spain, 4 National Institute of Bioinformatics, Parc Científic de Barcelona, Barcelona, Spain

Abstract

The Ras superfamily comprises many guanine nucleotide-binding proteins (G proteins) that are essential to intracellular signal transduction. The guanine nucleotide-dependent intrinsic flexibility patterns of five G proteins were investigated in atomic detail through Molecular Dynamics simulations of the GDP- and GTP-bound states (SGDP and SGTP, respectively). For all the considered systems, the intrinsic flexibility of SGDP was higher than that of SGTP, suggesting that Guanine Exchange Factor (GEF) recognition and nucleotide switch require higher amplitude motions than effector recognition or GTP hydrolysis. Functional mode, dynamic domain, and interaction energy correlation analyses highlighted significant differences in the dynamics of small G proteins and Ga proteins, especially in the inactive state. Indeed, SGDP of Ga is characterized by a more extensive energy coupling between nucleotide binding site and distal regions involved in GEF recognition compared to small G proteins, which attenuates in the active state. Moreover, mechanically distinct domains implicated in nucleotide switch could be detected in the presence of GDP but not in the presence of GTP. Finally, in small G proteins, functional modes are more detectable in the inactive state than in the active one and involve changes in solvent exposure of two highly conserved amino acids in switches I and II involved in GEF recognition. The average solvent exposure of these amino acids correlates in turn with the rate of GDP release, suggesting for them either direct or indirect roles in the process of nucleotide switch. Collectively, nucleotide binding changes the information flow through the conserved Ras-like domain, where GDP enhances the flexibility of mechanically distinct portions involved in nucleotide switch, and favors long distance allosteric communication (in Ga proteins), compared to GTP.

Citation: Raimondi F, Portella G, Orozco M, Fanelli F (2011) Nucleotide Binding Switches the Information Flow in Ras GTPases. PLoS Comput Biol 7(3): e1001098. doi:10.1371/journal.pcbi.1001098

Copyright: © 2011 Raimondi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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

* E-mail: fanelli@unimo.it

Introduction

The Ras superfamily comprises many guanine nucleotide-binding proteins (G proteins) that are essential to intracellular signal transduction [1,2]. These proteins act biologically as molecular switches cycling between ON and OFF states, thereby controlling a variety of processes ranging from cell growth and differentiation to vesicular and nuclear transport [1]. The switch-on process requires the release of the bound Guanosine Di-Phosphate (GDP) and the subsequent binding of the Guanosine Tri-Phosphate (GTP), an intrinsically slow process catalyzed by Guanine nucleotide Exchange Factors (GEFs) [1]. In the GTP-bound state (SGTP), the G proteins display high affinity for binding downstream effectors, interactions through which they exert their specific biologic functions. The switch-off process involves the hydrolysis of GTP to GDP, reaction that is accelerated by Guanine nucleotide Activating Proteins (GAPS) and leads to release of effector proteins (due to reduced affinity) and attenuation of downstream signaling [1]. Peculiar to the members of the Ga family is the fact that in the inactive GDP-bound state (SGDP) they form membrane-associated αβγ heterotrimers, with GDP bound to the α-subunit [3].

Nucleotide switch and hydrolysis in Ras GTPases are played by the conserved core, Ras-like domain (see Figure 1 legend for structure description). Selected members of the superfamily, such as the members of the Ga family, hold an extra-Ras α-helical domain (Figures 1 and 2) constituted by a long central helix surrounded by five shorter helices. This feature makes Ga proteins significantly larger than all the other members of the Ras superfamily, which are, hence, indicated as “small G proteins”.

The central role of Ras GTPases in cell function is testified by the proved involvement of selected members like Ras and Rho in many aspects of cancer development and tumor progression, which makes these proteins very interesting targets in cancer therapy [2]. This is why oncogenic Ras mutants have been the target of computational experiments aimed at unraveling the dynamic information encoded into the structure [4,5].

In a recent study, we combined Elastic Network Model (ENM) coarse grained simulations with Principal Component Analysis (PCA) on the experimental structures of representatives Ras...
Author Summary

The Ras superfamily comprises many guanine nucleotide-binding proteins (G proteins) that are essential to intracellular signal transduction. These proteins act biologically as molecular switches cycling between ON and OFF states, thereby controlling a variety of processes ranging from cell growth and differentiation to vesicular and nuclear transport. In spite of the extremely high biological and medical relevance of the Ras GTPase superfamily, a comprehensive structural/dynamic view of the trans-family and family-specific functioning mechanisms is still lacking. In this study, we gained insights into the functional dynamics of Ras GTPases by deciphering the dynamic information encoded in the topology of these proteins depending on the nucleotide-bound state, i.e. GDP- or GTP-bound (S\textsubscript{GDP} and S\textsubscript{GTP}, respectively). Collectively, nucleotide binding changes the information flow through the conserved Ras-like domain, where GDP enhances the flexibility of mechanically distinct portions involved in nucleotide switch, and favors long distance allosteric communication (in G\textsubscript{z} proteins), compared to GTP. Functional dynamics is instrumental in GDP switch, for the members of the G\textsubscript{z} family, different from small G proteins, requires allosteric communication between nucleotide and Guanine Exchange Factor binding sites.

GT\textsubscript{P}ases to decipher the physical and evolutionary deformability patterns that enable switching between active and inactive states [6,7]. The analysis highlighted functional separation of the conserved core into two lobes, as previously suggested by others [5,8]. The deformation modes involved in the switching function are conserved along evolution and are localized in lobe 1 portions close to the nucleotide (Figure 1). These modes lead to functional specialization when associated with evolution-driven deformations of protein portions essentially located in lobe 2, distal from the nucleotide, and involved in specific interactions with membrane, GEFs, or effectors [6]. Additional evidence that the Ras superfamily members share a set of switching dynamics was inferred by the identification of conserved hinge points throughout all subfamilies, which remark the bi-lobate dynamics of the conserved core [6].

