Title

A complete allosteric map of a GTPase switch in its native cellular network.

Permalink

https://escholarship.org/uc/item/3mf001rk

Journal

Cell Systems, 14(3)

Authors

Mathy, Christopher

Mishra, Parul

Flynn, Julia

et al.

Publication Date

2023-03-15

DOI

10.1016/j.cels.2023.01.003

Peer reviewed

*Correspondence: tanjakortemme@gmail.com, Dan.Bolon@umassmed.edu.

Author contributions

C.J.P.M., D.N.A.B. and T.K. identified and developed the core questions. C.J.P.M. performed the majority of the data analysis, with contributions from D.M., T.P., P.M., D.N.A.B. and T.K., and the biophysical experiments. P.M. performed the mutational scanning experiments. J.M.F. performed the spotting assays and western blot experiments. C.J.P.M. and T.K. wrote the manuscript with contributions from the other authors. D.N.A.B. and T.K. oversaw the project.

Supplemental Information

Document S1. Figures S1–S7, Table S1, and Methods S1.

Data S1. Gsp1 fitness scores with bins and raw read counts

Column definitions:

- *mutant*: descriptor of the allele (e.g. F28V)

- *aa_from*: wild-type amino acid (e.g. F)

- *position*: sequence position of the mutation (e.g. 28)

- *aa_to*: substituted amino acid (e.g. V)

- *counts_0gen*: number of reads corresponding to the allele at the beginning of selection

- *counts_6gen*: number of reads corresponding to the allele after 6 generations of selection

- *score*: computed fitness score, a log2-transformed changes in variant abundance relative to wild-type

- *bin*: assigned score bin

- *low_reads_flag*: boolean flag stating whether the allele had low read counts in the initial sample (< 2% of the average allele's number of reads)

Data S2. Gsp1   G data

Column definitions:

- *mutation*: descriptor of the mutation (e.g. F28V)

- *aa_from*: wild-type amino acid (e.g. F)

- *aa_to*: substituted amino acid (e.g. V)

- *pdb_id*: PDB ID of the crystal structure used for the calculation

- *species*: organism of the gene used for determination of the crystal structure

- *pos_Sc*: sequence position number of the corresponding residue in *S. cerevisiae* Gsp1

- *aa_Sc*: wild-type amino acid of the corresponding residue in *S. cerevisiae* Gsp1

- *pos_Hs*: sequence position number of the corresponding residue in *H. sapiens* Ran

- *aa_Sc*: wild-type amino acid of the corresponding residue in *H. sapiens* Ran

- *ddg*: computed   G of the mutation, in Rosetta Energy Units (REU), scaled so that 1 REU ~ 1 kcal/mol based on the benchmark data

Data S3. Benchmark   G data

Column definitions:

- *record_id*: unique ID used in this study for the benchmark mutation

- *pdb*: PDB ID of the crystal structure used for the calculation

- *mutation_full*: long form ID for the mutation made (e.g. A T 100 G corresponds to mutating the Threonine at position 100 in chain A to Glycine)

- *chain*: one-letter ID for the crystal structure chain used for the calculation

A complete allosteric map of a GTPase switch in its native cellular network

Christopher J.P. Mathy[1,2,3], **Parul Mishra**[4], **Julia M. Flynn**[4], **Tina Perica**[1,2], **David Mavor**[4], **Daniel N.A. Bolon**[4,*], **Tanja Kortemme**[1,2,3,5,6,*]

[1]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco; San Francisco, CA 94158, USA.

[2]Quantitative Biosciences Institute, University of California, San Francisco; San Francisco, CA 94158, USA.

[3]The UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, San Francisco; San Francisco, CA 94158, USA.

