Published in final edited form as:

J Mol Biol. 2009 July 3; 390(1): 142–153. doi:10.1016/j.jmb.2009.05.001.

The charge-dipole pocket: a defining feature of signaling pathway GTPase on-off switches

Andrew F. Neuwald

Institute for Genome Sciences and Department of Biochemistry & Molecular Biology, University of Maryland School of Medicine, HSF-I, Room 134, 20 Penn Street, Baltimore, MD 21201 USA

Abstract

Ras-like GTPases function as on-off switches in intracellular signaling pathways. Their on- or off-state is communicated through conformational changes in the so-called switch I and II regions. It is commonly believed that the distinguishing molecular features of these GTPases are well known. Here, however, I identify—through a Bayesian iterative analysis of GTPase evolutionary divergence—a previously undescribed switch II structural component that (along with previously described, functionally critical residues) most distinguish these signaling pathway on-off switches from other GTPases. In certain Ras-like GTPases this newly-identified component forms an aromatic pocket around the negative-dipole moment at the end of a switch II helix with a positively-charged residue inserted into the pocket. This helix is oriented in a specific direction away from the GTPase core, but is dramatically reoriented upon disruption of the charge-dipole pocket. The charge-dipole pocket occurs in both the on- and off-states and both the charge-dipole pocket and an alternative configuration occur within the unit cell of a single crystal structure of Rab5a GTPase in the off-state. Thus, the charge-dipole pocket configuration is closely associated, not with the on- or off-state, but rather with formation of the outward-oriented helix and, as a result, with restructuring of the switch II N-terminal region, which plays a critical role both in sensing the on-off state and in mediating GTP hydrolysis and nucleotide exchange.

Keywords

G protein alpha subunit; Rab; Arf-like; P-loop; Ran

Introduction

Advances in high-throughput sequencing technologies present the opportunity to infer aspects of biological mechanisms from vast amounts of sequence data—the cell's own language for encoding those mechanisms. Such inferences rely on the assumption that sequence patterns that have been conserved for a billion years or more reflect important structural or mechanistic similarities. More relevant to the analysis here, divergent patterns that are conserved in descendent proteins maintaining a particular divergent function likewise reflect important structural or mechanistic differences. Examining functionally divergent residues within available crystal structures thus can identify the molecular features responsible for important

Tel: 410-706-6724; Fax: 410-706-1482, E-mail: E-mail: aneuwald@som.umaryland.edu.

Supplemental Data: Supplementary Data associated with this article can be found, in the online version, at doi: xxx.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

functional distinctions and, as a result, suggest specific mechanisms to test experimentally. Moreover, such patterns may reflect aspects of underlying mechanisms that thus far have evaded detection owing to intrinsic experimental limitations. Just as careful statistical analyses of patterns of inherited traits allowed early geneticists to identify and characterize genes prior to the emergence of molecular biology, a statistical analysis of residue patterns may help identify and characterize the components of a protein's molecular machinery prior to the development and application of more direct methods.

Here I describe a Bayesian iterative strategy for comprehensive analysis of protein evolutionary divergence (Fig. 1a). This strategy can: (i) identify and accurately align vast numbers of related sequences; (ii) identify, based on statistically rigorous criteria, evolutionarily divergent patterns and the sequences conserving those patterns; and (iii) identify corresponding protein structural features. A procedure for performing step (i) of this strategy is introduced here (see Materials and Methods); procedures for performing steps (ii) and (iii) have been described previously¹. As a stringent test of its ability to identify important but previously undescribed molecular features, here I apply this strategy to the well characterized class of phosphate-binding loop (P-loop) GTPases and, as a proof of concept, present one outcome of this analysis: the identification of a previously undescribed structural component that is highly distinctive of Ras-like GTPase on-off switches.

P-loop GTPases bind to GDP or GTP through amino acid residues that show up at the sequence level as highly conserved motifs². These include the Walker A (G-K-[ST]) motif³, which corresponds to the P-loop⁴, and a Walker B (D-x-x-G) motif, the conserved aspartate (D) of which interacts (indirectly through a water molecule) with the Mg⁺⁺ ion coordinating with guanine nucleotide phosphate groups. An important subgroup of P-loop GTPases is the extended Ras-like superfamily⁵, which includes Ras, Rab, Rho/Rac, Ran, Arf, Arf-like (Arl), and Sar GTPases and α subunits of heterotrimeric G proteins (Fig. 1b) and which are referred to here as the Ras-like GTPases. Ras-like GTPases function as on-off switches within eukaryotic signaling pathways that regulate diverse cellular processes, including vesicle transport, embryonic development, the sensation of vision, odor, taste and pain, microtubule assembly and cell division. These GTPases are associated both with guanine nucleotide exchange factors, which turn them on by mediating the exchange of GTP for GDP, and with GTPase activating proteins, which turn them off by stimulating hydrolysis of GTP to GDP. The on- or off-state of a Ras-like GTPase is communicated through conformational changes within their switch I and II regions, which detect the presence or absence of the γ phosphate of bound guanine nucleotide (Fig. 1c).

Results and Discussion

Ras-like GTPase functional divergence

Using the strategy described in Fig. 1a and in Materials and Methods, approximately a quarter of a million P-loop GTPase sequences were identified, multiply aligned, classified into evolutionarily divergent categories and, when atomic coordinates were available, functionally divergent residues were structurally analyzed. After removing nearly-identical and fragmented sequences, the final alignment contained 91,406 sequences (the most conserved features of which are highlighted in Fig. S1 of Supplemental Data). As mentioned in the Introduction, the focus here is on the most characteristic features of Ras-like GTPases. These were identified by optimally splitting this final P-loop GTPase alignment (using Bayesian partitioning with pattern selection⁶; Fig. 1a) into two major evolutionarily-divergent subgroups: Ras-like GTPases versus other P-loop GTPases (plus a few atypical Ras-like GTPases). Fig. 2 highlights the divergent residues that most distinguish typical Ras-like GTPases from other P-loop GTPases.