In this study, we analyze in atomistic detail the intrinsic flexibility patterns of G proteins through Molecular Dynamics (MD) simulations of selected members of the five major families of Ras GTPases (i.e. Arf1, G\textsubscript{z}, also called transducin), Sec4, H-Ras, and RhoA). Functional dynamics of these molecular switches was investigated through the analyses of essential motions, functional modes, interaction energy correlations and dynamic domains in relation to the functional states, i.e. S\textsubscript{GDP} or S\textsubscript{GTP}. Our simulations confirm the complexity of the deformability patterns of Ras GT\textsubscript{P}ases and the careful tuning that evolution has made on it. In G proteins, functional dynamics is suggested to be instrumental in GDP switch, which, for the members of the G\textsubscript{z} family, different from the small G proteins, requires allosteric communication between nucleotide and GEF binding sites [3,9,10].

Results

Analysis of the intrinsic flexibility of Ras GT\textsubscript{P}ases

Remarkable functional-state dependent features shared by the five homologous GT\textsubscript{P}ases reside in the Cz-root Mean Square Fluctuation (Cz-RMSF) profiles, showing higher flexibility for the inactive forms compared to the active ones, with sw\textsubscript{I} and sw\textsubscript{II} being the lobe 1 portions that better define the intrinsic trans-family flexibility, as they account for major differences between S\textsubscript{GDP} and S\textsubscript{GTP} (Table 1 and Figure 3). Such an increase in flexibility of sw\textsubscript{I} and sw\textsubscript{II} has already been shown for Ras p21, as a feature linked to the absence of the \gamma-phosphate [11]. For G\textsubscript{z}, the higher flexibility of S\textsubscript{GDP} resides also in the z-helical domain (Figures 3 and S1). Such a flexibility trend between inactive and active states is mostly shown also by the average Cz-root Mean Square Deviations (Cz-RMSD; Table S1), which for all GT\textsubscript{P}ases but H-Ras is higher in S\textsubscript{GDP} compared to S\textsubscript{GTP}.

The analysis of concerted motions was carried out both on single and on concatenated Cz-atom trajectories of the inactive and active forms. Consistent with the Cz-RMSF profiles, the total variance computed by summing the eigenvalues from PCA on single trajectories is always higher for the inactive forms compared to the active ones (Table 1, Figure 3). As for PCA on the concatenated trajectories, the first eigenvector (i.e. principal component 1; PC1) separated the structures visited along the trajectory into two clusters, corresponding to S\textsubscript{GDP} and S\textsubscript{GTP} (this separation is especially clear for all small G-proteins; Figure 4 and Figure S1). Displacements along second and third eigenvectors (PC2 and PC3) generally reflect the higher motility of S\textsubscript{GDP} compared to S\textsubscript{GTP} (Figure 4). For the small G proteins, Cz-atom projections along the three PCs concern collective motions of sw\textsubscript{I}, sw\textsubscript{II}, and inter-switch (Figures S1, S2, and S3). In contrast, for G\textsubscript{z}, the collective motions of switches and inter-switch are associated with motions of the \alpha/\beta loops and of the z-helical domain (Figures S1, S2, and S3).

Collectively, Ras GT\textsubscript{P}ases share a similar essential dynamics that is more amplified in S\textsubscript{GDP} and involves sw\textsubscript{I}, sw\textsubscript{II}, and inter-switch. Such a functional state-dependent dynamics is presumably linked to the nucleotide switch mechanism that pertains to S\textsubscript{GDP}. The additional essential motions differentiating G\textsubscript{z} from the small G proteins may be considered as expressions of family-specific functional dynamics finalized to the nucleotide switch.

Functional modes in the small G proteins are associated with changes in solvent accessibility of lobe 1 portions

In an attempt to find a trans-family structural indicator of the differences between S\textsubscript{GDP} and S\textsubscript{GTP}, we computed the Solvent Accessible Surface Area (SASA) of all the highly conserved amino acids in the G-boxes (Figure 1). The analysis highlighted the SASA computed over T\textsubscript{G24} and G\textsubscript{C34} (SASA\textsubscript{TG}; see Figure 1B legend for the numbering explanation) as the only functional state descriptor valid for all the five GT\textsubscript{P}ase families (Table 1). Indeed, the SASA\textsubscript{TG} averaged over all the trajectory frames (SASA\textsubscript{TGavg} in Table 1) or plotted as time series (Figure 5) is always greater for the inactive form than for the active one. This is due to the breakage of the interactions between the side chain of T\textsubscript{G24} and both the Mg\textsuperscript{2+} ion and a \gamma-phosphate oxygen atom, and between the backbone nitrogen atom of G\textsubscript{C34} and a \gamma-phosphate oxygen atom on going from S\textsubscript{GTP} to S\textsubscript{GDP} (Figure 6). Such a breakage of intermolecular interactions is expected to contribute, at least in part, to the higher flexibility of the two switches in the inactive state compared to the inactive one (Figure 3), suggesting the existence of a link between SASA\textsubscript{TG} and RMSF profiles.

According to the results of Functional Mode Analysis (FMA) (see Methods), for the small G proteins SASA\textsubscript{TG} correlates extremely well with the first twenty PCs derived from PCA on the single Cz-atom trajectories of S\textsubscript{GDP}, G\textsubscript{z} showing a lower correlation coefficient (Table 1). Collectively, correlations are lower for S\textsubscript{GTP} compared to S\textsubscript{GDP} (Table 1).