[4]Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Medical School; Worcester, MA 01605, USA

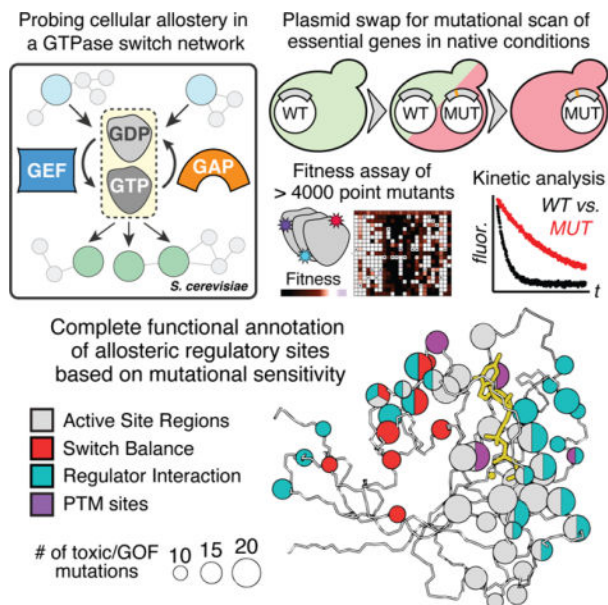[5]Chan Zuckerberg Biohub; San Francisco, CA 94158, USA.

[6]Lead contact

Summary

Allosteric regulation is central to protein function in cellular networks. A fundamental open question is whether cellular regulation of allosteric proteins occurs only at a few defined positions or at many sites distributed throughout the structure. Here we probe the regulation of GTPases - protein switches that control signaling through regulated conformational cycling - at residue-level resolution by deep mutagenesis in the native biological network. For the GTPase Gsp1/Ran, we find that 28% of the 4,315 assayed mutations show pronounced gain-of-function responses. Twenty of the sixty positions enriched for gain-of-function mutations are outside the canonical GTPase active site switch regions. Kinetic analysis shows that these distal sites are allosterically coupled to the active site. We conclude that the GTPase switch mechanism is broadly sensitive to cellular allosteric regulation. Our systematic discovery of new regulatory sites provides a functional map to interrogate and target GTPases controlling many essential biological processes. A record of this paper's Transparent Peer Review process is included in the Supplemental Information.

Graphical Abstract

- *mutation*: short form descriptor of the mutation (e.g. T100G)

- *position*: sequence position of the mutation (e.g. 100)

   ddg_expt: experimental   G of the mutation, in kcal/mol

- *ddg_calc*: computed   G of the mutation, in unscaled Rosetta Energy Units (REU)

- *ddg_calc_adj*: computed   G of the mutation, in REU, scaled down so that 1 REU ~ 1 kcal/mol

Complete functional annotation of allosteric regulatory sites based on mutational sensitivity

☐ Active Site Regions
☐ Switch Balance
☐ Regulator Interaction
☐ PTM sites

# of toxic/GOF mutations   10 15 20

eTOC blurb:

Allostery is essential to protein regulation in cellular networks. Using saturation mutagenesis of the GTPase Gsp1/Ran in its cellular context, we find allosteric sites to be prevalent throughout the GTPase fold. We show that perturbations at these sites impact function by tuning the kinetics of switching between the GTPase states.

Keywords

GTPases; Allostery; Mutational Scanning; Gsp1; RAN

Introduction

Allostery, the process by which perturbations at one site of a protein exert functional effects at distal sites, is a central regulatory mechanism in cells.[1] Protein or ligand binding, posttranslational modifications, and mutations can allosterically alter enzymatic activities or subsequent binding events to control metabolism[2] or signaling,[3] making allosteric regulation a driver of disease and attractive target for therapeutic drug design.[4] While it has been suggested that a considerable fraction of sites in a protein may be primed for allosteric regulation[5] and this priming may enable the evolution of new functional protein-protein interactions,[6] it remains an open question how prevalent allosteric sites are in a protein structure. Moreover, while biophysical aspects of allostery have been mapped using technological advances,[5] the role of allosteric perturbations on cellular function in physiological networks has not been mapped comprehensively even for single proteins. One contributor is a lack of unbiased methods for discovering sites of allosteric regulation in the cellular context, thus limiting the identification of new targets for drug development and the reprogramming of functions in cellular networks.

