

Integrating energy calculations with functional assays to decipher the specificity of G protein–RGS protein interactions

Mickey Kosloff¹, Amanda M Travis¹, Dustin E Bosch², David P Siderovski² & Vadim Y Arshavsky¹

The diverse Regulator of G protein Signaling (RGS) family sets the timing of G protein signaling. To understand how the structure of RGS proteins determines their common ability to inactivate G proteins and their selective G protein recognition, we combined structure-based energy calculations with biochemical measurements of RGS activity. We found a previously unidentified group of variable ‘Modulatory’ residues that reside at the periphery of the RGS domain–G protein interface and fine-tune G protein recognition. Mutations of Modulatory residues in high-activity RGS proteins impaired RGS function, whereas redesign of low-activity RGS proteins in critical Modulatory positions yielded complete gain of function. Therefore, RGS proteins combine a conserved core interface with peripheral Modulatory residues to selectively optimize G protein recognition and inactivation. Finally, we show that our approach can be extended to analyze interaction specificity across other large protein families.

RGS proteins have a critical role in many G protein–dependent signaling pathways. RGS proteins ‘turn off’ heterotrimeric ($\alpha\beta\gamma$) G proteins and thereby determine the duration of G protein–mediated signaling events^{1–5}. Like many signaling proteins, RGS proteins comprise a large and diverse family. In humans, about 20 ‘canonical’ RGS proteins downregulate activated G proteins of the G_i and G_q subfamilies^{6,7}. In these RGS proteins, the ~120-residue RGS homology domain functions as a GTPase-activating protein (GAP) for GTP-bound $G\alpha$ subunits^{3–5}. RGS proteins have been implicated in a wide range of pathologies, including cancer, hypertension, arrhythmias, drug abuse and schizophrenia^{7–10}, making them promising drug targets^{7,8}. Therefore, identifying the determinants of G protein recognition by RGS proteins is essential for understanding these signaling pathways and for eventually manipulating them with drugs.

Although multiple RGS proteins are often expressed in the same cell, only particular RGS proteins mediate a given biological function^{11–17}. This has generated considerable interest in understanding the interaction specificity of RGS proteins. In many cases this specificity may originate from precise subcellular targeting, contributions from additional noncatalytic domains, adaptor proteins or participation in scaffolded protein complexes^{7,9,13,15,18,19}. However, in some cases the ability to recognize a given G protein is defined by the RGS domain itself^{7,9,13,15}. Nevertheless, the only two well-studied examples of RGS domain specificity are RGS9, whose specific recognition of $G\alpha_q$ requires the adaptor protein PDE γ ^{18,20}, and RGS2, which specifically downregulates G proteins of the G_q , but not G_i , subfamilies^{21,22} (compare ref. 23). The key determinants of RGS2 specificity have been identified²² by analysis of the multiple sequence alignment of RGS proteins

in the context of the RGS4– $G\alpha_{i1}$ crystal structure²⁴. This alignment shows three crucial positions that are highly conserved in the RGS family, but are different in RGS2. Changing these three RGS2 residues to their counterparts in RGS4 yields a gain-of-function phenotype that enables RGS2 to efficiently downregulate $G\alpha_i$ ^{22,25}. Additional studies showed that the GAP activity of individual RGS proteins toward a given $G\alpha$ may vary (reviewed in refs. 6–8,13), but the molecular determinants for this selectivity have not been identified.

Critical insights into the GAP activity of RGS proteins have been obtained using X-ray crystallography. So far, eight different structures of $G\alpha$ subunits in complex with canonical RGS domains have been solved^{24–28}. These studies, combined with biochemical examinations, have established that high-activity RGS domains bind $G\alpha$ subunits and stabilize their catalytic residues allosterically in a conformation optimal for GTP hydrolysis^{6,24,29–31}. RGS protein residues in the vicinity of the $G\alpha$ –RGS domain interface show substantial diversity, suggesting that they may set interaction specificity. However, low sequence identity among RGS domains (as low as 30%; **Supplementary Table 1**) makes it difficult to pinpoint RGS domain residues that determine selective interaction with a specific $G\alpha$ subunit^{27,32}.

In this study, we combined functional assays with structure-based computations to determine the structural features within a large array of human RGS proteins that control their ability to inactivate a representative G protein, $G\alpha_o$ (also known as GNAO1). We combined the experimental benchmark of the ability of ten RGS domains to activate $G\alpha_o$ GTPase with comparative structural analysis, electrostatic calculations of interaction energies using the finite-difference Poisson–Boltzmann (FDPB) method and *in silico* mutagenesis. Using a

¹Duke Eye Center, Duke University Medical Center, Durham, North Carolina, USA. ²Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. Correspondence should be addressed to V.Y.A. (vadim.arshavsky@duke.edu).

Received 5 November 2010; accepted 7 April 2011; published online 19 June 2011; doi:10.1038/nsmb.2068



Figure 1 The GAP activities of ten representative RGS domains are not correlated with their subfamily classification. **(a)** The k_{gap} constant for each domain was calculated as described in **Supplementary Methods** from single-exponential fits to the time course of GTP hydrolyzed by $G\alpha_0$ (400 nM) with or without added RGS protein (20 nM). Data are mean \pm s.e.m.; $n \geq 4$. High-activity, low-activity and no-activity RGS proteins are green, magenta and red (RGS2 only), respectively. **(b)** Phylogenetic tree of 19 human RGS domains. RGS proteins whose activity was tested in this study are colored as in **a**.

consensus approach across the eight available RGS domain–G protein crystal structures, we developed a structure-to-sequence map predicting which residues within the RGS domains are essential for GAP function and which residues can modulate specific interactions with the cognate $G\alpha$ subunit. We validated these predictions by site-specific mutagenesis of critical residues in this map that allowed us to impair GAP function in high-activity RGS proteins and completely restore this function in low-activity RGS proteins. Finally, we explored the general utility of this approach by applying it to the interaction between the *Escherichia coli* colicin E7 and its inhibitory immunity proteins, a well-established system for studying protein–protein interaction specificity. Our computational analysis pinpointed not only specificity determinants found in earlier computational studies of these proteins, but also some previously identified only by *in vitro* evolution. Therefore, our approach extends the analysis of interaction specificity to whole families and complements existing protein design methodologies.

RESULTS

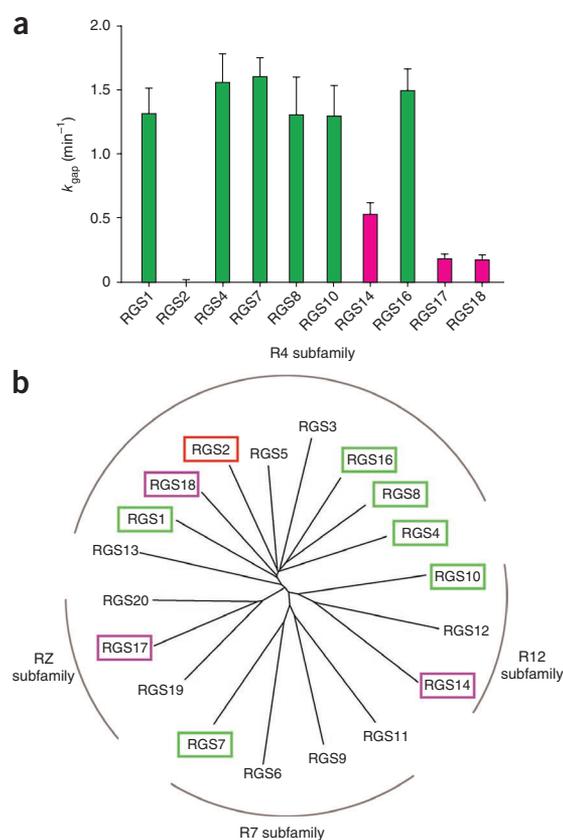
RGS proteins differ in their ability to activate $G\alpha_0$ GTPase

We measured the GAP activity of ten individual human RGS domains using single-turnover GTPase assays with the G protein $G\alpha_0$ (**Fig. 1a**). Six of these domains (RGS1, RGS4, RGS7, RGS8, RGS10 and RGS16) had the same high GAP activity, one (RGS2) had no measurable activity (as we expected from refs. 21,22,25) and three (RGS14, RGS17 and RGS18) had low but discernible activities. Notably, there was no correlation between the GAP activities of individual RGS domains and the degree of their sequence identity. Indeed, the sequence identity among the six highly active RGS domains typically ranged 37–60%, with only one pair sharing 73% identity (**Supplementary Table 1**). This is the same range as the identity among the sequences of no-activity (RGS2) and low-activity (RGS14, RGS17 and RGS18) RGS domains (37–56%), or between the sequences of the no- or low-activity and high-activity groups of RGS domains (36–60%). Therefore, sequence identity among RGS domains does not reliably predict RGS protein GAP activity on $G\alpha_0$.