Ras-like GTPase residues with previously proposed functions

As expected, some of the residues highlighted in Fig. 2 were previously proposed to perform critical functions in Ras-like GTPases. Because the aim here is to identify Ras-like residues of unknown function, these residues merely serve as positive controls for this analysis and thus are mentioned only briefly here. These residues are located near the γ -phosphate of bound GTP and include: (i) A threonine residue (Thr43-Rab11a in Fig. 2), which is believed to mediate Switch I conformational changes and which coordinates with the Mg^{++} ion that is associated with guanine nucleotide phosphate groups⁷. (ii) A glutamine residue (Gln70-Rab11a), which is proposed to play a critical catalytic role in GTP hydrolysis⁷. (iii) An acidic residue (Glu71-Rab11a), which typically interacts with the Walker A lysine whenever Mg^{++} or guanine nucleotide is absent and which has been proposed to play a role in Mg^{++} release and nucleotide exchange^{8; 9}. (iv) A glycine or alanine residue (Ala68-Rab11a), for which-in GTPases harboring an alanine at this position-the side-chain methyl group is believed to block the Mg^{++} binding site and thereby promote release of the Mg^{++} ion during nucleotide exchange¹⁰. And (v) an arginine residue within the switch II region (Arg74-Rab11a), which, for Arf, Arl and G α GTPases, was proposed to sense the GTP-bound state¹¹.

A Switch II C-terminal (swII-CT) structural component

In addition to residues with previously proposed functions, Fig. 2 reveals three additional patterns specific to Ras-like GTPases: (i) the pattern [RK]-x-[ILV] preceding the P-loop, (ii) the pattern [WF] directly preceding the Walker B aspartate residue (this aspartate is highlighted in Fig. S1 of Supplementary Data), and (iii) the pattern [YF]-[YF] at the C-terminal end of the switch II region. (Incidentally, among the five deleterious Rab GTPase missense mutations thus far identified in humans¹², one changes the tryptophan associated with the [WF] pattern to glycine resulting in Griscelli syndrome, which is typified by pigmentation defects and unusual activation of lymphocytes and macrophages.) Roughly two-thirds of all Ras-like GTPases match all three patterns exactly and about one-half (mostly Rab and Ran GTPases) conserve the most typical pattern residues. Rho/Rac, Ras and G α GTPases, however, conserve alternative residues at one or more pattern positions, presumably owing to additional functional divergence. Although these patterns occur within three distinct regions in the sequences, within available crystal structures the corresponding residues cluster near the switch II region's C-terminal (swII-CT) end and exhibit diverse structural configurations, which appear to be facilitated by alternative interactions involving the aromatic residues of the [WF] and [YF]-[YF] patterns.

Interactions between aromatic units play a central role in molecular recognition and self-assembly^{13; 14} and can accommodate diverse conformational forms inasmuch as aromatic rings can interact in four distinct geometric orientations: displaced face-to-face, face-to-edge, parallel staggered, and herringbone¹⁵. Moreover, aromatic-aromatic interactions are believed to play an important role in stabilizing protein tertiary structure^{16; 17}. Aromatic residues stabilize protein structure by lowering conformational entropy inasmuch as their side chains have dramatically fewer degrees of freedom relative to the number of hydrophobic interactions they can establish. They occur more frequently in thermophilic proteins than in mesophilic homologs¹⁸, likewise suggesting that aromatic clusters can counteract structurally instability (due, in that case, to high temperature). The swII-CT aromatic residues often interact with other aromatic residues that are highly conserved within specific Ras-like families; these include, for example, a phenylalanine in Rab GTPases that interacts with the swII-CT tryptophan (not shown) and a tryptophan specifically conserved within Arf, Arl and G α GTPases¹¹. Aromatic tyrosine residues, which are favored within the swII-CT [YF]-[YF] pattern, can further contribute to structural diversity and stability¹⁹ by forming internal hydrogen bonds involving their side-chain -OH groups, which can serve as both hydrogen donors and acceptors. Thus

the swII-CT aromatic residues may facilitate alternative switch II conformational states, while also counteracting the destabilizing effects of conformational plasticity.

The charge-dipole pocket and an outward-directed switch II α -helix

One configuration (Fig. 3a) can explain why the swII-CT residues are conserved inasmuch as, within this configuration, they form a pocket around the negative-dipole moment of a switch II α helix with the positively-charged residue of the [RK]-x-[ILV] pattern inserted into this pocket (Fig. 3a inset). The atomic interactions forming this charge-dipole pocket include: a CH- π interaction between the aromatic residues of the [YF]-[YF] pattern²⁰; a cation- π interaction^{21; 22} between the positively-charged residue and the tryptophan (or, at times, a phenylalanine) that directly precedes the Walker B aspartate, and—when the positively-charged residue is a lysine—a CH- π interaction²⁰ between these residues as well; van der Waals interactions between the aliphatic residue of the [RK]-x-[ILV] pattern and the other swII-CT residues; and—whenever the first residue of the [YF]-[YF] pattern is a tyrosine, which is typically the case—a hydrogen bond between its side-chain -OH group and a specific backbone oxygen.

An association between the charge-dipole pocket and switch II restructuring is suggested by comparisons between typical monomeric forms of Rab family GTPases and an unusual, homodimeric form of GDP-bound Rab27. In the homodimeric form, which is observed in four distinct crystal structures²³ (pdb_ids: 2f7s, 2.70 Å; 2iey, 3.18 Å; 2iez, 2.80 Å; 2if0, 2.80 Å), the inter-switch region (which connects switch I to switch II) is exchanged between subunits so that, for each subunit, this region is donated by the other subunit (Fig. 3b). As a result, the switch II region forms a long α helix that is directed away from the structural core of one subunit and toward the structural core of the other subunit. Presumably this helix lacks the conformational strain typically imposed on monomeric GTPases—the switch II regions of which need to bend around and reconnect to the structural core. Moreover, for each Rab27 homodimeric subunit the swII-CT residues form the charge-dipole pocket configuration with nearly ideal geometry. Hence, presumably the charge-dipole pocket is structurally compatible with formation of this unusual, outward-directed switch II helix—a structural theme that occurs repeatedly in the following analysis of other Ras-like GTPases. Incidentally, effector-bound forms of (monomeric) Rab27-GTP^{24; 25} (not shown) lack both the charge-dipole pocket configuration and the outward-directed switch II helix, thereby demonstrating that Rab27 is able to form alternative swII-CT configurations.