The Cz-atom projections of the linear combination of the first twenty PCs shows that the portions that contribute the most to the
Figure 1. Structure and sequence features of the five GTPases. A: cartoons of the H-Ras structure (PDB code: 5P21) in its GTP-bound state are shown. The Ras superfamily GTPases share a common domain, the Ras-like domain. The latter, according to CATH [34], is characterized by a Rossmann fold with a 3-layer(aba) sandwich architecture, where helices $a_1$ and $a_5$; the secondary structure elements in the Ras-like domain are labeled according to the Noel’s nomenclature [35]) lay on one side, whereas $a_2$, $a_3$, and $a_4$ lay on the other side of the central five-stranded parallel b-sheet (i.e. comprising the $b_1$ and $b_3$-$b_6$ strands, Figures 1 and 2). The helices $a_1$ and $a_3$ lay on the opposite side of the sheet due to the inversion in the order of the preceding strands, $b_1$ and $b_3$, respectively, which are adjacent to each other. The $b_1$-$b_3$ loop, i.e. phosphate binding loop (P loop), and the region comprising $a_2$ as well as the preceding and following loops (i.e. switch II (swII)) participate in the binding of the nucleotide phosphates (Figures 1 and 2). The architecture of this superfamily is such that $b_1$ is also adjacent to $b_4$. The $b_1$-$b_4$ interface divides the Ras-like domain into two lobes: i) lobe 1 (i.e. the N-terminal half of the domain, magenta) includes the $b_1$-$b_3$ strands, the P-loop and the two switches, and ii) lobe 2 (blue), which includes the $b_4$-$b_6$ strands and the $a_3$-$a_5$ helices. Another structural feature of the conserved Ras domain is that $b_2$ forms a b-hairpin with $b_3$, the loop that connects the two antiparallel strands being directed towards the opposite side of the nucleotide binding cleft (Figures 1 and 2) [3]. The $b_2$-$b_3$ hairpin is also called “inter-switch” (i.e. delimited by a green oval) because the loops that enter $b_2$ and exit from $b_3$ constitute, respectively, the swI and swII regions. The loops connected to the C-term of $b_1$ and the N-term of $a_2$-helical and Ras-like domain constitutes the nucleotide binding cleft. Incidentally, among the small G proteins RhoA has a structural peculiarity consisting of a ten amino acid $\alpha$-helical insertion ($\alpha$-insert) on the $b_5$-$a_4$ loop like the $\alpha$G segment shared by the members of the G family. B: the multiple sequence alignment derived from the multiple structure alignment of representatives of the GTP state of Arf1 (PDB code: 1O3Y), G$_{t}$ (PDB code: 1TND), Sec4 (PDB code: 1G17), H-Ras (PDB code: 5Q21), and RhoA (PDB code: 1KMQ) is shown (i.e. achieved by the Multiprot-Staccato software) [36]. Helices, strands, and loops are, respectively, violet, yellow, and cyan. Ultra-conserved sequences involved in nucleotide binding (G boxes) are delimited by black boxes. Black numbers on the left side of the alignment refer to the sequential numbering, whereas black numbers above the sequences indicate the beginning of a secondary structure/G box motif. The fully conserved residues in such boxes are red and marked by an asterisk. In order to facilitate trans-family comparisons of the MD simulation outputs, an arbitrary numbering was set characterized by the label of the secondary structure segment followed by the amino acid position in that segment. In those cases where the G-boxes overlap with the secondary structure segment, positions refer to the G-boxes.

doi:10.1371/journal.pcbi.1001098.g001
combined essential motions in the S\textsuperscript{GDP} state include sw1, swII, and inter-switch (Figure 7). G\textsubscript{2} shows additional collective motions of \alpha-helical domain and \beta/\beta' loop, whereas RhoA is characterized by additional essential motions of the \beta-insert (Figures 1 and 7). Remarkably, the highly conserved amino acids that contribute to SASA\textsubscript{TG} lay just on sw1 (T\textsuperscript{G2/3}) and on swII (G\textsuperscript{G3/4}). Incidentally, in small G proteins, the movement of sw and, by a lower extent, of swII, marked, respectively, by the increase in solvent accessibility of T\textsuperscript{G2/3} and G\textsuperscript{G3/4} on going from S\textsuperscript{GTP} to S\textsuperscript{GDP}, is instrumental in GEF recognition (Figure 6 and Figure S4). The fact that those modes, which correlate with SASA\textsubscript{TG}, concern portions of the small G proteins implicated in GEF recognition, and that correlations pertain mostly to the S\textsuperscript{GDP} state, which is engineered to recognize GEFs, suggests that the correlated modes are, indeed, functional modes related to the nucleotide switch mechanism in the small G proteins. These inferences are also supported by the correlation between SASA\textsubscript{TG} and rate of GDP release (Table 1), the linear correlation coefficient for the five considered systems being 0.86) [12–17]. In this respect, Sec4 shows the highest SASA\textsubscript{TG} and GDP release rates in the absence of GEF (Table 1).

**Comparative analyses of interaction energy correlations**

Further insights into the dynamic properties of S\textsuperscript{GDP} and S\textsuperscript{GTP} involved in functional specialization of selected families were gained through the analysis of correlated non-bonded interaction energies. Highly correlated interacting pairs are markers of protein regions that communicate between each other. A functional state-independent feature of the five G proteins detected by correlated energies is the lack of interaction energy correlations between \beta4 amino acids and any amino acid from lobe 1 (Figures S5 and S6), which reflects the low flexibility of the \beta-strand [6]. Other functional-state independent features, shared by Arf1 and G\textsubscript{2}, are the diffuse energy correlations involving \beta4/\beta6 loop and both lobe 1 and lobe 2 portions (Figures S5 and S6).

The patterns of correlated interaction energies show that the representative members of the five G protein families considered in this study, less evident in Arf1, share intra-lobe 1 energy correlations in their inactive state, which are lost upon nucleotide exchange. On the contrary, inter-lobe energy coupling between \alpha3, on one side, and swI and inter-switch, on the other one, is more marked in S\textsuperscript{GTP} compared to S\textsuperscript{GDP}. A singularity of H-Ras with respect to the other small G proteins is the energy coupling between \beta2/\beta3 turn and C-term of \alpha5, which is more marked in S\textsuperscript{GDP} compared to S\textsuperscript{GTP} (Figure S5).

Family specialization in the context of the energetic coupling between amino acid pairs is more evident by comparing G\textsubscript{2} with the four small G proteins (Figures S5 and S6). In this respect, a singularity of G\textsubscript{2} is the extended coupling between C-term of \alpha5 and lobe 1 portions in S\textsuperscript{GDP}, which is lost upon activation (Figure S6). Other singularities of the inactive state of G\textsubscript{2}, is the inter-domain coupling between \alphaE/linker1, on one side, and inter-switch, swII, \alpha3, \beta4/\beta6 loop and \alpha5, on the other, which is more marked in the inactive state. Finally, in S\textsuperscript{GDP} of G\textsubscript{2}, the \alpha/\beta loops that are involved in GEF recognition show more marked energy coupling compared to the \beta/\beta' loops. Taken together, these features reflect the bi-domain structural organization of the protein as well as the singular GEF recognition mode and GEF-catalyzed nucleotide switch mechanism, which makes the difference from the small G proteins.