Consequently, the GAP activity of these ten RGS domains was not correlated with their sequence alignment–based classification into subfamilies (**Fig. 1b**; we reached the same subfamily classification on the basis of the identity of additional noncatalytic domains in the corresponding full-length RGS proteins^{6,7,33}). We observed large differences in GAP activity within the same subfamily (for example, among RGS4, RGS18 and RGS2), but similar activities in RGS domains representing different subfamilies (for example, among RGS4, RGS7 and RGS10). This analysis demonstrates that RGS protein GAP function is determined at a finer resolution (that is, the individual-residue level) than provided by current RGS protein classifications.

Residue-level energy analysis of RGS–G protein interactions

To map the contributions of individual RGS domain residues to their GAP activity, we characterized the eight available crystal structures of canonical RGS domains bound to $G\alpha$ subunits, using a comparative



structural and energetic analysis (**Fig. 2a,b**). There are many RGS protein residues in the vicinity of the RGS domain– $G\alpha$ interface (for example, the eight crystal structures contain 62–67 RGS domain residues within 10 Å of the $G\alpha$ subunit), and the sequence diversity among these residues is considerable²⁷. Therefore, building upon a previously described approach³⁴, we coupled the FDPB method with *in silico* mutagenesis to calculate which RGS protein residues make substantial electrostatic contributions ($\Delta\Delta G_{\text{elec}}$) to the interaction with the cognate $G\alpha$ partner. In these calculations, we considered all residues within 15 Å of the RGS domain– $G\alpha$ interface (89–93 residues per RGS domain). We separated the electrostatic contributions of each residue into those coming from the side chain and/or those originating from the main chain (**Supplementary Fig. 1**; see Online Methods for details). We also estimated the nonpolar energetic contributions of each residue by converting surface area buried in the complex to the equivalent energy contribution³⁴. Because these energetic contributions were calculated in a static snapshot of a complex, we did not expect the obtained per-residue $\Delta\Delta G$ values to exactly match experimentally determined $\Delta\Delta G$ values (see refs. 34,35 for a detailed discussion). Rather, we aimed to generate a list of residues probably important for interactions with a $G\alpha$ partner. Therefore, we constructed a residue-level sequence ‘map’ that listed all RGS protein residues predicted to contribute substantially (by ≥ 1 kcal mol⁻¹) to the interaction (see Online Methods). We classified these residues into two major groups. The first group, ‘Significant & Conserved’ residues, make the same type of substantial energy contribution in the majority of structures (red asterisks, **Fig. 2a**). If the energy contribution comes only from the residue backbone, amino acids in Significant & Conserved positions may not be conserved at the sequence level (for example, position 131). The second group, putative ‘Modulatory’ residues, make substantial energy contributions only in some of the structures and are not conserved across the

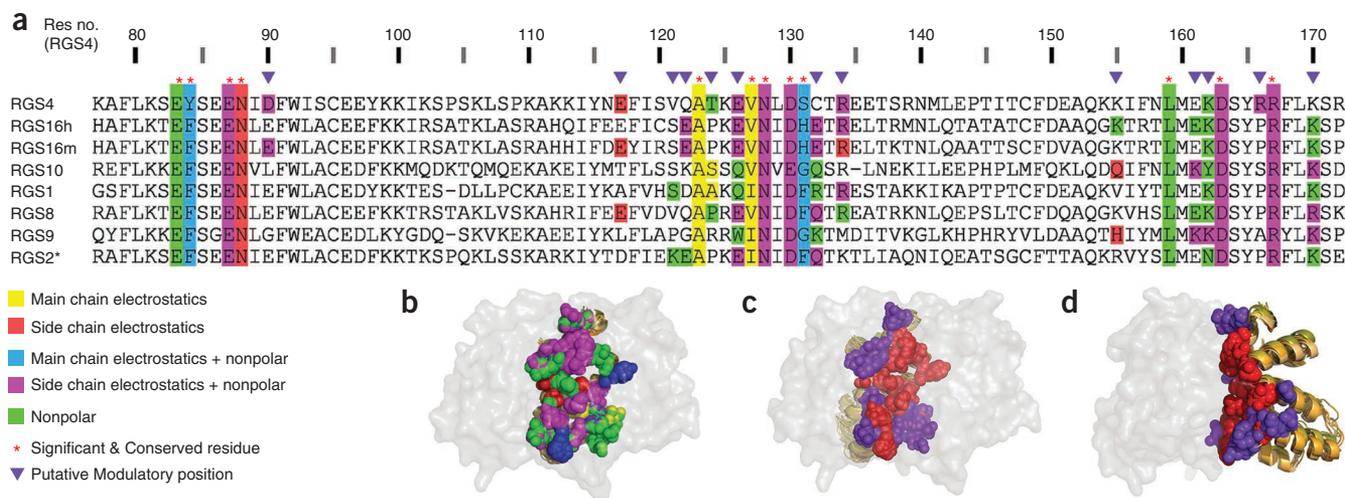


Figure 2 Positions of Significant & Conserved and Modulatory residues in multiple RGS proteins. **(a)** Residue-level sequence map summarizing structure analysis and energy calculations of eight RGS- α crystal structures with PDB codes 1AGR (RGS4); 2IK8 (RGS16h, human RGS16); 3C7K (RGS16m, mouse RGS16); 2IHB (RGS10); 2GTP (RGS1); 2ODE (RGS8); 1FQJ (RGS9); 2V4Z (RGS2*, gain-of-function RGS2 triple mutant; see Online Methods). The sequences in the multiple sequence alignment are taken from the crystal structures. RGS protein residues that contribute substantially to the interaction with the α subunit are color-coded in the panel according to the type of their energetic contribution (see key). Putative Significant & Conserved and Modulatory positions are marked above the alignment by red asterisks and purple triangles, respectively. **(b)** 3D visualization of the different types of energetic contributions by individual RGS protein residues (spheres, colored as in **a**). The eight superimposed RGS domain structures are viewed through the semitransparent surface of α . **(c)** Significant & Conserved and Modulatory residues in the eight superimposed RGS domain structures are red and purple spheres, respectively. Orientation is the same as in **b**. **(d)** 3D visualization as in **c**, rotated 90° about the y axis.

structures (purple triangles, **Fig. 2a**). We identified 12 RGS domain residues as Significant & Conserved and between 6 and 8 residues in each structure as Modulatory.

Notably, Significant & Conserved residues are located mainly in the center of the RGS domain- α interface, whereas putative Modulatory residues are located mostly at the periphery of this interface (**Fig. 2c,d**). This arrangement suggests that Significant & Conserved residues could be essential for RGS protein GAP activity, whereas different combinations of Modulatory residues may further fine-tune RGS domain- α interactions, ultimately defining whether a given RGS protein is a good or a poor GAP; we tested this hypothesis in this study.

Comparison of predictions with previous mutagenesis studies

To evaluate whether a substantial energetic contribution of an RGS protein residue (**Fig. 2a**) reliably predicts its importance in RGS GAP function, we first used published mutagenesis studies. In a comprehensive mutagenesis study of 39 RGS4 residues, analyzed using GTPase assays and/or the inhibition of G protein signaling in yeast, 23 mutants did not affect RGS4 function³⁶. Consistent with these experiments, our calculations showed no substantial energetic contribution for 22 of these residues. The only exception was Lys162, which was predicted to make a conserved nonpolar energetic contribution in all RGS domain structures (**Fig. 2a**). The K162A mutation did not impair RGS4 activity in the earlier study³⁶, although it was tested only in the less-direct yeast assay.