Monomeric Ras-like GTPases that form the charge-dipole pocket (Fig. 4a,d,g and Fig. S2a in Supplementary Data) or nearly so (Fig. 4b,e and Fig. S2b in Supplementary Data) consistently possess a switch II helix that traces out a path close to that of the Rab27 homodimeric helix (see structural superpositions in Fig. 4 and in Fig. S2 of Supplementary Data). However, in these monomeric forms this helix is truncated and tends to deviate slightly relative to the homodimeric form, presumably because the switch II backbone needs to bend around and reconnect to the structural core.

In contrast, monomeric Ras-like GTPases that lack the charge-dipole pocket configuration typically possess a switch II helix that more directly bridges the sites of attachment with the core and that thus traces out a path distinct from that of the Rab27 homodimeric helix (Fig. 4c,f,i and Fig. S3a,c,e in Supplementary Data). Less often, Ras-like GTPases exhibit atypical swII-CT alternative configurations, the switch II helix of which likewise follows a path distinct from that of the homodimer (Fig. S3b,d,f in Supplementary Data). Structures of various non-Ras-like GTPases (Fig. S4 in Supplementary Data) also lack the outward-facing switch II helix. Taken together, these observations indicate an association between the charge-dipole pocket and the outward-oriented α -helix that, based on a ‘back-of-the-envelope’ calculation¹, is highly statistically significant.

Possible role for the charge-dipole pocket

An analysis of available crystal structures reveals that, although the charge-dipole pocket configuration is most often associated with the GDP-bound state, it also is observed in the GTP-bound state and may be present or absent in both states. For example, for Rab11a the charge-dipole pocket occurs in both states (Fig. 4d,e,g), whereas for Rab5a an alternative configuration occurs in both states (Fig. 4c,f). Moreover, within the unit cell of a single crystal of GDP-bound Rab5a, both the charge-dipole pocket and an alternative configuration occur²⁶ (Fig. 4a,c). In addition, for GDP-bound Rab11a and Rab23 relatively minor deviations from the charge-dipole pocket configuration are associated with shortening (Fig. 4g,e) or lengthening (Fig. S2a,b in Supplementary Data)(respectively) of the outward-directed switch II helix—suggesting that this helix can extend and retract. Together, these and the previous observations suggest that the role of the charge-dipole pocket is to promote formation of the outward-directed helix, which, in turn, is associated with restructuring of the switch II N-terminal region and, as a result, with repositioning of co-conserved residues implicated in GTP hydrolysis⁷ and nucleotide exchange^{8; 9; 10} (see Fig. S5 in Supplementary Data). These residues correspond to the pattern [AG]-x-Q-[DE] in Fig. 2.

Possible tests of the proposed role of the charge-dipole pocket

The role of the charge-dipole pocket may be explored experimentally by determining: (i) whether preventing formation of the charge-dipole pocket abolishes or significantly inhibits GTP hydrolysis or nucleotide exchange and (ii) whether the outward facing switch II helix can form in the absence of the charge-dipole pocket. In particular, I propose two experiments involving unnatural amino acid mutagenesis²⁷ of human Rab5a and of human Rab11a in conjunction with NMR spectroscopic²⁸ monitoring of ¹³C and ¹⁵N labeled backbone conformations. Such analyses are facilitated by the numerous Ras-like GTPase structures that have been previously determined through both NMR spectroscopy and X-ray crystallography.

For both of these experiments, the first tyrosine residue of the [YF]-[YF] motif (Fig. 2) (Tyr89 and Tyr81 within Rab5a and Rab11a, respectively) would be mutated to o-nitrobenzyl-tyrosine (oNBtyr) using an *in vivo* system for unnatural amino acid mutagenesis (either in the yeast *Saccharomyces cerevisiae*²⁹ or in mammalian cells³⁰). Within Rab structures exhibiting the charge-dipole pocket configuration, the side chain OH group of this tyrosine forms a hydrogen bond with the backbone of the β -strand directly preceding the P-loop (Figs 3a and 4a,d,e,g). However, within Rab structures exhibiting the typical alternative swII-CT configuration (Figs 4c,f), the tyrosine side chain is flipped in the other direction and, as a result, the OH group is solvent exposed. Thus, mutation of this tyrosine to oNBtyr in Rab11a or in Rab5a is predicted to lock the swII-CT residues into this alternative configuration because the nitrobenzyl group—which is covalently bound to the tyrosine side chain oxygen within oNBtyr—prohibits the typical tyrosine hydrogen bond from forming and, in any case, is too bulky to accommodate the charge-dipole pocket configuration. Formation of the alternative SwII-CT configuration in this way can be confirmed through NMR spectroscopy. A critical feature of the oNBtyr mutant residue is that it can be rapidly converted back to the native tyrosine through UV-irradiation,

¹Determining the statistical significance of the association between the charge-dipole pocket configuration and the outward facing switch II helix is complicated by the fact that various structural features associated with the swII-CT residues were examined during the analysis. This requires that the statistical significance be adjusted for multiple hypotheses—though it's not clear by how much. As a rough approximation, the following 'back-of-the-envelope' calculation was performed: Assume (conservatively) that about a million hypotheses were tested implicitly while examining about 120 distinct Ras-like GTPase structures: At least two dozen distinct structures exhibit the charge-dipole pocket and the outward-oriented switch II helix (or very nearly so); at least ninety distinct structures exhibit an alternative swII-CT configuration without the outward-oriented helix; and no more than a half dozen of these structures exhibit features intermediate between these two configurations. These intermediate forms can be categorized (conservatively) by viewing them as three charge-dipole pocket configurations without an outward-oriented helix and three alternative configurations with an outward-oriented helix. Based on these observations, Fisher's exact test rejects the hypothesis that there is no association between the charge-dipole pocket and the outward-oriented switch II helix with a *p*-value of $\sim 10^{-12}$ (after adjusting for a million hypotheses using a Bonferroni correction).

which cleaves the nitrobenzyl-photocage from oNB^{Tyr} ³¹. The feasibility of this approach is supported by recent studies of oNB^{Tyr} mutants within *Escherichia coli* ³².