We also analyzed functional-state dependent changes in the patterns of the amino acid pairs whose interaction energies correlate with the pairwise interaction energies involving the nucleotide (i.e. amino acid pairs energetically coupled with nucleotide-mediated interaction(s)). Incidentally, the average interaction energy profile of the nucleotide in the two different functional states of the five representatives shows remarkable trans-family conservation (Figure S7). In more detail, for both the S\textsuperscript{GDP} and S\textsuperscript{GTP} forms of the small G proteins, the G1 and G4 boxes (Figure 1) give, respectively, the first and second strongest contributions in terms of attractive interaction energy values, whereas the G2 and G3 boxes contribute only in the S\textsuperscript{GTP} state. Peculiar to both functional forms of G\textsubscript{2}, is the fact that the G2 box gives the second strongest contribution due to R174 (G\textsuperscript{G2/3}).

In general, in S\textsuperscript{GDP} of the small G proteins, the amino acids involved in direct interaction with the nucleotide essentially lay on \alpha1 and the P-loop, whereas in S\textsuperscript{GTP}, the nucleotide interaction sphere extends to sw1 and swII amino acids (Figure 8). Upon activation, for Arf1, Sec4, and RhoA the energy coupling between nucleotide binding site and other protein portions increases. In contrast, for H-Ras, GTP substitution for GDP reduces the network of electrostatic interactions energetically coupled with GDP found between the \beta2/\beta3 hairpin and \alpha5 (i.e. including the salt bridges between D47 in the \beta2/\beta3 turn and both R161 (G\textsuperscript{G2/3}) and R164 (G\textsuperscript{G2/3}) and between E49 (G\textsuperscript{G2/3}) and R164 (G\textsuperscript{G2/3}) (Figure 8).

**Table 1. Correlations between shape descriptor and PCs.**

| Family | State | Var\textsuperscript{a} | R\textsuperscript{b} | SASA\textsubscript{TGmax} (Å\textsuperscript{2})\textsuperscript{c} | GDP release\textsuperscript{d} |
|--------|-------|-----------------|-----------------|--------------------------|------------------|
| Arf1   | S\textsuperscript{GDP} | 91.00 | 0.85 | 91.71±20.64 | 0.04 [12] |
|        | S\textsuperscript{GTP} | 65.63 | 0.32 | 8.34±2.37 | |
| G\textsubscript{1} | S\textsuperscript{GDP} | 390.37 | 0.60 | 87.32±18.21 | 0.00072 [13] |
|        | S\textsuperscript{GTP} | 285.11 | 0.40 | 16.85±4.48 | |
| Sec4   | S\textsuperscript{GDP} | 157.88 | 0.86 | 151.27±29.10 | 0.21 [14] |
|        | S\textsuperscript{GTP} | 79.31 | 0.52 | 58.95±4.05 | |
| H-Ras  | S\textsuperscript{GDP} | 87.19 | 0.79 | 120.88±21.58 | 0.025 [16] |
|        | S\textsuperscript{GTP} | 69.96 | 0.65 | 20.93±7.04 | |
| RhoA   | S\textsuperscript{GDP} | 111.70 | 0.83 | 107.68±20.76 | 0.0072 [17] |
|        | S\textsuperscript{GTP} | 103.93 | 0.57 | 20.49±6.88 | |

\textsuperscript{a}Total variance obtained by summing the eigenvectors from PCA.

\textsuperscript{b}Correlation coefficient between SASA\textsubscript{TG} and a combination of the first twenty PCs.

\textsuperscript{c}SASA\textsubscript{TG} index averaged over the 40000 frames constituting the 40 ns trajectories.

\textsuperscript{d}Rate of GEF-independent GDP release (min\textsuperscript{-1}); the relative bibliographic source is in square brackets.

doi:10.1371/journal.pcbi.1001098.t001
The energetic coupling in \( \text{G}_{\text{at}} \) shows a singular behavior compared to the small G proteins (Figures 8 and 9). In fact, both \( \text{SGDP} \) and \( \text{SGTP} \) show a significantly higher number of correlated pairs compared to the small G proteins (Figure 9). These pairs are both inter- and intra-domain located and undergo a change in distribution upon activation (Figure 9). As for \( \text{SGDP} \), inter-domain correlated pairs essentially involve R172 (in the \( \beta2/\beta3 \) loop) and the fully conserved D268 (in the \( \alpha1/\alpha2 \) loop), which interacts also with the nucleotide, as well as D146 (in the \( \alpha3/\alpha4 \) loop) and K256 (in the \( \beta3/\beta4 \) loop) (Figure 9). As for the intra-\( \alpha \)-helical domain-located pairs, only ten

---

**Figure 3. RMSF profile from MD trajectories of the \( \text{SGDP} \) and \( \text{SGTP} \) forms of the five Ras GTPases.** Green and red lines refer to the \( \text{SGDP} \) and \( \text{SGTP} \) forms, respectively, of Arf1, \( \text{G}_{\text{at}} \), Sec4, H-Ras, and RhoA. RMSF profiles refer to the 40000 frames constituting 40 ns trajectories. The secondary structure elements are shown on the abscissa, following nomenclature and color code described in Figure 1.

doi:10.1371/journal.pcbi.1001098.g003
Comparative analysis of dynamic domains

The five G proteins (in both active and inactive forms) were also compared by the Dynamic Domain (DD) method, which clusters the Cx-atom of the system according to their propensity to be part of mechanically coherent domains in the trajectory frames [18]. Not surprisingly, the first cluster separates active vs inactive G-proteins, stressing the existence of differential effects of GDP and GTP binding on G protein dynamics. In S\textsubscript{GDP} of the small G proteins, mechanistically distinct domains include swI and swII, which are directly implicated in GEF recognition and the nucleotide switch, whereas S\textsubscript{GTP} does not show common clustering patterns. Furthermore, the S\textsubscript{GDP} of Ga\textsubscript{q} shows unique features since the \(\alpha\)-helical domain appears mechanically distinct from the Ras-like domain (Figure 10).