Among the 16 positions substantially impairing RGS4 activity³⁶, 7 are not located near the RGS domain- α interface and instead are a part of the hydrophobic core of the RGS domain, which is conserved across all available crystal structures (**Supplementary Fig. 2**). Presumably, mutating these large hydrophobic residues to alanines impaired RGS4 GAP activity indirectly through improper folding of these mutants. All of the other nine activity-impairing mutations (three of which were also identified in refs. 24,29,37,38) corresponded

to positions we marked as Significant & Conserved, confirming the predictions of our computational analysis. The remaining three RGS4 residues we identified as Significant & Conserved (Ala124, Val127 and Ser131) have not been mutated in earlier studies. However, the energetic contributions of these residues originate from their backbones rather than their side chains and thus are not amenable to straightforward validation by mutagenesis of side chains. Therefore, earlier mutagenesis studies fully agree with the predictions of our computationally derived residue-level map.

Design of loss-of-function RGS4 and RGS16 mutants

Next, we tested whether the putative Modulatory residues listed in our map (**Fig. 2a**) have a role in RGS protein GAP activity. Almost none of these residues have been mutated in earlier studies, probably because the lack of conservation at these positions suggested that they have no functional role. For these mutagenesis experiments, we picked representative Modulatory positions in human RGS4 and RGS16 (**Fig. 3a,b**). Single alanine substitutions of Modulatory residues in RGS4 had either a minor or a moderate effect on GAP activity (**Fig. 3c**). However, the loss-of-function effect was additive: the GAP activity of a triple mutant (RGS4d) was abolished. Therefore, mutations in a sufficient number of Modulatory residues cause complete loss of function, comparable to the effect of a mutation in the Significant & Conserved residue Asn128 (RGS4e in **Fig. 3c**), which is critical for the functions of RGS4 (refs. 24,29) and RGS16 (refs. 37,38).

Similarly, mutating individual Modulatory residues in RGS16 had either no effect or a moderate effect on its GAP activity (**Fig. 3d**). But, as we observed in RGS4, the effect of double or triple mutants was additive and impaired the ability of RGS16 to activate G_{α_o} GTPase to a much higher degree than single mutations. These results underscore the importance of Modulatory residues in attaining the maximal GAP activity of RGS proteins, and thereby validate our approach for pinpointing critical residues using our structure-to-sequence map.

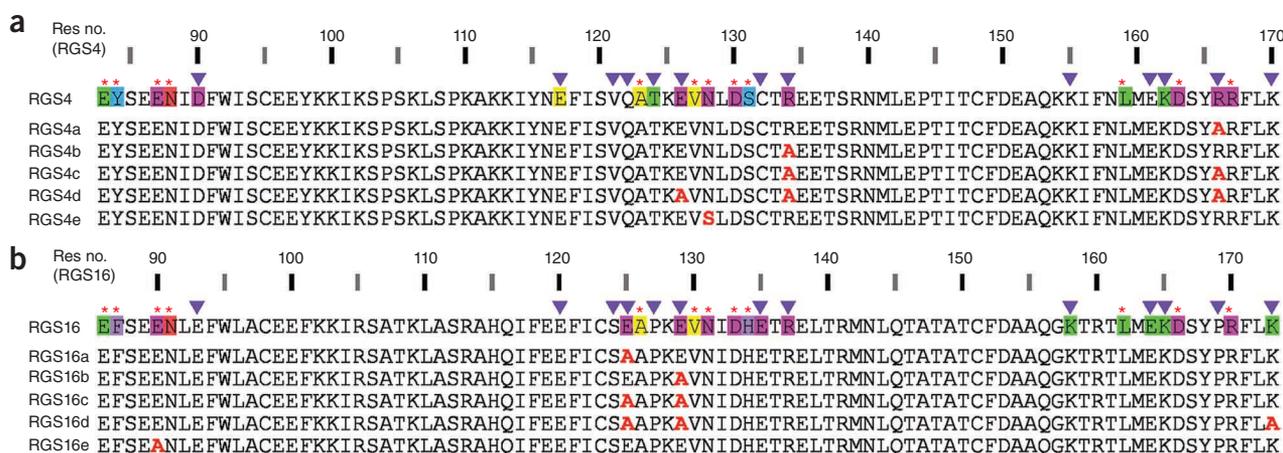
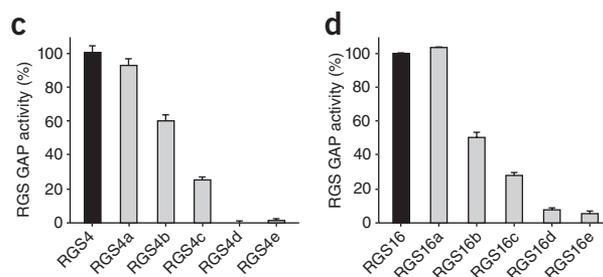


Figure 3 Mutations in Modulatory positions impair the GAP activities of RGS4 and RGS16 in an additive manner. (a) Sequences of RGS4 mutations in Modulatory positions (RGS4a–RGS4d) and a Significant & Conserved position (RGS4e). (b) Sequences of RGS16 mutations in Modulatory positions (RGS16a–RGS16d) and a Significant & Conserved position (RGS16e). The annotated sequences of wild-type RGS4 and RGS16 in a and b are from Figure 2a. (c) GAP activities of RGS4 mutants determined by single-turnover GTPase assays. GTP-loaded $G\alpha_o$ (400 nM) was incubated with or without RGS4 (40 nM) for 1 min. GAP activities are expressed as a percentage of wild-type RGS4 activity. Values are mean \pm s.e.m. ($n \geq 4$). (d) GAP activities of the RGS16 mutants, determined as in c. Experiments were conducted in triplicate.



Design of gain-of-function RGS17 and RGS18 mutants

We tested the utility of our energy-contribution map by taking low-GAP-activity RGS proteins and redesigning them into mutants with high GAP activity (Figs. 4 and 5). We selected two low-activity RGS proteins representing different subfamilies, RGS17 and RGS18. The high-activity template for redesign was RGS16, as it is best represented in available RGS domain– $G\alpha$ crystal structures^{27,28}. The RGS domain of RGS16 is different from those of RGS17 and RGS18 in 70 and 56 positions, respectively. To identify which of these residues in RGS17 and RGS18 are responsible for their impaired GAP activity,

we focused on the positions defined as either Significant & Conserved or Modulatory, reducing the candidate residues to 13 in RGS17 and 8 in RGS18. To further reduce the number of positions to mutate, we dismissed residues found at the corresponding positions in any of the high-activity RGS proteins (bold black, Figs. 4a and 5a). For example, Arg154 in RGS17 corresponds to a glutamic acid in RGS16; yet in the high-activity RGS1 this position is also an arginine, suggesting that Arg154 in RGS17 is not related to its low GAP activity.

We first applied these residue selection criteria to RGS17 and identified four sites that could be responsible for its low GAP activity:

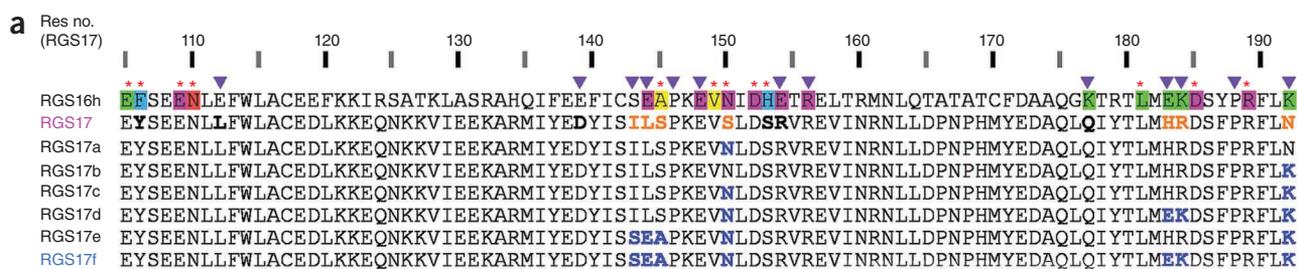


Figure 4 Redesign of RGS17 gain-of-function mutants.