The first experiment utilizes the Rab5a oNB^{Tyr}89 mutant. Since available crystal structures of both the GTP-bound and the GDP-bound states of Rab5a can exhibit the alternative swII-CT configuration (Figs 4c,f), this mutant can be used to test whether formation of the charge-dipole pocket is important for transitioning between states—as measured by monitoring GTP hydrolysis and nucleotide exchange. At the same time, NMR spectroscopy can be used to monitor whether formation of the outward-directed switch II helix can occur when the swII-CT residues are prohibited from forming the charge-dipole pocket. As a control, the oNB^{Tyr}81 mutation can be converted back to the native tyrosine via UV-irradiation; this will ensure that any unusual biochemical properties of the protein (including an inability to form the outward-directed switch II helix) is due solely to this mutation. This experiment serves as a stringent test of the proposed hypothesis: If transitions between states are not significantly inhibited in the mutant relative to the native protein, then this would rule out a functionally-significant restructuring role for the charge-dipole pocket in GTP hydrolysis and nucleotide exchange.

The second experiment utilizes the Rab11a oNB^{Tyr}81 mutation, and was inspired by the observation that all five currently available structures of Rab11a exhibit the charge-dipole pocket configuration and the outward-directed switch II helix (pdb identifiers and states: 1oiv, bound to GDP; 1oiw, bound to GTP γ S; 1oix, bound to GDP + Pi; 1yzk, bound to GppNHp; and 2f9l bound to GDP). In this experiment, NMR spectroscopy would be used to monitor the Rab11a oNB^{Tyr}81 mutant backbone configuration. If the outward-directed switch II helix forms in the (mutation-induced) alternative swII-CT configuration, then this would falsify the proposed hypothesis that formation of the charge-dipole pocket is required to form this unusual helix. Again, as a control, the UV-irradiation can be used to convert the mutant back to the native protein.

The swII-CT region outside of the Rab and Ran families

The typical alternative swII-CT configuration that is observed for Rab GTPases (Fig. 4c,f and Fig. S3a in Supplementary Data) is also observed for Arf (Fig. S3c in Supplementary Data), G α (Fig. 4i), and other Ras-like GTPases outside of the Rab family. However, for non-Rab family GTPases bound only to GDP or to GTP, the charge-dipole pocket configuration thus far is observed (albeit with suboptimal geometry) only for Ran (Fig. 4b). This is despite the fact that most Arf, Arl and Sar GTPases match the swII-CT patterns exactly. One explanation is that, for these GTPases, some of the swII-CT residues have assumed other functions that could influence formation of the charge-dipole pocket. For example, Arf, Arl and Sar possess an ‘inter-switch toggle’ device ³³ that is absent from other Ras-like GTPases and that, in the GDP-bound state, shifts the swII-CT tryptophan residue away from its normal position. Of course, such family-specific mechanisms are not inconsistent with formation of the charge-dipole pocket in certain conformational states that have not yet been crystallized.

To explore this possibility, Fig. 5a shows a hypothetical charge-dipole pocket configuration for Arl8 based on molecular modeling and, for comparison, Fig. 5b shows an empirically-based crystal structure of Arl8 bound to GDP. In the Arl8-GDP structure and in other crystal structures of Arf and Arl GTPases bound to GDP the residue corresponding to Trp77 of Arl8 interacts with the (inter-switch toggle-shifted) swII-CT tryptophan (Trp65 of Arl8) ¹¹. By contrast, in available structures of Arf, Arl and G α GTPases bound to GTP the residues corresponding to Trp77 and Arg74 in Arl8 typically interact with the N-terminal switch II region involved in sensing of the γ -phosphate of GTP ¹¹ (not shown). Thus, it is noteworthy that, for the hypothetical Arl8 charge-dipole pocket configuration in Fig. 5a, Arg74 and Trp77 are directed away from the switch II N-terminal region due to formation of the outward-facing switch II helix, which—by restructuring the switch II region—could disrupt these interactions.

Support for the notion that formation of the charge-dipole pocket may, in some cases, require an additional, interacting component is provided by crystal structure analysis of G α i (Fig. 4h), which like Arf and Arl GTPases, conserves switch II residues corresponding to Arg74 and Trp77 of Arl8¹¹. The heterotrimeric G protein α subunit, G α i, when bound to both GDP and the GDP-selective peptide KB752³⁴ forms a (suboptimal) charge-dipole pocket with a switch II helix that traces out a path close to that of the Rab27 homodimer (Fig. 4h). This occurs despite the fact that within G α i the first aromatic residue of the [YF]-[YF] pattern is replaced by a cysteine. Notably, the crystallographic data for the G α i-GDP-KB752 complex³⁴ is consistent with this cysteine and the adjacent aromatic residue establishing an SH- π interaction^{35,36} that is chemically analogous to the typical face-to-edge CH- π interaction between the adjacent aromatic residues (Fig. 3a inset).

Why then might this G α i-GDP-KB752 complex exhibit charge-dipole-pocket-like features? The KB752 peptide stimulates low-level exchange of GTP for GDP in G α i by mimicking the role of the heterotrimeric G protein $\beta\gamma$ complex in receptor-promoted nucleotide exchange^{34,37}. Thus the G α i-GDP-KB752 configuration appears to represent a transitional form required for G α i activation. If so, then perhaps the cysteine substitution ensures tight regulation of G α i activation by requiring the participation of G $\beta\gamma$ for charge-dipole-pocket-mediated switch II restructuring. Of course, this possibility, the formation of the charge-dipole pocket in Ras-like GTPases for which it has not yet been observed, and the extent to which it is generally associated with switch II restructuring remains to be determined.

Conclusion

This study has yielded the following observations: (i) The most distinguishing feature of Ras-like GTPases at the sequence level are five conserved residue positions (at three distinct sequence locations) that are noted for the first time here—along with five other conserved residue positions whose likely biological roles in Ras-like GTPases were noted previously. (ii) At the structural level, these newly identified residues come together to form either a ‘charge-dipole pocket’ configuration or a variety of alternative configurations. (iii) The specific atomic interactions forming the charge-dipole pocket configuration require the specific amino acids conserved at these positions. (iv) The charge-dipole pocket configuration occurs in both the GDP- and the GTP-bound states and both the charge-dipole pocket and alternative configurations can occur in the same state. (v) The formation of the charge-dipole pocket is highly correlated with formation within the switch II region of an outward-oriented α -helix. And (vi) the formation of this helix is associated with structural changes in the switch II N-terminal region, which senses the γ phosphate of GTP and which harbors three previously-noted Ras-like residues involved in GTP hydrolysis and nucleotide exchange. Taken together, these observations suggest that the charge-dipole pocket may play a critical role in the switching mechanism by facilitating restructuring of the switch II N-terminal region.