Discussion

MD simulations done in this study helped unraveling the functional dynamics of Ras superfamily GTPases, providing atomistic details that were not reached by previous evolutionary or coarse-grained studies [6]. Indeed, current simulations clearly show that the intrinsic flexibility of S\textsubscript{GDP} is higher than that of S\textsubscript{GTP}, suggesting that GEF recognition and nucleotide switch mechanism require higher amplitude motions than effector recognition or GTP hydrolysis.

Novel trans-family features pertaining to functional dynamics were inferred from the analysis of interaction energy correlations. The latter revealed intra-lobe 1 correlations in all the five G proteins; such correlations attenuated upon activation. This may relate with lobe 1 being heavily involved in function-representation dynamics. Furthermore, interaction energy correlations highlighted the largest contribution of correlations involving the \(\beta\)4 segment in both functional states of the five G proteins. This feature remarks the bi-lobate nature of the conserved Ras-like domain, which is related to \(\beta\)4 being the holder of the strongest and most conserved hinge point [6].

The results of this study suggest that, in small G proteins, functional modes, i.e. collective motions directly related to function, are more evident in the inactive state rather than in the active one. These modes, which involve swI and swII in lobe 1, correlate with changes in solvent exposure of T\textsuperscript{G2:4} and G\textsuperscript{G3:4}, which, in the small G proteins, with emphasis on T\textsuperscript{G2:4}, are involved in GEF recognition. These results, together with the existence of a correlation between SASSA\textsubscript{TRGSY} and rate of GDP release, suggest that the two conserved G-box amino acids participate either directly or indirectly in the mechanism of nucleotide switch in small G proteins. Functional modes involving swI and swII are less detectable for Ga\textsubscript{q} proteins, suggesting that biologically relevant modes in large and small G proteins are different. Indeed, the analyses of essential motions and dynamic domains support previous inferences that in Ga\textsubscript{q} proteins the nucleotide switch involves concerted motions of the \(\alpha\)-helical domain with respect to the Ras-like domain, following allosteric GEF recognition by the \(\alpha\)/\(\beta\) loops [19]. Consistent with these results, the inactive state of Ga\textsubscript{q} is characterized by a significantly more extensive communication between nucleotide binding site and distal regions involved in GEF recognition compared to the small G proteins. This was inferred from S\textsubscript{GDP} of Ga\textsubscript{q}, showing an evident energy coupling between nucleotide binding site and distal portions like the \(\beta\)2/\(\beta\)3 turn, \(\alpha\)5, and the three \(\alpha\)/\(\beta\) loops, which participate in GEF recognition [3]. Remarkably, this energy coupling is absent in S\textsubscript{GTP}, in which function, i.e. effector binding and GTP-hydrolysis, does not require long distance communication as it involves regions like swI and swII that participate in the nucleotide binding site. Thus, the energy coupling between nucleotide and GEF binding sites likely pertains to the nucleotide switch mechanism in Ga\textsubscript{q} proteins and not in small G proteins. The evident communication between nucleotide binding site and inter-switch-C term of Ga\textsubscript{q} is in line with the results of previous simulations of the receptor-G protein complex suggesting that the receptor-induced detachment of \(\alpha\)5 from the inter-switch promotes a cascade of structural changes in Ga\textsubscript{q} that propagate from the C-term to the \(\alpha\)-helical domain through lobe 1 portions of the Ras-like domain [19]. These changes, indeed, culminate with the formation of a nucleotide exit route in between the \(\alpha\)2/\(\alpha\)5 loop [19].

The analysis of the interaction energy correlations highlights novel family-specific features. These include interaction energy correlations involving \(\alpha\)/\(\beta\) loop and the remaining portions of Arf1 and Ga\textsubscript{q}, related to the higher length and flexibility of this loop in the two proteins compared to the other three G proteins. Moreover, a singularity of H-Ras compared to the other small G proteins is a reduction in energy coupling between nucleotide binding site and distal regions like \(\alpha\)5, on going from S\textsubscript{GDP} to S\textsubscript{GTP}. This may relate, at least in part, to the postulated implication of inter-switch and \(\alpha\)5 of H-Ras in a novel switch mechanism involving a nucleotide-independent change in the membrane anchoring of the protein, operated through \(\alpha\)4 and the hypervariable region (HVR) in \(\alpha\)5 [20].

In summary, in Ras GTPases, the intrinsic dynamics oriented to functional specialization essentially pertains to the inactive state rather than to the active one and clearly separates the small G proteins from Ga\textsubscript{q}. Indeed, mechanically distinct domains implied in the mechanism of nucleotide switch could be detected in the presence of GDP but not in the presence of GTP. S\textsubscript{GDP} is engineered to respond to the GEF’s request to release GDP and this follows different mechanisms in the small G proteins compared to the members of the Ga\textsubscript{q} family. Whereas in small G proteins GEF binds directly to swI and swII, which are implicated in nucleotide exit, in the Ga\textsubscript{q} family GEF binds to protein portions that are distal from the nucleotide, thus implying allosteric communication for GEF-catalyzed nucleotide release. The latter is expected to involve also a concerted motion of the \(\alpha\)-helical domain with respect to the Ras-like domain, feature intrinsic to S\textsubscript{GDP} and amplified by GEF. In contrast, for S\textsubscript{GTP} of all G proteins, in which nucleotide and effector binding sites are quite
Figure 5. Time series of the SASA index. Time series of the SASA index computed over $T^{(G2-4)}$ and $G^{(G3-4)}$ (SASABP) are shown for the $S^{GDP}$ (green lines) and $S^{GTP}$ (red lines) representatives of the five considered GTPases.