(a) Sequences of RGS16 (annotated as in Fig. 2a), RGS17 and its mutants. RGS17 residues in Significant & Conserved or Modulatory positions that are different from RGS16 are marked as follows: those predicted to interfere with high GAP activity are orange, and those appearing in other high-activity RGS proteins are bold black. Residues in RGS17 mutants that were replaced by RGS16 residues are blue. (b) Positions of the four RGS17 sites mutated in the redesign experiments. The RGS16 residues used as the template for the redesign are visualized on the superimposed structures of $G\alpha_{11}$ -RGS16 (PDB 2IK8) and $G\alpha_o$ -RGS16 (PDB 3CK7), viewed through the semitransparent surface of the $G\alpha$ subunit. Corresponding RGS17 residue numbers are in parentheses. (c) GAP activities of the redesigned RGS17 mutants compared to activities of wild-type proteins. k_{gap} values were determined as in Figure 1 and are mean \pm s.e.m. ($n \geq 4$).

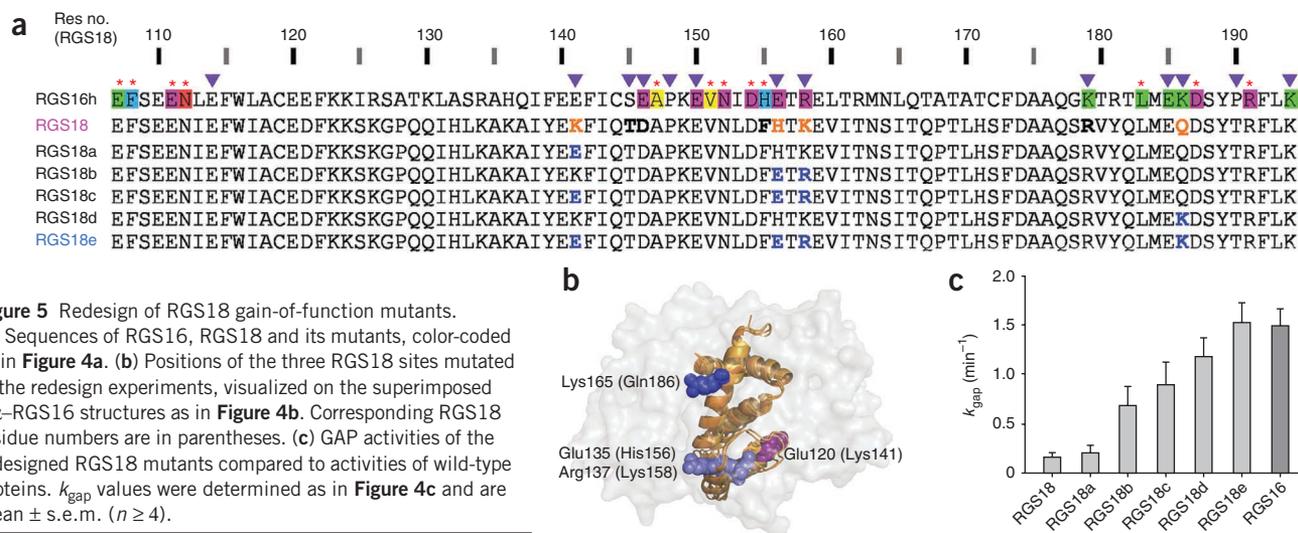


Figure 5 Redesign of RGS18 gain-of-function mutants.

(a) Sequences of RGS16, RGS18 and its mutants, color-coded as in **Figure 4a**. (b) Positions of the three RGS18 sites mutated in the redesign experiments, visualized on the superimposed $G\alpha$ -RGS16 structures as in **Figure 4b**. Corresponding RGS18 residue numbers are in parentheses. (c) GAP activities of the redesigned RGS18 mutants compared to activities of wild-type proteins. k_{gap} values were determined as in **Figure 4c** and are mean \pm s.e.m. ($n \geq 4$).

positions 143–145, 150, 183–184 and 192 (**Fig. 4a,b**). Two of these sites were predicted to impair activity because they lack side chains directly interacting with $G\alpha$ in high-activity RGS proteins. Ser150 is found at the RGS17 position occupied by a Significant & Conserved asparagine in all high-activity RGS proteins (**Fig. 2**). Indeed, the corresponding N128S mutation in RGS4 abolished its GAP function (**Fig. 3c** and ref. 29). Similarly, Asn192 in RGS17 corresponds to a lysine in all high-activity RGS proteins. The two remaining RGS17 sites (residues 143–145 and 183–184), containing mostly Modulatory residues, probably affect its GAP activity indirectly by displacing neighboring residues that interact with $G\alpha$ directly. Ser145, despite occupying the position of a Significant & Conserved alanine in high-activity RGS proteins, presumably affects the GAP activity of RGS17 indirectly: while the backbone of the corresponding RGS16 alanine interacts favorably with the $G\alpha$ subunit, the aliphatic side chain points into the RGS domain core. Thus, a serine in this position would probably necessitate a local repacking of the RGS protein, thereby affecting interactions with the $G\alpha$ subunit indirectly.

We measured the GAP activity of representative RGS17 mutants bearing different combinations of amino acid replacements at these four sites (**Fig. 4c**). Notably, the RGS17-to-RGS16 replacements of both ‘direct’ contributors (S150N or N192K), separately or together, did not increase RGS17 GAP activity at all (**Fig. 4c**). Even combining the S150N N192K double mutation with the replacement of the entire residue 143–145 site (containing Ser145) caused only a minor increase in activity. However, simultaneous substitution of all four RGS17 sites led to the same high GAP activity as in RGS16. Therefore, optimizing Modulatory positions in this protein was critical for achieving complete gain of function.

We applied a similar redesign to RGS18, also using the RGS16 template. Unlike RGS17, RGS18 has no Significant & Conserved positions that are different from those in high-activity RGS proteins. RGS18 does have four Modulatory positions in three distinct sites that could potentially impair its GAP activity: 141, 156+158 and 186 (**Fig. 5a,b**). In contrast to the minimal effect of partial mutagenesis in RGS17, two of three single-site mutants in RGS18 (H156E K158R and Q186K) markedly increased its GAP activity (**Fig. 5c**). Combining H156E K158R with K141E caused a slight additional improvement, and mutating all three sites simultaneously yielded full gain of function.

To test whether the increased GAP activity of the redesigned gain-of-function mutants was a result of increased affinity for the $G\alpha$ subunit, we assessed the binding of the series of redesigned RGS18

mutants (**Fig. 5a,c**) to $G\alpha$ using surface plasmon resonance (SPR) spectroscopy (**Table 1** and **Supplementary Fig. 3**). In accordance with their low GAP activity, the K_d values of RGS18 and its K141E mutant for $G\alpha$ were each $>3 \mu\text{M}$. However, the redesigned mutants that showed higher GAP activity had lower K_d values, with the highest-activity mutant (RGS18e) having the lowest K_d of 69 nM. These measurements show a strong correlation between GAP activity and $G\alpha$ -binding affinity for each RGS18 mutant. Taken together, our data demonstrate that optimizing Modulatory residues is sufficient for the restoration of maximal GAP activity of RGS18.