Materials and Methods

Iterative strategy for analysis of evolutionarily-divergent patterns

Fig. 1a outlines a strategy for iteratively enlarging and improving a multiple sequence alignment while at the same time categorizing the sequences based on evolutionarily-divergent patterns. This strategy relies on four procedures: (i) a Bayesian partitioning with pattern selection (BPPS) procedure⁶; (ii) a routine to aid recursive application of the BPPS procedure; (iii) a search procedure that uses “Multiply-Aligned Profiles for Global Alignment of Protein Sequences” (MAPGAPS); and (iv) an automated procedure for examining the structural locations of the patterns identified by the BPPS procedure. The first and last of these procedures are implemented in the CHAIN program¹, which is available at www.chain.umaryland.edu; programs implementing other procedures, which are first described here, are available at

<http://mapgaps.igs.umaryland.edu>. This strategy iteratively enlarges and improves the input alignment insofar as additional sequences are detected and aligned in Step 4 of each iteration and new subfamilies are identified and incorporated into the multiple profile alignment in Steps 1-3 of the next iteration (see Fig. 1a). In principle, an alignment can be progressively improved in this way until reaching the inherent limitations associated with the currently available data.

Within each subfamily, sequences typically are so closely-related that they can be accurately aligned using standard alignment procedures, such as MUSCLE³⁸. Across families, however, sequences can be impossible to align correctly using standard procedures. Consider, for example, a conserved arginine that forms a salt bridge with an acidic residue within Ras, Rho, Rab and Ran GTPases⁶: In Rho GTPases (relative to these other GTPases) a two-residue deletion directly precedes this arginine and a 15-residue insertion directly follows it, so that conventional alignment methods fail to align this arginine correctly. However, once the subfamily profiles are accurately aligned, the MAPGAPS program can easily align such residues (as described below). For this study, creation of a highly accurate multiple profile alignment was facilitated by over 400 GTPase structures in the pdb database and, as a result, the MAPGAPS program was able to obtain a very accurate alignment compared to automated methods. For example, the MAPGAPS program correctly aligns five weakly conserved sequence motifs within 32 distantly related GTPases of known structure, whereas MUSCLE misaligns 77 of these regions (unpublished).

A multiple profile alignment consists of a set of consensus sequences, one for each profile, each of which is aligned against a master consensus sequence representing the profile alignment itself. For the analysis here, accurate alignment of these consensus sequences was based on the locations of evolutionarily-conserved and evolutionarily-divergent patterns in over 400 P-loop GTPase structures (using the automated structural analysis procedure mentioned above) and on alignment of subtle, yet statistically significant similarities within large numbers of GTPase sequences using publically-available Bayesian sequence analysis methods^{39; 40; 41; 42; 43}.

The iterative strategy was applied to sequences in the NCBI nr, env_nr and translated EST databases⁴⁴. After convergence, the P-loop GTPase alignment contained 278,101 sequences (193,785 unique sequences). After removing closely-related and fragmented sequences, the resulting alignment contained 91,406 P-loop GTPase sequences. The BPPS procedure was applied to this final alignment to obtain the 'contrast alignment' shown in Fig. 2.

The BPPS procedure

The BPPS procedure starts by arbitrarily partitioning a multiple sequence alignment (given as input) into two sets (termed the foreground and the background) and by arbitrarily choosing a sequence pattern associated with the foreground set. For example, it might start by randomly selecting half of the sequences for the foreground set and the other half for the background set and by randomly selecting the residue patterns [YF], [ILV] and [ST] at column positions 20, 37 and 68, respectively. Next a Markov chain Monte Carlo strategy⁴⁵, called Gibbs sampling, iteratively reassigns sequences to one of the partitioned sets and adds or removes pattern residues at alignment positions⁶. Such changes are sampled with probability proportional to the degree to which they improve the underlying statistical model, which measures how well the pattern distinguishes the foreground from the background sequences. This process is continued until convergence on a (local or global) optimum or nearly-optimum partition-pattern pair.

The BPPS procedure is specifically designed to find local optima because otherwise it would miss biologically important divergent patterns. Take Ran GTPases, for example. Ran conserves certain residues unique to the Ran family, other residues unique to the Ran, Rab, Rho and Ras families, and others shared by all Ras-like GTPases. Based on the posterior probability

distribution two of these three categories of conserved residues will show up as local optima whereas a third category will show up as a global optimum. Thus it is often desirable to find local optima as well as the global optimum. Simulated annealing⁴⁶ is applied to drop into a (typically local) optimum, which corresponds to two groups of proteins that have strikingly diverged from each other at pattern positions. For mathematical details see⁶. For input alignments composed of multiple subgroups, the BPPS procedure can identify these multiple subcategories by using an option where multiple runs are initialized with distinct seed patterns, as previously described^{1; 6}.

The BPPS procedure returns as output a contrast alignment, such as that shown in Fig. 2. Of course, inherent uncertainties are associated with such a contrast alignment due to: (i) the specific sequences included in the input alignment; (ii) regions of alignment uncertainty; (iii) statistical uncertainty regarding the pattern-partition pair; and (iv) the choice of parameters specifying the degree of pattern conservation and the expected number of pattern positions. Given these uncertainties, it is impossible to guarantee that the BPPS procedure will find the biologically most meaningful pattern-partition pair. Nevertheless, the robustness of a pattern-partition pair can be established by running the BPPS procedure multiple times using distinct random seeds, various parameter settings, and various input sequence sets. Fig. S6 of Supplemental Data illustrates how such inherent variability can be explored by generating three contrast alignments, each of which utilizes a different random seed and a different (50%) subset of the aligned sequences used for the analysis in Fig. 2: Although there is some variability, in each case the predominant residue pattern obtained is very similar and does not alter the main conclusions of the analysis. The issue of alignment uncertainty is addressed in the next section.