doi:10.1371/journal.pcbi.1001098.g005
Figure 6. Cartoons of three different functional forms of the five GTPases. Left, central, and right panels show, respectively, the GDP-, GEF-, and GTP-bound forms of Arf1, Sec4, H-Ras, and RhoA. The GEF protein is colored cyan with helices represented as cylinders. The SASA computed on T(G2:4) and G(G3:4) is shown by green dots. The T(G2:4) side chain and the nucleotide are represented as sticks. Dashed lines indicate the distances between either the side chain oxygen atom of T(G2:4) or the backbone oxygen atom of G(G3:4) and an interacting partner on the GEF molecule.

doi:10.1371/journal.pcbi.1001098.g006
Figure 7. Cα-atoms projections along the first 20 PCs. The Cα-atoms projections along the linear combination of the first twenty PCs from the trajectories referred to the $S^{\text{GDP}}$ (left) and $S^{\text{GTP}}$ (right) states are shown. doi:10.1371/journal.pcbi.1001098.g007
close, the functioning mechanism does not require long distance communication nor inter-domain motions that are, hence, attenuated compared to S\textsubscript{GDP}. In this respect, the intrinsic flexibility of small G proteins and G\textsubscript{z} proteins shares more commonalities in the active state compared to the inactive one.

Current and previous results [6] suggest that the Ras superfamily utilizes a hierarchical organization of its structural flexibilities; lobe 1 motions associated with its switching function must be retained in order to accomplish the primary G protein function of changing its affinity for GEFs, effector proteins, and GAPs through different bound nucleotides, but additional motions across both lobes of the protein are family-specific and play a role in determining the unique functional characteristics of specific members. Molecular communication between lobe 1 and lobe 2 portions (intra-Ras communication) and between \alpha-helical and Ras-like domains, for G\textsubscript{z}, represents the way to accomplish functional specialization.

Taken as a whole the results of this investigation reveal that the topology of the conserved Ras-like domain is such that it allows for differential flow of information depending on the bound nucleotide, i.e. GDP enhances the flexibility of mechanically distinct portions involved in nucleotide switch, and favors long distance allosteric communication (in G\textsubscript{z} proteins), compared to GTP.

**Methods**

**Simulated proteins**

The following S\textsubscript{GDP} and S\textsubscript{GTP} representatives of the Arf1, G\textsubscript{z}, Sec4, H-Ras, and RhoA families were selected as input of MD simulations: Arf1 (PDB codes: 1HUR and 1O3Y), G\textsubscript{z} (PDB codes: 1TAG and 1TND), Sec4 (PDB codes: 1G16 and 1G17), H-Ras (PDB codes: 4Q21 and 5Q21), and RhoA (PDB codes: 1FTN and 1KMQ) based on the following criteria: i) homogeneity of the molecular specie, i.e. same sequence, for the two functional states, and ii) existence of a crystal structure for the isolated state, i.e. not in complex with other proteins. In a few cases, wild-type structures (in one or both S\textsubscript{GDP} and S\textsubscript{GTP} states) were generated by manipulation of mutant structures: L71Q mutant for Arf1 S\textsubscript{GTP}, N25F, and N25F/L63Q for the S\textsubscript{GDP}, and S\textsubscript{GTP} states of RhoA, respectively. Due to the intrinsic GTPase activity, the GTP-bound forms had been crystallized with an hydrolysis-resistant ion originally present in the 1G16 structure. The two different metal ions are expected to be interchangeable with no substantial structural effects [21].

Set-up of MD simulations

MD simulations on the five representative Ras GTPases (Arf1, G\textsubscript{z}, Sec4, H-Ras, and RhoA) were carried out using the GROMACS4 simulation package [22] with the AMBER03 all molecular species, i.e. same sequence, for the two functional states, and 1KMQ) based on the following criteria: i) homogeneity of the molecular specie, i.e. same sequence, for the two functional states, and ii) existence of a crystal structure for the isolated state, i.e. not in complex with other proteins. In a few cases, wild-type structures (in one or both S\textsubscript{GDP} and S\textsubscript{GTP} states) were generated by manipulation of mutant structures: L71Q mutant for Arf1 S\textsubscript{GTP}, N25F, and N25F/L63Q for the S\textsubscript{GDP}, and S\textsubscript{GTP} states of RhoA, respectively. Due to the intrinsic GTPase activity, the GTP-bound forms had been crystallized with an hydrolysis-resistant analogue of GTP, i.e. GppNHp for 5P21, 1G17, and 1KMQ, and GTP\textsubscript{y}S for 1TND. In our simulations the GTP analogues were converted to GTP.

Finally, all the simulated systems contained the Mg\textsuperscript{2+} ion together with the coordinating water molecules. In the case of the GDP-bound form of Sec4, the Mg\textsuperscript{2+} ion substitutes for the Co\textsuperscript{2+} ion originally present in the 1G16 structure. The two different metal ions are expected to be interchangeable with no substantial structural effects [21].

**Analysis of trajectories**

Trajectories were subjected to a variety of analyses. RMSDs, RMSF, and SASA were performed using the Wordom software [29]. Detailed information concerning the essential motions of the proteins was obtained by using Essential Dynamics (ED) analyses. In this respect, covariance matrices (C\textsubscript{z}-atoms in our case) were built, using either isolated or concatenated MD trajectories, and diagonalized to a set of eigenvectors (describing the essential deformation modes) and associated eigenvalues (which explain the amount of each-mode variance). As for ED on the concatenated trajectories, the C\textsubscript{z}-atom structure averaged over all the concatenated frames was used as a reference for the building of the co-variance matrix. In order to facilitate comparison along
$G_{\alpha_t}$
different proteins, a minimum consensus length for the inactive and active states was chosen for each family, i.e. 18-177 for Arf1, 29-339 for Gαz, 20-183 for Sec4, 3-165 for H-Ras, and 5-178 for RhoA.

The search for correlations between structural features and essential modes was carried out through the FMA [30] tool, by using the Linear Mutual Information (LMI) estimator [31]. In this framework, a number of size/shape and intermolecular interaction descriptors were correlated with linear combinations of a variable number of PCs.