Comparison to alternative computational approaches

We compared our computational approach to other methods that predict residues contributing significantly to protein-protein interactions. We applied Rosetta’s computational alanine scanning³⁹ to the RGS domain- $G\alpha$ structures analyzed above. This method identified potential hot spots in each RGS protein corresponding to between five and eight of our Significant & Conserved residues and between zero and two Modulatory residues (**Supplementary Table 2**). As expected from an alanine-scanning protocol, Rosetta did not identify residues making substantial energy contributions via their backbones, but it also did not identify most Modulatory residues. This suggests that the majority of Modulatory positions in RGS domains do not make sufficient energy contributions to be identified as hot spots by computational alanine scanning. Indeed, we typically had to mutate multiple Modulatory residues to observe large changes in RGS activity (**Fig. 3**). Another reason why our approach identified more critical RGS residues may be that long-range electrostatics, which are not explicitly taken into account by Rosetta, have an important role at the RGS

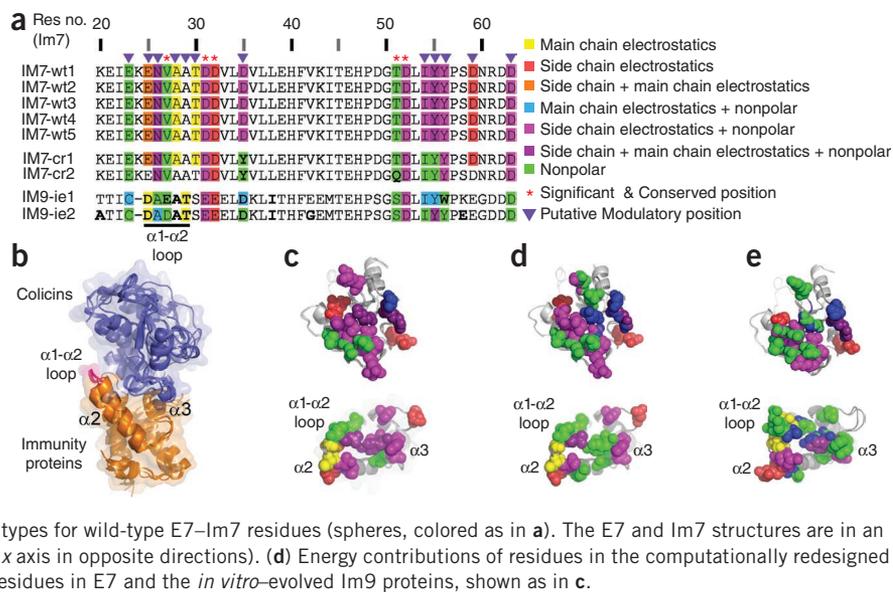
Table 1 Quantification of $G\alpha$ binding affinity of RGS18 mutants

	K_d (nM) ^a
RGS18	$>3,000^b$
RGS18a	$>3,000^b$
RGS18b	240 ± 70
RGS18c	215 ± 45
RGS18d	230 ± 80
RGS18e	69 ± 6

$G\alpha$ is bound to the transition state analog GDP-aluminum fluoride.

^a K_d values were calculated as weighed averages \pm s.e.m. from two or three independent experiments (see Online Methods for details). ^bThese K_d values are underestimates because dose response saturation was not reached.

Figure 6 Residues contributing substantially to colicin E7–immunity protein interactions. Energy calculations were carried out on the following structures: wild-type E7–Im7 complexes (wt1–wt5, PDB 7CEI, 2JAZ, 2JB0, 2JBG and 1ZNV); computationally redesigned E7–Im7 (cr1 and cr2, PDB 1UJZ and 2ERH); and E7 bound to Im9 proteins evolved *in vitro* to bind E7 with high affinity (ie1 and ie2, PDB 3GJN and 3GKL). (a) Residue-level sequence map of wild-type and engineered immunity proteins. Sequences in the multiple sequence alignment are taken from the crystal structures. Residues that contribute substantially to the interaction are color-coded according to type of energy contribution (see key). Consensus analysis was applied to the five wild-type proteins and Significant & Conserved and Modulatory positions were determined for all nine structures as in **Figure 2**. (b) The nine E7–Im structures, superimposed via the Im proteins. (c) Visualization of the energy-contribution types for wild-type E7–Im7 residues (spheres, colored as in a). The E7 and Im7 structures are in an ‘open book’ view (rotated 90° relative to b about the x axis in opposite directions). (d) Energy contributions of residues in the computationally redesigned E7–Im7, shown as in c. (e) Energy contributions of residues in E7 and the *in vitro*-evolved Im9 proteins, shown as in c.



domain–G α interface. Therefore, the physics-based energy calculations used in this study seem better suited to identifying residues in RGS proteins that are engaged in modulatory interactions.

We next used ConSurf⁴⁰ to test whether a sequence-based approach, which searches for phylogenetic relations between close homologs, can identify RGS residues that contribute to interactions with G α subunits. ConSurf calculated that the majority of Significant & Conserved residues had a conservation score above average, as we expected from residues that share a similar functional role among all high-activity RGS proteins. Seven additional residues at or near the RGS domain–G α interface were also identified as evolutionary conserved, although mutations in most of these residues had no effect on GAP function³⁶. The vast majority of RGS Modulatory residues had average or below average conservation scores and therefore were not pinpointed by this analysis.

A previous study obtained a more complete result using the Evolutionary Trace method³². This study identified an evolutionary privileged surface containing 17 RGS domain residues, 10 of which form a cluster of well-conserved contact residues judged not to have a role in determining specificity (we classified 8 of them as Significant & Conserved). Five of the seven remaining residues were defined as a second cluster of ‘class-specific’ residues (we classified four of them as Modulatory). In the case of RGS9, this cluster was suggested to form a binding site for the RGS9 adaptor protein, PDE γ , a concept experimentally confirmed in a subsequent study¹⁸. However, this study did not address the role of these evolutionary privileged residues in setting RGS–G protein specificity. Rather it highlighted that certain class-specific residues can participate in specific interactions with proteins other than G α subunits (for example, RGS9 interaction with PDE γ). This sequence-level superposition of overlapping interaction surfaces may provide an additional challenge for sequence-based methods (such as ConSurf and Evolutionary Trace), but not for structure-based methods like the approach used in our study.

Computational analysis of colicin E7–Im protein interactions

To explore the general applicability of our approach, we considered the interaction between the DNase colicin E7 (E7) and the inhibitory immunity protein Im7, a system used extensively to study specificity determinants in protein–protein interactions^{41,42}, interface specificity redesign^{43–45} and *in vitro* evolution studies⁴⁶. To map the contributions of individual residues to the interaction, we applied our

consensus-based comparative structural and energetic analysis to the five available crystal structures of E7–Im7 complexes (**Fig. 6** and **Supplementary Fig. 4**). These structures contained no E7 mutations near the Im7 interface and therefore were considered wild-type proteins in regard to Im7 binding (see Online Methods). We also applied our comparative analysis to the two structures of computationally redesigned E7–Im7 (refs. 43,44) and to the two structures of E7 bound to noncognate Im9 proteins selected through *in vitro* evolution for high E7 affinity⁴⁶.

Using the same criteria as for RGS proteins, we identified eight E7 positions and five Im positions as Significant & Conserved and seven E7 positions and 12 Im positions as Modulatory (**Fig. 6a** and **Supplementary Fig. 4**). The majority of these positions were shown to contribute to colicin–immunity protein binding and specificity^{41,47}. Notably, both the computationally redesigned and the *in vitro*-evolved protein pairs seem to use essentially the same complement of energetically important residues as the wild-type proteins. A minority of the residues in the computationally redesigned E7–Im7 use a different energy type of interaction (**Fig. 6a** and **Supplementary Fig. 4**; compare **Fig. 6c,d**); for example, the E7 K528Q mutation leads to a loss of electrostatic side chain contribution (**Supplementary Fig. 4**). In contrast, the *in vitro*-evolved Im9 proteins show a markedly different map of energy contributions (**Fig. 6a,e**).

Strikingly, our analysis identified residues in the Im7 α 1- α 2 loop as substantial contributors to the interactions with E7. This loop, located at the periphery of the E7–Im7 interface (**Fig. 6b**), has not been identified in earlier computational analyses and only recently has been implicated as having a role in binding specificity by the *in vitro* evolution study⁴⁶. We observe substantial contributions from these residues in all structures with a consistent theme of main chain electrostatic contributions. However, the overall pattern of energy contributions from residues in this loop is quite different in the *in vitro*-evolved Im9 proteins, suggesting that *in vitro* evolution revealed an alternative mode of interaction using this Im substructure.