The MAPGAPS procedure

The MAPGAPS procedure takes as input a multiple alignment of hidden Markov Model (HMM) profiles^{47; 48}, searches a protein sequence database for significant hits to at least one of these profiles, and outputs a global alignment of the matching sequences. The HMM profiles are generated from multiple alignments obtained in step 1 of Fig. 1a. The MAPGAPS procedure uses PSI-BLAST heuristics and statistics⁴⁹ to speed up the search and to ensure appropriate sensitivity and selectivity. The search time depends on the number of profiles, their lengths, and the size of the protein database being search, but typically runs in a couple hours on half a dozen grid nodes when searching the NCBI nr database against about 150 profiles—roughly the number used for the GTPase analysis performed here. If desired, a MAPGAPS program option can be used to eliminate certain protein families from the output alignment when the goal is to analyze functionally divergent residues that distinguish one specific subgroup from another specific subgroup within a larger protein class.

After a search, the MAPGAPS program globally aligns all database sequences with a significant match to at least one profile using both the pair-wise alignment of each database sequence against its highest scoring profile and that profile's alignment against the other profiles—as specified by the multiple profile alignment. More specifically, if position x in a matching sequence aligns with position y in profile Y and if position y in profile Y aligns with position z in the multiple profile alignment, then position x in the sequence is aligned with position z in the global alignment. In this way, as long as the multiple profile alignment is correct, evolutionarily and functionally related residues that would otherwise be impossible to align using standard procedures, can be aligned correctly and automatically. The MAPGAPS program and its application will be described in detail elsewhere.

Structural analysis

The Reduce program⁵⁰ was used to predict hydrogen atom positions within structural coordinate files; this step is required for automated detection of hydrogen bonds by the CHAIN

program¹. Molecular images were created using Rasmol⁵¹. To eliminate extraneous influences on GTPase conformations, all of the structures shown in the figures, except for the G α i structure in (Fig. 4h), are bound either only to GDP or only to GTP or a GTP analog. The structural alignments between monomeric GTPases and the Rab27 homodimer in Fig. 4 and in Supplementary Data were obtained by minimizing the least square distances between the monomeric and homodimeric α -carbons of residues corresponding to the swII-CT patterns [RK]-x-[ILV] and [WF]. Molecular modeling of the charge-dipole pocket configuration for Arl8 GTPase was performed by applying the GROMACS molecular dynamics program⁵² (for energy minimization with hydrogen bond constraints) and the KiNG program (<http://kinemage.biochem.duke.edu/software/king.php>) (for obtaining favorable side-chain and backbone conformations and for resolving steric clashes) starting with the atomic coordinates for 1YZG (structural genomics consortium).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I thank L. Aravind, Leo Chavas, Tony Hunter, Kentaro Ihara, Vishvanath Nene, Ji-Joon Song and David J. Weber for comments and suggestions. This work was funded by the NIH Division of General Medicine Grant GM078541.

References

1. Neuwald AF. The CHAIN program: forging evolutionary links to underlying mechanisms. *Trends Biochem Sciences* 2007;32:487–493.
2. Wittinghofer, A. The Functioning of Molecular Switches in Three Dimensions. In: Hall, A., editor. *GTPases*. Oxford University Press; Oxford: 2000. p. 244–310.
3. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* 1982;1:945–51. [PubMed: 6329717]
4. Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* 1990;15:430–4. [PubMed: 2126155]
5. Leippe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* 2002;317:41–72. [PubMed: 11916378]
6. Neuwald AF, Kannan N, Poleksic A, Hata N, Liu JS. Ran's C-terminal, basic patch and nucleotide exchange mechanisms in light of a canonical structure for Rab, Rho, Ras and Ran GTPases. *Genome Res* 2003;13:673–692. [PubMed: 12671004]
7. Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. *Science* 2001;294:1299–304. [PubMed: 11701921]
8. Thomas C, Fricke I, Scrima A, Berken A, Wittinghofer A. Structural evidence for a common intermediate in small G protein-GEF reactions. *Mol Cell* 2007;25:141–9. [PubMed: 17218277]
9. Gasper R, Thomas C, Ahmadian MR, Wittinghofer A. The role of the conserved switch II glutamate in guanine nucleotide exchange factor-mediated nucleotide exchange of GTP-binding proteins. *J Mol Biol* 2008;379:51–63. [PubMed: 18440551]
10. Margarit SM, Sondermann H, Hall BE, Nagar B, Hoelz A, Pirruccello M, Bar-Sagi D, Kuriyan J. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* 2003;112:685–95. [PubMed: 12628188]
11. Neuwald AF. $G\alpha$ - $G\beta\gamma$ dissociation may be due to retraction of a buried lysine and disruption of an aromatic cluster by a GTP-sensing Arg-Trp pair. *Protein Science* 2007;16:2570–2577. [PubMed: 17962409]
12. Corbeel L, Freson K. Rab proteins and Rab-associated proteins: major actors in the mechanism of protein-trafficking disorders. *Eur J Pediatr* 2008;167:723–9. [PubMed: 18463892]