The dynamical/mechanical properties of the simulated systems were characterized by means of the DD method implemented into the PCASUITTE package (http://mmb.pcb.ub.es/software/pcasuite/) [32].

Non bonded interaction energies involving each protein residue, the nucleotide, and the Mg\(^{2+}\) ion were monitored every 20 ps along the trajectory. Only the non bonded interactions whose average values along the simulations were greater than 2 kcal mol\(^{-1}\) were considered relevant for the analysis and were processed to find potential pair-correlations according to the following equation [33]:

\[
C_{ijkl} = \frac{\sum_{i=1}^{N} (E_{ij} - \bar{E}_{ij})(E_{kl} - \bar{E}_{kl})}{\sqrt{\sum_{i=1}^{N} (E_{ij} - \bar{E}_{ij})^2 \sum_{i=1}^{N} (E_{kl} - \bar{E}_{kl})^2}}
\]

(1)

Where \(E_{ij}\) and \(\bar{E}_{ij}\) (as well as \(E_{kl}\) and \(\bar{E}_{kl}\)) stand for instantaneous (i.e. frame\(^a\)-associated) and average interaction energy values between particles i and j (as well as k and l), \(A \times N\) correlation matrix is thus built, where \(N\) is the total number of relevant interacting pairs. Only the elements with absolute correlation coefficient \(\geq 0.4\) were further considered in the analysis. Note that by mapping the inter-atomic energy correlation matrix into the residue space, a residue correlation matrix can be derived:

\[
RC_{ij} = \frac{1}{k} \sum_{m=1}^{N} \sum_{n} |C_{mn} \times \delta_{mn}^{ij}| \quad (2)
\]

where \(N\) is the dimension of the energy correlation matrix and \(k\) is the number of relevant interactions, whose correlation value is above the correlation threshold, in which residues \(i\) and \(j\) are involved. \(\delta_{mn}^{ij}\) is 1 if residue \(i\) and \(j\) are involved in interactions with \(m\) and \(n\) (\(n\) and \(m\)), otherwise \(\delta_{mn}^{ij}\) is set to 0.

**Supporting Information**

**Figure S1** Cβ-atoms projections along PC1. Cβ-atoms projections along the PC1 from the trajectories obtained by concatenating the trajectories relative to the S\(^{\text{GDP}}\) (left panel) and S\(^{\text{GTP}}\) (right panel) bound representatives of the five families are shown. A number of conformations were generated by displacing the Cβ-atoms of the first frame of the S\(^{\text{GDP}}\) and S\(^{\text{GTP}}\) trajectories from the minimum to the maximum displacements observed along PC1 in the relative cluster. For S\(^{\text{GDP}}\) and S\(^{\text{GTP}}\), the color changes, respectively, from green to blue, and from red to blue. Labels mark the structural portions in the Ras-like domain, which are more involved in such displacements.

Found at: doi:10.1371/journal.pcbi.1001098.s001 (3.13 MB TIF)

**Figure S2** Cβ-atoms projections along PC2. Cβ-atoms projections along the PC2 from the trajectories obtained by concatenating the trajectories relative to the S\(^{\text{GDP}}\) (left panel) and S\(^{\text{GTP}}\) (right panel) bound representatives of the five families are shown. See the legend to Figure S1 for an explanation of this figure.

Found at: doi:10.1371/journal.pcbi.1001098.s002 (3.22 MB TIF)

**Figure S3** Cβ-atoms projections along PC3. Cβ-atoms projections along the PC3 from the trajectories obtained by concatenating the trajectories relative to the S\(^{\text{GDP}}\) (left panel) and S\(^{\text{GTP}}\) (right panel) bound representatives of the five families are shown. See the legend to Figure S1 for an explanation of this figure.

Found at: doi:10.1371/journal.pcbi.1001098.s003 (3.11 MB TIF)

**Figure S4** Superimposed structures of the S\(^{\text{GDP}}\) (green), S\(^{\text{GTP}}\) (red), and GEF-bound forms of the four small G proteins.

Found at: doi:10.1371/journal.pcbi.1001098.s004 (2.64 MB TIF)

**Figure S5** Interaction energy correlations for the small G proteins. The inter-residue interaction energy correlation matrices for Arf1, Sec4, H-Ras, and RhoA are shown. The dimension of each symmetric matrix corresponds to the number of residues shared by the two functionally different states of the protein. Each column or row represents a specific residue. The regions above and below the matrix main diagonal concern S\(^{\text{GDP}}\) and S\(^{\text{GTP}}\), respectively. The secondary structure elements are shown, following nomenclature and color code described in Figure 1. Interaction energy and correlation coefficient cutoffs of 2 kcal mol\(^{-1}\) (in absolute value) and \(\geq 0.4\), respectively, were employed in the analysis. The color scale is from 0 (blue) to 1 (red), where 0 corresponds to a 0.4 correlation coefficient.

Found at: doi:10.1371/journal.pcbi.1001098.s005 (0.81 MB TIF)

**Figure S6** Interaction energy correlations for Gαz. The inter-residue interaction energy correlation matrix for Gαz is shown. The dimension of the symmetric matrix corresponds to the number of residues shared by the two functionally different states of the protein. Each column or row represents a specific residue. The regions above and below the matrix main diagonal concern S\(^{\text{GDP}}\) and S\(^{\text{GTP}}\), respectively. The secondary structure elements are shown, following nomenclature and color code described in Figure 1. Interaction energy and correlation coefficient cutoffs of 2 kcal mol\(^{-1}\) (in absolute value) and \(\geq 0.4\), respectively, were employed in the analysis. The color scale is from 0 (blue) to 1 (red), where 0 corresponds to a 0.4 correlation coefficient.

Found at: doi:10.1371/journal.pcbi.1001098.s006 (0.71 MB TIF)

**Figure S7** Average interaction energy profiles of the nucleotide. Green and red lines refer to S\(^{\text{GDP}}\) and S\(^{\text{GTP}}\), respectively of Arf1,
Gzα, Sec4, H-Ras, and RhoA. The secondary structure elements are shown on the abscissa, following nomenclature and color code described in Figure 1.