DISCUSSION

Our study presents a new approach to pinpointing structural determinants that are critical for fine-tuning protein–protein interaction specificity. After recent successes in redesigning interaction affinity

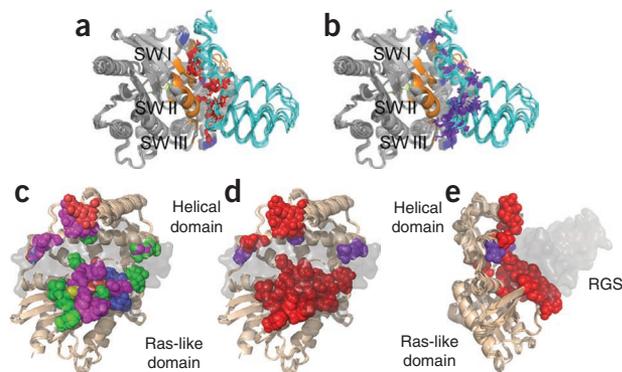


Figure 7 Positions of Significant & Conserved and Modulatory residues in the $G\alpha$ subunits interacting with RGS domains. **(a)** Significant & Conserved RGS residues (red) interact with all three $G\alpha$ switch regions (SW I–SW III). **(b)** Modulatory RGS residues (purple) interact with switch II and III and the helical domain of the $G\alpha$ subunits. In both **a** and **b**, Significant & Conserved positions within the $G\alpha$ subunits are orange and Modulatory positions are blue. **(c)** The different types of energetic contributions by individual $G\alpha$ residues (spheres, colored as in Fig. 2). The eight superimposed $G\alpha$ subunits are viewed through the semitransparent surface of the RGS domain and are rotated 90° about the y axis and 30° about the x axis relative to **a**. **(d)** Significant & Conserved and Modulatory residues in the $G\alpha$ structures (red and purple spheres, respectively). **(e)** Same as in **d**, rotated 90° about the y axis.

and/or specificity by combining computational analysis with experimental validation (for example, refs. 48–50), we combined the experimental benchmark of enzymatic assays with physics-based energy calculations using a consensus approach across multiple crystal structures. We find that RGS proteins contain a previously uninvestigated group of nonconserved residues that contribute to selective functional recognition of $G\alpha$. Accordingly, mutations of these Modulatory residues in two high-activity RGS proteins severely impaired their ability to accelerate $G\alpha$ GTPase, whereas redesigning low-activity RGS proteins by mutating critical Modulatory residues increased their GAP activity to the level observed in the highest-activity RGS domains.

We found that the typical quantitative impact of a single Modulatory residue on RGS GAP activity was smaller than that of Significant & Conserved residues. However, multiple Modulatory residues affected GAP function in a synergistic manner. In some cases, each single Modulatory residue made a small incremental contribution. In others, several Modulatory residues had to be mutated simultaneously to affect GAP activity substantially. The former is best represented by the loss-of-function mutants of RGS4 and 16; the latter is exemplified by the all-or-none gain-of-function effect of the redesigned RGS17 mutants.

Modulatory residues are located mostly at the periphery of the $G\alpha$ –RGS domain interface, where they contribute to $G\alpha$ subunit recognition. The center of this interface is occupied by Significant & Conserved residues that are thought to have the primary role in accelerating $G\alpha$ GTPase by stabilizing $G\alpha$ in a conformation optimal for GTP hydrolysis³¹. This arrangement probably enables RGS proteins to share a common mechanism of GAP function concomitantly with divergent levels of selectivity toward a given $G\alpha$ subunit. Furthermore, Significant & Conserved and Modulatory RGS residues show different patterns of $G\alpha$ interactions (Fig. 7). In the eight structures we analyzed, Significant & Conserved RGS residues interact with all three $G\alpha$ switch regions (Fig. 7a,b), as we expected from the pivotal role of the switch regions in GTP hydrolysis^{30,31}. Modulatory RGS residues interact with switch regions II and III, and with multiple residues in

the $G\alpha$ all-helical domain. We find the latter particularly intriguing because of the growing interest in the role of the all-helical domain in facilitating $G\alpha$ interactions with its regulatory partners^{27,33}. Notably, some Modulatory residues may interact with proteins other than $G\alpha$, as exemplified by RGS9 interactions with PDE γ .

In contrast to the variability of Modulatory residues among RGS proteins, the energy contributions of $G\alpha$ residues forming the reciprocal side of this interface are highly conserved (compare Figs. 7d,e and 2c,d). Almost all of these $G\alpha$ residues are classified by our energy-based calculations as Significant & Conserved, probably reflecting the fact that $G\alpha$ subunits analyzed in our calculations are all from the G_i subfamily ($G\alpha_{11}$, $G\alpha_{13}$, $G\alpha_0$ and $G\alpha_4$). This conservation may explain why some RGS proteins, whose isolated catalytic domains show similarly high GAP activity toward these $G\alpha$ subunits, rely on additional noncatalytic domains or adaptor proteins to discriminate among individual G_i family members^{7,13,15,20}. Multiple sequence alignment³¹ shows that other $G\alpha$ subfamilies (for example, G_s and $G_{12/13}$) are quite different from G_i at the positions interacting with RGS Modulatory residues. This hints at how the specificity of RGS domain recognition may be achieved across the entire $G\alpha$ family, which can be investigated in future studies.

From a methodological perspective, our approach to redesigning protein-protein interactions bypasses the computational bottleneck of commonly used protein design methods—searching both sequence and three-dimensional (3D) structure space simultaneously to find promising design candidates^{45,51}. Rather, we used comparative information across the RGS protein family (via our sequence-level map) as a shortcut to identify the RGS domain sites that were most attractive for redesign mutagenesis. Furthermore, our approach does not depend on improving protein-protein interactions by mutating individual residues one at a time and combining mutations showing notable individual experimental effects, the approach used in some of the most successful earlier studies (reviewed in ref. 45). Using such a strategy for RGS17 would have failed because individual mutations in this protein did not measurably increase its GAP activity. Our successes in redesigning RGS domain interactions and in predicting the determinants of interactions between colicins and immunity proteins suggest that physics-based energy functions can complement the engineered energy functions commonly used in protein design, both in analyzing design templates and assessing design products.

In conclusion, our work provides a quantitative framework for understanding the determinants of selective RGS protein interactions with G proteins and enables structure-based redesign of protein-protein interactions at the family level. It can be extended to design a variety of RGS protein and G protein mutants with distinct activities and selectivities as tools to decipher G protein signaling networks in living cells. Given the growing number of available structures of representative protein-protein complexes (for example, ref. 52), this methodology can be easily adapted to study interaction specificity across other large protein families.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This research was supported by US National Institutes of Health grants EY012859 (V.Y.A.) and GM082892 (D.P.S.), by a core grant for vision research to Duke University (EY5722), by the US National Science Foundation through TeraGrid

resources (TG-MCB080085T; M.K.) and by a long-term postdoctoral fellowship from the Human Frontier Science Program (M.K.). We thank the Duke Shared Cluster Resource and the San Diego Supercomputer Center for computational resources, S.A. Baker, S. Farsiu, N.P. Skiba, E.S. Lobanova and D. Reichmann for helpful suggestions, B. Honig for insightful guidance (M.K.) and F. Sheinerman, R. Rohs, S. Fleishman and E. Alexov for helpful discussions.