13. Claessens CG, Stoddart JF. pi-pi interactions in self-assembly. *J Physical Organic Chem* 1998;10:254–272.
14. McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. *J Biol Chem* 1998;273:15458–63. [PubMed: 9624131]
15. Sun S, Bernstein ER. Aromatic van der Waals clusters: structure and nonrigidity. *J Phys Chem* 1996;100:13348–13366.
16. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 1985;229:23–8. [PubMed: 3892686]
17. Serrano L, Bycroft M, Fersht AR. Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J Mol Biol* 1991;218:465–75. [PubMed: 2010920]
18. Kannan N, Vishveshwara S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng* 2000;13:753–61. [PubMed: 11161106]
19. Pace CN, Horn G, Hebert EJ, Bechert J, Shaw K, Urbanikova L, Scholtz JM, Sevcik J. Tyrosine hydrogen bonds make a large contribution to protein stability. *J Mol Biol* 2001;312:393–404. [PubMed: 11554795]
20. Weiss MS, Brandl M, Suhnel J, Pal D, Hilgenfeld R. More hydrogen bonds for the (structural) biologist. *Trends Biochem Sci* 2001;26:521–3. [PubMed: 11551776]
21. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* 1999;96:9459–64. [PubMed: 10449714]
22. Anderson MA, Ogbay B, Arimoto R, Sha W, Kisselev OG, Cistola DP, Marshall GR. Relative strength of cation-pi vs salt-bridge interactions: the G α (340-350) peptide/rhodopsin system. *J Am Chem Soc* 2006;128:7531–41. [PubMed: 16756308]
23. Chavas LM, Torii S, Kamikubo H, Kawasaki M, Ihara K, Kato R, Kataoka M, Izumi T, Wakatsuki S. Structure of the small GTPase Rab27b shows an unexpected swapped dimer. *Acta Crystallogr D Biol Crystallogr* 2007;63:769–79. [PubMed: 17582168]
24. Kukimoto-Niino M, Sakamoto A, Kanno E, Hanawa-Suetsugu K, Terada T, Shirouzu M, Fukuda M, Yokoyama S. Structural basis for the exclusive specificity of Slac2-a/melanophilin for the Rab27 GTPases. *Structure* 2008;16:1478–90. [PubMed: 18940604]
25. Chavas LM, Ihara K, Kawasaki M, Torii S, Uejima T, Kato R, Izumi T, Wakatsuki S. Elucidation of Rab27 recruitment by its effectors: structure of Rab27a bound to Exophilin4/Slp2-a. *Structure* 2008;16:1468–77. [PubMed: 18940603]
26. Zhu G, Zhai P, Liu J, Terzyan S, Li G, Zhang XC. Structural basis of Rab5-Rabaptin5 interaction in endocytosis. *Nat Struct Mol Biol* 2004;11:975–83. [PubMed: 15378032]
27. Xie J, Schultz PG. A chemical toolkit for proteins—an expanded genetic code. *Nat Rev Mol Cell Biol* 2006;7:775–82. [PubMed: 16926858]
28. Markwick PR, Malliavin T, Nilges M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol* 2008;4:e1000168. [PubMed: 18818721]
29. Chen S, Schultz PG, Brock A. An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J Mol Biol* 2007;371:112–22. [PubMed: 17560600]
30. Liu W, Brock A, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* 2007;4:239–44. [PubMed: 17322890]
31. Deiters A, Groff D, Ryu Y, Xie J, Schultz PG. A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* 2006;45:2728–31. [PubMed: 16548032]
32. Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B, Spraggon G, Brock A, Ryu Y, Uno T, Schultz PG, Geierstanger BH. In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* 2008;130:9268–81. [PubMed: 18576636]
33. Pasqualato S, Renault L, Cherfils J. Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for ‘front-back’ communication. *EMBO Rep* 2002;3:1035–41. [PubMed: 12429613]
34. Johnston CA, Willard FS, Jezyk MR, Fredericks Z, Bodor ET, Jones MB, Blaesiuss R, Watts VJ, Harden TK, Sondek J, Ramer JK, Siderovski DP. Structure of G α (i1) bound to a GDP-selective

- peptide provides insight into guanine nucleotide exchange. *Structure* 2005;13:1069–80. [PubMed: 16004878]
35. Duan GL, Smith VH, Weaver DF. Characterization of aromatic-thiol pi-type hydrogen bonding and phenylalanine-cysteine side chain interactions through ab initio calculations and protein database analyses. *Molecular Physics* 2001;99:1689–1699.
 36. Reid KSC, Lindley PF, Thornton JM. Sulphur-aromatic interactions in proteins. *FEBS letters* 1985;190:209–213.
 37. Johnston CA, Siderovski DP. Structural basis for nucleotide exchange on G alpha i subunits and receptor coupling specificity. *Proc Natl Acad Sci U S A* 2007;104:2001–6. [PubMed: 17264214]
 38. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113. [PubMed: 15318951]
 39. Liu JS, Neuwald AF, Lawrence CE. Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *J Am Stat Assoc* 1995;90:1156–1170.
 40. Liu JS, Neuwald AF, Lawrence CE. Markovian structures in biological sequence alignments. *JASA* 1999;94:1–15.
 41. Neuwald AF, Liu JS. Gapped alignment of protein sequence motifs through Monte Carlo optimization of a hidden Markov model. *BMC Bioinformatics* 2004;5:157. [PubMed: 15504234]
 42. Neuwald AF, Liu JS, Lawrence CE. Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci* 1995;4:1618–32. [PubMed: 8520488]
 43. Neuwald AF, Liu JS, Lipman DJ, Lawrence CE. Extracting protein alignment models from the sequence database. *Nucleic Acids Research* 1997;25:1665–1677. [PubMed: 9108146]
 44. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008;36:D13–21. [PubMed: 18045790]
 45. Liu, JS. Monte Carlo Strategies in Scientific Computing Springer Series in Statistics. Springer-Verlag; New York: 2008.
 46. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* 1983;220:671–680. [PubMed: 17813860]
 47. Eddy SR. Multiple alignment using hidden Markov models. *Ismb* 1995;3:114–20. [PubMed: 7584426]
 48. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;14:755–63. [PubMed: 9918945]
 49. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402. [PubMed: 9254694]
 50. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 1999;285:1735–47. [PubMed: 9917408]
 51. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 1995;20:374. [PubMed: 7482707]
 52. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26:1701–18. [PubMed: 16211538]
 53. Stroupe C, Brunger AT. Crystal structures of a Rab protein in its inactive and active conformations. *J Mol Biol* 2000;304:585–98. [PubMed: 11099382]
 54. Pasqualato S, Senic-Matuglia F, Renault L, Goud B, Salamero J, Cherfils J. The structural GDP/GTP cycle of Rab11 reveals a novel interface involved in the dynamics of recycling endosomes. *J Biol Chem* 2004;279:11480–8. [PubMed: 14699104]
 55. Stewart M, Kent HM, McCoy AJ. The structure of the Q69L mutant of GDP-Ran shows a major conformational change in the switch II loop that accounts for its failure to bind nuclear transport factor 2 (NTF2). *J Mol Biol* 1998;284:1517–27. [PubMed: 9878368]