Table S1  Details of the simulated systems and average Cα-RMSD.

Found at: doi:10.1371/journal.pcbi.1001098.s008 (0.04 MB PDF)

References

1. Vetter IR, Wittinghofer A (2001) The guanine nucleotide-binding switch in three dimensions. Science 294: 1299–1304.
2. Colicelli J (2004) Human RAS superfamily proteins and related GTPases. Sci STKE 2004: RE13.
3. Oldham WM, Hamm HE (2006) Structural basis of function in heterotrimeric G proteins. Q Rev Biophys 39: 117–166.
4. Futatsugi N, Tsuda M (2001) Molecular dynamics simulations of Gly-12—>Val mutants of G12α: dynamic inhibition mechanism. Biophys J 81: 3483–3489.
5. Gorfe AA, Grant BJ, McCammon JA (2008) Mapping the nucleotide and isoform-dependent and structural dynamics of ras proteins. Structure 16: 895–896.
6. Raimondi F, Orozco M, Fanelli F (2010) Deciphering the deformation modes associated with protein function retention and specialization in members of the Ras superfamily. Structure 18: 402–414.
7. Van Wynsberghe AW, Cui Q (2010) Conservation and variation of structural flexibility in protein families. Structure 18: 281–283.
8. Gram BJ, Gorfe AA, McCammon JA (2009) Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics. PLoS Comput Biol 5: e1000325.
9. Oldham WM, Hamm HE (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. Nat Rev Mol Cell Biol 9: 60–71.
10. Bay Y, Ugrn O, Liu T, Hilser VJ, Onaran O (2008) Exploring aistoric coupling in the alpha-subunit of heterotrimeric G proteins using evolutionary and ensemble-based approaches. BMC Struct Biol 8: 23.
11. Ma JP, Karplus M (1997) Ligand-induced conformational changes in ras p21: A normal mode and energy minimization analysis. J Mol Biol 274: 114–131.
12. Paris S, Berard-Dafour S, Robinseau S, Bigay J, Antony B, et al. (1997) Role of protein-phospholipid interactions in the activation of ARF1 by the guanine nucleotide exchange factor Arno. J Biol Chem 272: 22221–22226.
13. Thomas TO, Bae H, Medkova M, Hamm HE (2001) An intramolecular contact in Galpha transducin that participates in maintaining its intrinsic GDP release rate. Mol Cell Biol Res Commun 4: 282–291.
14. Kabeckel AK, Goud B, Northrup JK, Novick PJ (1999) Binding and hydrolysis of guanine nucleotides by Sec61p, a yeast protein involved in the regulation of vesicular traffic. J Biol Chem 274: 9566–9572.
15. Feuerstein J, Kalbitzer HR, John J, Goody RS, Wittinghofer A (1987) Characterisation of the metal-ion-GDP complex at the active sites of transforming and nontransforming p21 proteins by observation of the 170-Mn superhyperfine coupling and by kinetic methods. Eur J Biochem 162: 49–55.
16. Neal SE, Eccleston JF, Hall A, Webb MR (1988) Kinetic analysis of the hydrolysis of GTP by p21-N-ras. The basal GTPase mechanism. J Biol Chem 263: 19718–19722.
17. Hutchinson JP, Eccleston JF (2000) Mechanism of nucleotide release from Rho by the GDP dissociation inhibitor protein. Biochemistry 39: 11348–11359.
18. Naviez I, Cailliez F, Lavery R (2004) Probing protein mechanisms: Residue-level properties and their use in defining domains. Biophys J 87: 1426–1435.
19. Raimondi F, Sebeer M, De Benedetti PG, Fanelli F (2006) Mechanisms of inter-and intramolecular communication in GPCRs and G proteins. J Am Chem Soc 130: 4310–4325.
20. Abuikova D, Hanzal-Bayer M, Arishti N, Flossman SJ, Gorfe AA, et al. (2008) A novel switch region regulates H-ras membrane orientation and signal output. Embo J 27: 727–735.
21. Stroupe C, Brunger AT (2000) Crystal structures of a Rab protein in its inactive and active conformations. J Mol Biol 304: 585–590.
22. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4: 435–447.
23. Case DA, Cheatham TE, Darden T, Merz KM, Simmerling C, et al. (2005) The Amber biomolecular simulation programs. J Comput Chem 26: 1668–1688.
24. Sorin EJ, Parse VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J 88: 2472–2493.
25. Meagher KL, Redman LT, Carlson HA (2003) Development of polyphosphate parameters for use with the AMBER force field. J Comput Chem 24: 1016–1025.
26. Hess B, Bekker H, Berendsen HJC, Fraaije JGM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18: 1463–1472.
27. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126: 041101.
28. Berendsen HJC, Postma JPM, Vlugteren WF, Dinola A, Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 81: 3884–3890.
29. Secker M, Cecchini M, Rao F, Setzam G, Caflich A (2007) Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics 23: 2625–2627.
30. Hub JS, de Groot BL (2009) Detection of functional modes in protein dynamics. PLoS Comput Biol 5: e1000480.
31. Lange OF, Grauhmuller H (2006) Generalized correlation for biomolecular dynamics. Proteins 62: 1053–1061.
32. Meyer T, Ferrer-Costa C, Perez A, Rueda M, Bidon-Chanal A, et al. (2006) Embo J 27: 727–735.
33. Oldham WM, Hamm HE, Sigler PB (1993) The 2.2 A crystal structure of transducin-GDP. Nature 366: 654–663.
34. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, et al. (1997) CATH - a hierarchic classification of protein domain structures. Structure 5: 1993–1998.
35. Noel JP, Hamm HE, Sigler PB (1993) Two distinct lower symmetry axes in the Cα structures of transducin-α and transducin-β. Structure 1: 23–29.
36. Shatskii M, Nussinov R, Widlon HJ (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56: 143–156.

Acknowledgments

We are grateful to Michele Sebeer for his technical support. We also thank Ongun Onaran for helpful discussion.

Author Contributions

Conceived and designed the experiments: FR FF. Performed the experiments: FR GP. Analyzed the data: FR FF. Wrote the paper: FR MO FF.