AUTHOR CONTRIBUTIONS

M.K. designed and carried out computational analysis and biochemical experiments, analyzed data and prepared the manuscript, A.M.T. carried out experiments and prepared the manuscript, D.E.B. carried out experiments and prepared the manuscript, D.P.S. supervised the project and prepared the manuscript and V.Y.A. supervised the project and analysis and prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Siderovski, D.P., Hessel, A., Chung, S., Mak, T.W. & Tyers, M. A new family of regulators of G-protein-coupled receptors? *Curr. Biol.* **6**, 211–212 (1996).
- Koelle, M.R. & Horvitz, H.R. EGL-10 regulates G protein signaling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell* **84**, 115–125 (1996).
- Berman, D.M., Wikie, T.M. & Gilman, A.G. GAIP and RGS4 are GTPase-activating proteins for the Gi subfamily of G protein α subunits. *Cell* **86**, 445–452 (1996).
- Hunt, T.W., Fields, T.A., Casey, P.J. & Peralta, E.G. RGS10 is a selective activator of G α_q GTPase activity. *Nature* **383**, 175–177 (1996).
- Watson, N., Linder, M.E., Druey, K.M., Kehrl, J.H. & Blumer, K.J. RGS family members: GTPase-activating proteins for heterotrimeric G-protein α -subunits. *Nature* **383**, 172–175 (1996).
- Ross, E.M. & Wilkie, T.M. GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. *Annu. Rev. Biochem.* **69**, 795–827 (2000).
- Hollinger, S. & Hepler, J.R. Cellular regulation of RGS proteins: modulators and integrators of G protein signaling. *Pharmacol. Rev.* **54**, 527–559 (2002).
- Neubig, R.R. & Siderovski, D.P. Regulators of G-protein signalling as new central nervous system drug targets. *Nat. Rev. Drug Discov.* **1**, 187–197 (2002).
- Neitzel, K.L. & Hepler, J.R. Cellular mechanisms that determine selective RGS protein regulation of G protein-coupled receptor signaling. *Semin. Cell Dev. Biol.* **17**, 383–389 (2006).
- Hurst, J.H. & Hooks, S.B. Regulator of G-protein signaling (RGS) proteins in cancer biology. *Biochem. Pharmacol.* **78**, 1289–1297 (2009).
- Wang, Q., Liu, M., Mullah, B., Siderovski, D.P. & Neubig, R.R. Receptor-selective effects of endogenous RGS3 and RGS5 to regulate mitogen-activated protein kinase activation in rat vascular smooth muscle cells. *J. Biol. Chem.* **277**, 24949–24958 (2002).
- Tang, K.M. *et al.* Regulator of G-protein signaling-2 mediates vascular smooth muscle relaxation and blood pressure. *Nat. Med.* **9**, 1506–1512 (2003).
- Xie, G.X. & Palmer, P.P. How regulators of G protein signaling achieve selective regulation. *J. Mol. Biol.* **366**, 349–365 (2007).
- Cifelli, C. *et al.* RGS4 regulates parasympathetic signaling and heart rate control in the sinoatrial node. *Circ. Res.* **103**, 527–535 (2008).
- Bansal, G., Druey, K.M. & Xie, Z. R4 RGS proteins: regulation of G-protein signaling and beyond. *Pharmacol. Ther.* **116**, 473–495 (2007).
- Bansal, G., Xie, Z., Rao, S., Nocka, K.H. & Druey, K.M. Suppression of immunoglobulin E-mediated allergic responses by regulator of G protein signaling 13. *Nat. Immunol.* **9**, 73–80 (2008).
- Laroché, G., Giguère, P.M., Roth, B.L., Trejo, J. & Siderovski, D.P. RNA interference screen for RGS protein specificity at muscarinic and protease-activated receptors reveals bidirectional modulation of signaling. *Am. J. Physiol. Cell Physiol.* **299**, C654–C664 (2010).
- Sowa, M.E. *et al.* Prediction and confirmation of a site critical for effector regulation of RGS domain activity. *Nat. Struct. Biol.* **8**, 234–237 (2001).
- Martemyanov, K.A. & Arshavsky, V.Y. Biology and functions of the RGS9 isoforms. *Prog. Mol. Biol. Transl. Sci.* **86**, 205–227 (2009).
- Skiba, N.P., Hopp, J.A. & Arshavsky, V.Y. The effector enzyme regulates the duration of G protein signaling in vertebrate photoreceptors by increasing the affinity between transducin and RGS protein. *J. Biol. Chem.* **275**, 32716–32720 (2000).
- Heximer, S.P., Watson, N., Linder, M.E., Blumer, K.J. & Hepler, J.R. RGS2/GOS8 is a selective inhibitor of G α_q function. *Proc. Natl. Acad. Sci. USA* **94**, 14389–14393 (1997).
- Heximer, S.P. *et al.* G protein selectivity is a determinant of RGS2 function. *J. Biol. Chem.* **274**, 34253–34259 (1999).
- Ingi, T. *et al.* Dynamic regulation of RGS2 suggests a novel mechanism in G-protein signaling and neuronal plasticity. *J. Neurosci.* **18**, 7178–7188 (1998).
- Tesmer, J.J., Berman, D.M., Gilman, A.G. & Sprang, S.R. Structure of RGS4 bound to AIF4-activated G α_{i1} : stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251–261 (1997).
- Kimple, A.J. *et al.* Structural determinants of G-protein α subunit selectivity by regulator of G-protein signaling 2 (RGS2). *J. Biol. Chem.* **284**, 19402–19411 (2009).
- Slep, K.C. *et al.* Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 Å. *Nature* **409**, 1071–1077 (2001).
- Soundararajan, M. *et al.* Structural diversity in the RGS domain and its interaction with heterotrimeric G protein α -subunits. *Proc. Natl. Acad. Sci. USA* **105**, 6457–6462 (2008).
- Slep, K.C. *et al.* Molecular architecture of G α_o and the structural basis for RGS16-mediated deactivation. *Proc. Natl. Acad. Sci. USA* **105**, 6243–6248 (2008).
- Posner, B.A., Mukhopadhyay, S., Tesmer, J.J., Gilman, A.G. & Ross, E.M. Modulation of the affinity and selectivity of RGS protein interaction with G α subunits by a conserved asparagine/serine residue. *Biochemistry* **38**, 7773–7779 (1999).
- Kosloff, M. & Selinger, Z. GTPase catalysis by Ras and other G-proteins: insights from Substrate Directed Superimposition. *J. Mol. Biol.* **331**, 1157–1170 (2003).
- Sprang, S.R., Chen, Z. & Du, X. Structural basis of effector regulation and signal termination in heterotrimeric G α proteins. *Adv. Protein Chem.* **74**, 1–65 (2007).
- Sowa, M.E., He, W., Wensel, T.G. & Lichtarge, O. A regulator of G protein signaling interaction surface linked to effector specificity. *Proc. Natl. Acad. Sci. USA* **97**, 1483–1488 (2000).
- Siderovski, D.P. & Willard, F.S. The GAPs, GEFs, and GDIs of heterotrimeric G-protein α subunits. *Int. J. Biol. Sci.* **1**, 51–66 (2005).
- Sheinerman, F.B., Al-Lazikani, B. & Honig, B. Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J. Mol. Biol.* **334**, 823–841 (2003).
- Sheinerman, F.B. & Honig, B. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.* **318**, 161–177 (2002).
- Srinivasa, S.P., Watson, N., Overton, M.C. & Blumer, K.J. Mechanism of RGS4, a GTPase-activating protein for G protein α subunits. *J. Biol. Chem.* **273**, 1529–1533 (1998).
- Natochin, M., McEntaffer, R.L. & Artemyev, N.O. Mutational analysis of the Asn residue essential for RGS protein binding to G-proteins. *J. Biol. Chem.* **273**, 6731–6735 (1998).
- Wieland, T., Bahtijari, N., Zhou, X.B., Kleuss, C. & Simon, M.I. Polarity exchange at the interface of regulators of G protein signaling with G protein α -subunits. *J. Biol. Chem.* **275**, 28500–28506 (2000).
- Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA* **99**, 14116–14121 (2002).
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
- Kühlmann, U.C., Pommer, A.J., Moore, G.R., James, R. & Kleantous, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178 (2000).
- Schreiber, G. & Keating, A.E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2010).
- Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379 (2004).
- Joachimski, L.A., Kortemme, T., Stoddard, B.L. & Baker, D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol.* **361**, 195–208 (2006).
- Mandell, D.J. & Kortemme, T. Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* **5**, 797–807 (2009).
- Levin, K.B. *et al.* Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol.* **16**, 1049–1055 (2009).
- Li, W. *et al.* Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J. Mol. Biol.* **337**, 743–759 (2004).
- Lippow, S.M., Wittrup, K.D. & Tidor, B. Computational design of antibody-affinity improvement beyond *in vivo* maturation. *Nat. Biotechnol.* **25**, 1171–1176 (2007).
- Skerker, J.M. *et al.* Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
- Grigoryan, G., Reinke, A.W. & Keating, A.E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864 (2009).
- Karanicolas, J. & Kuhlman, B. Computational design of affinity and specificity at protein-protein interfaces. *Curr. Opin. Struct. Biol.* **19**, 458–463 (2009).
- Edwards, A. Large-scale structural biology of the human proteome. *Annu. Rev. Biochem.* **78**, 541–568 (2009).