56. Scapin SM, Carneiro FR, Alves AC, Medrano FJ, Guimaraes BG, Zanchin NI. The crystal structure of the small GTPase Rab11b reveals critical differences relative to the Rab11a isoform. *J Struct Biol* 2006;154:260–8. [PubMed: 16545962]
57. Terzyan S, Zhu G, Li G, Zhang XC. Refinement of the structure of human Rab5a GTPase domain at 1.05 Å resolution. *Acta Crystallogr D Biol Crystallogr* 2004;60:54–60. [PubMed: 14684892]
58. Coleman DE, Sprang SR. Structure of Gα1.GppNHp, autoinhibition in a Gα protein-substrate complex. *J Biol Chem* 1999;274:16669–72. [PubMed: 10358003]

Fig. 1. Analysis and attributes of Ras-like GTPases. **(a)** Iterative strategy used for comprehensive analysis of GTPase evolutionary divergence. **Step 1:** Use Bayesian partitioning with pattern selection (BPPS) ^{1, 6} to optimally partition an initial sequence alignment into “foreground” and “background” sub-alignments such that the foreground strikingly conserves a pattern (set of blue vertical bars) that is non-conserved in (or that contrasts with) the background at pattern positions (white vertical bars). This output is termed a “contrast alignment”. When appropriate, more than one contrast alignment is returned. **Steps 1a,1b:** Apply the BPPS procedure recursively to previously-identified foreground and background alignments; in effect, this identifies every (reasonably-sized) evolutionarily-divergent subgroup. **Step 2:** For each

foreground alignment create a hidden Markov Model (HMM) profile⁴⁸. **Step 3:** Globally align the HMM profiles based on structural analysis of pattern residues and on Bayesian sequence analysis. **Step 4:** Perform a MAPGAPS search to identify and globally align database sequences with significant profile matches. Finally, iteratively repeat steps 1-4 to enlarge and improve the alignment. See Materials and Methods for details. **(b)** Relative sizes of major Ras-like GTPase families that were identified by the analysis in (a). **(c)** Relevant structural features of Ras-like GTPases. The γ phosphorus atom of GTP is shown in orange with covalently bound oxygen atoms in red. The structure of Sec4 is shown (pdb_id: 1g17)⁵³.

Fig. 2.

Contrast alignment identifying sequence patterns that most distinguish Ras-like GTPases from other P-loop GTPases. Representative foreground sequences are aligned with protein names and pdb identifiers in the leftmost column and with pattern residues highlighted as previously described⁶. The red bars above the alignment quantify the selective pressure that is imposed on the foreground relative to the background at pattern positions, as previously defined⁶. The large red and blue dots indicate pattern positions corresponding to the newly-identified swII-CT component and to residues with previously proposed functions, respectively. Characteristic foreground residues at each position are shown directly below the alignment, and, directly below these, corresponding residue frequencies are given in integer tenths; for example, a '7' indicates that the corresponding pattern residue occurs in 70%-80% of the foreground sequences. Below this, predominant residue types and frequencies for background sequences at each position are shown. Alignment positions correspond to the first sequence, Rab11a.

Fig. 3. The charge-dipole pocket. **(a)** GDP-bound Rab11a (pdb_id: 1oiv; resolution, 1.98 Å)⁵⁴. The inset indicates the characteristic atomic interactions of the charge-dipole pocket. **(b)** Homodimeric GDP-bound Rab27 (pdb_id: 2if0; resolution, 2.8 Å)²³. The inter-switch regions of the two subunits are exchanged and, as a result, each of the switch II regions forms an α helix that is directed away from the GTPase structural core. The backbone of the adjacent subunit's switch and inter-switch regions are shown in gray. General color scheme: backbone of the P-loop, switch I and adjacent regions, dark red; backbone of the switch II region and adjacent β strand, orange; side chains of the Walker A lysine and Walker B aspartate residues, magenta; side chains of residues forming the swII-CT component, orange; oxygen, nitrogen,

and hydrogen atoms involved in hydrogen bonds are red, blue and white, respectively. Hydrogen bonds are indicated by dotted lines and aromatic and electrostatic interactions by dot clouds.

Fig. 4.

Structural superposition of the switch II regions of various monomeric Ras-like GTPases with that of the Rab27 homodimer. Backbone traces of relevant regions for monomeric GTPases are colored as in Fig. 3a and for the two homodimeric subunits (pdb_id: 2if0; resolution, 2.80 Å)²³ in light and dark gray. Boxes of the same color around images denote identical (or nearly identical) proteins. (a) Rab5A bound to GDP (pdb_id, 1tu4A; resolution, 2.20 Å)²⁶. (b) Ran bound to GDP (1byu; 2.15 Å)⁵⁵. (c) Rab5A bound to GDP (1tu4B; 2.20 Å)²⁶. (d) Rab11 bound to a GTP analog (GTPγS) (1oiw; 2.05 Å)⁵⁴. (e) Rab11 bound to GDP (2f9l; 1.55 Å)⁵⁶. (f) Rab5A bound to a GTP analog (GNP) (1r2q; 1.05 Å)⁵⁷. (g) Rab11 bound to GDP (1oiv; 1.98 Å)⁵⁴. (h) G_{α_i} bound to GDP and a GDP-selective peptide (1y3a; 2.50 Å)³⁴. Within G_{α_i} the side-chain of a conserved cysteine residue that replaces the first residue of the [YF]-[YF] pattern is shown in cyan with the sulphur atom in yellow. (i) G_{α_i} bound to a GTP analog (GppNHp) (1cip; 1.50 Å)⁵⁸.

Fig. 5. Hypothetical and observed conformations of the Arl8 switch II region. See legend to Fig. 3 for descriptions. **(a)** Structural model of Arl8 switch II region in the charge-dipole pocket configuration. This model was obtained by restructuring the switch I and II regions of the Arl8-GDP structure in (b) based on the structure of Rab11 bound to GDP (Fig. 3a), and then subjecting the model to energy minimization and to manual adjustments using the GROMACS and KiNG programs, respectively (see Materials and Methods). **(b)** Atypical alternative swII-CT configuration observed in the crystal structure of Arl8 bound to GDP (pdb_id, 1yzg; resolution, 2.00 Å)(structural genomics consortium). In this structure the swII-CT tryptophan residue (Trp65) has shifted two positions (via an inter-switch toggle device³³) and, as a result, interacts with another switch II tryptophan residue (Trp77)(yellow sidechain) that is conserved within Arf, Arl and Gα GTPases.