ONLINE METHODS

Atomic structural models. The atomic models of the RGS domain–G α complexes used in the calculations were taken from the following PDB entries: 1AGR (G α_{11} –RGS4); 2IK8 (G α_{11} –RGS16); 3C7K (G α_6 –RGS16); 2IHB (G α_{13} –RGS10); 2GTP (G α_{11} –RGS1); 2ODE (G α_{13} –RGS8); 1FQJ (G $\alpha_{11/1}$ –RGS9); 2V4Z (G α_{13} –RGS2 C106S N184D E191K triple mutant)^{24–28}. Atomic models of colicin–immunity protein complexes were taken from the following PDB entries: 7CEI, 2JAZ, 2JB0, 2JBG and 1ZNV (wild-type E7–Im7^{53–55}, although some of these E7 proteins contain point mutations, these mutations are far from the Im7-binding site and therefore these chains were considered wild type); 1UJZ and 2ERH (computationally redesigned E7–Im7)^{43,44}; 3GJN and 3GKL (E7 bound to *in vitro*-evolved Im9)⁴⁶. Missing short segments in PDB entries 2IK8 (G α_{11} residues 112–118), 2IHB (RGS10 residues 103–113) and 2GTP (G α_{11} residues 112–118) were modeled on the basis of the G α_{11} –RGS4 structure (PDB 1AGR) using the program Nest⁵⁶ and partial or missing side chains were modeled using Scap⁵⁷. Similarly, a short missing E7 segment in the following structures was modeled on the basis of PDB 7CEI: PDB entries 2JBG (residues 547–554), 2JAZ (residues 548–554), 2JB0 (residues 551–552), 1ZNV (residues 547–554), 3GJN (residues 549–554) and 3GKL (residues 548–554).

Hydrogen atoms were added using CHARMM⁵⁸, and the structures were subjected to conjugate gradient minimization with a harmonic restraint force of 50 kcal mol⁻¹ Å⁻² applied to the heavy atoms.

Calculating residue-level electrostatic and nonpolar free energy contributions.

Electrostatic potentials and free energies were calculated using the DelPhi program⁵⁹. DelPhi yields finite-difference solutions to the Poisson-Boltzmann equation (the FDPB method) for a system in which the solvent is described in terms of a bulk dielectric constant and concentrations of mobile ions, whereas the solutes are described in atomic detail by the coordinates of individual atoms, atomic radii and partial charges. The proteins were mapped onto a fine 3D grid, in which each small cube represents a small region of the protein or solvent. Charges and radii were taken from the CHARMM22 parameter set. Regions inside the molecular surfaces of the proteins were assigned a dielectric constant of 2, and those outside a dielectric constant of 80, combined with an ion exclusion layer of 2 Å around the solute. These particular parameters have been optimized for energetic calculations of protein-protein interactions and have been validated extensively for numerous systems (see refs. 34,35 and references therein). The ionic strength was set to 100 mM to approximate the experimental conditions. The numerical calculation of the potential was iterated to convergence, defined as the point at which the potential changes <10⁻⁵ kT e⁻¹ between successive iterations. A sequence of focusing runs of increasing resolution was used to calculate the electrostatic potentials (for example, 0.375, 0.75, 1.5 and 3.0 grids per angstrom). Electrostatic energies were obtained using the calculated potentials, and the net electrostatic energy of a protein-protein interaction was determined as the difference between (i) the electrostatic free energy of the proteins in complex and (ii) the electrostatic free energies of each of the proteins infinitely far apart (that is, calculated separately).

We used the FDPB method as described^{34,35}, coupled with *in silico* mutagenesis, to calculate the net electrostatic and polar energetic contributions ($\Delta\Delta G_{\text{elec}}$) of a residue to the interaction with its protein partner resulting from the removal of partial and real charges of each residue. This would correspond to an *in silico* residue that is identical in shape and dielectric permittivity to the original residue, but is now partially or completely nonpolar. For each residue, this was repeated twice, once neutralizing backbone and side chain and once neutralizing the side

chain only. We thereby differentiated between energetic contributions from the side chain versus the main chain (**Supplementary Fig. 1**). We considered all residues within 15 Å of the RGS domain–G protein interface; this distance threshold (~1.5 debye lengths) was a compromise between identifying electrostatic contributions from residues distal to the interface and avoiding excessively long computational times. We checked the consequences of this distance threshold by repeating the calculations for G α_{11} –RGS4 without any distance threshold. All residues >15 Å from the interface contributed <1 kcal mol⁻¹ to the interaction.

The nonpolar energetic contribution ($\Delta\Delta G_{\text{np}}$) of each residue was calculated as a surface area–proportional term, obtained by multiplying the per-residue surface area buried upon complex formation by a surface tension constant of 0.05 kcal mol⁻¹ Å⁻² (**Supplementary Fig. 1**)³⁴. Solvent-accessible surface areas were calculated using the *surf* program⁶⁰.

Test calculations using small translations (0.1–0.2 Å), rotations (5°) of the proteins, or changes in the grid size estimated the numerical error in $\Delta\Delta G_{\text{elec}}$ calculations as <0.5 kcal mol⁻¹. According to the more stringent criteria of ref. 34, we defined residues that made substantial electrostatic contributions to the interactions with their cognate partners as those contributing $\Delta\Delta G_{\text{elec}} \geq 1$ kcal mol⁻¹ to binding. Similarly, residues contributing $\Delta\Delta G_{\text{np}} \geq 1$ kcal mol⁻¹ (≥ 20 Å² buried upon complex formation) were selected as substantial nonpolar energetic contributors. To reduce false positives and negatives, we used a consensus approach: residues conserved across all structures with comparable GAP activities (for RGS domains) or affinities (for colicin–immunity protein complexes) and that were calculated to have substantial interactions in the majority of structures were considered to contribute substantially to the interaction in all these structures. Residues conserved across all such structures calculated to have substantial interactions in fewer than two structures were considered false positives. This consensus approach improved the accuracy of our predictions, as we encountered several false positives and negatives owing to a different side chain rotamer found in only one structure, despite that residue being strictly conserved and in a comparable 3D neighborhood (see **Fig. 2** and **Supplementary Fig. 1**). RGS domain residues thus determined to contribute substantially were mapped onto a sequence map (for example, **Fig. 2a**).

Full methods. Methods for protein expression, purification, GTPase assays and SPR assays are in **Supplementary Methods**.

53. Doudeva, L.G. *et al.* Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn²⁺ or inhibitor/Ni²⁺. *Protein Sci.* **15**, 269–280 (2006).
54. Ko, T.P., Liao, C.C., Ku, W.Y., Chak, K.F. & Yuan, H.S. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* **7**, 91–102 (1999).
55. Huang, H. & Yuan, H.S. The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *J. Mol. Biol.* **368**, 812–821 (2007).
56. Petrey, D. *et al.* Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* **53** (Suppl. 6) 430–435 (2003).
57. Xiang, Z. & Honig, B. Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* **311**, 421–430 (2001).
58. MacKerell, A.D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
59. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
60. Sridharan, S., Nicholls, A. & Honig, B. A new vertex algorithm to calculate solvent accessible surface areas. *FASEB J.* **6**, A174 (1992).