A class of proteins thought to be regulated through allosteric mechanisms are switches, which cycle between "on" and "off" states in response to signals, are ubiquitous in biological regulation,[7] and whose misregulation is often associated with disease.[8] In small GTPase switches, interconversion between a GTP-bound on-state and a GDP-bound off-state is intrinsically slow but is accelerated by two opposing regulators: GTPase-activating (GAP) proteins that activate GTP hydrolysis and guanine nucleotide exchange factor (GEF) proteins that accelerate nucleotide replacement. Perturbations at a very limited number of allosteric sites distal from the active site, which comprises the nucleotide binding region and the switch loops,[9] have been shown to affect the kinetics of biochemical switching function *in vitro*[3] and to lead to switch overactivation[10] and altered cellular function.[3] Structural and computational studies in the GTPase Ras have identified allosteric sites,[11,12] for example a distal calcium acetate binding site whose occupancy results in ordering of the active site loop switch II,[13] as well as correlated motions between the active site and secondary structure units near the membrane-binding interface.[14] Additionally, an allosteric site of Ras adjacent to switch II has been successfully targeted by small molecule inhibitors.[15] Despite these key findings, the vast majority of GTPase sites remain untested for allosteric regulation in their native biological networks[16] when the functional context of opposing regulators, posttranslational modifications, interaction partners, and downstream signaling pathways is preserved (Figure 1A).

Here we introduce an approach to generate a complete functional map of the essential eukaryotic GTPase switch Gsp1/Ran in the native context of its *in vivo* interaction network in *S. cerevisiae* based on comprehensive mutational perturbation.[17,18] Gsp1/Ran uses a single pair of regulators, the GAP Rna1 and the GEF Srm1, but an extended network of adaptor and effector proteins, whose interactions with Gsp1/Ran are dependent on switch state, control diverse processes including nucleocytoplasmic transport, cell cycle progression and RNA processing.[3] Gsp1 is highly conserved, with 82% of its amino acid sequence identical to the human homolog Ran. With some notable exceptions,[10,19] prior mutational scanning experiments have revealed a tolerance to mutations even among highly conserved proteins,[20] suggesting missing biological context.[21] In contrast, for Gsp1 in its physiological network, here we report that cellular function is affected by mutations at a large number of previously uncharacterized positions outside the active site, identifying widespread sensitivity of a central GTPase to allosteric regulation. Moreover, our results suggest a unifying mechanism where perturbations to the balance between the on and off states of the switch underlie the cellular effects.

Results

Gsp1 function in the cellular context is broadly sensitive to mutational perturbations.

To systematically measure the effect of all Gsp1 mutations on cellular function (Figure 1A), we developed an approach derived from our EMPIRIC (extremely methodical and parallel investigation of randomized individual codons) method[22] but with a generalizable plasmid dropout selection to probe the function of essential genes (Figure 1B, Methods). We transformed a chromosomal GSP1 knockout strain with the wild-type (WT) GSP1 allele under the control of its native promoter on a URA selectable plasmid harboring

constitutively expressed GFP, and confirmed Gsp1 protein expression via western blot (Figure S1). We introduced a library of all possible single Gsp1 mutants, also expressed from the native Gsp1 promoter, using a HIS selectable plasmid harboring constitutively expressed mCherry. We sorted for cells expressing mCherry (library plasmid) but not GFP (WT plasmid) and compared allele abundances from the initial population to the population after six generations of growth to compute fitness scores for all 19 possible single amino acid substitutions as well as WT synonymous (WT-syn) and STOP codons at every position in Gsp1 (Figure 1C, Methods). This approach interrogates variant fitness both in the presence and absence of a WT copy with the potential to inform on both gain of toxic function and loss of normal function.

We categorized the fitness score of each mutation relative to the distributions of fitness scores for WT-syn and STOP codons (Figure 1D, Methods). Compared to the WT-syn distribution, 48.5% of all mutations showed deleterious fitness effects, while very few mutations (15/4315 or 0.35%) appeared beneficial (primarily attributable to low read counts for these mutations in our library pre-selection, see Methods and Figure S2). We observed strongly deleterious mutations in the GTPase active site, which we define as the highly conserved G1–5 functional regions of the Ras superfamily of small GTPases (including the switch loops that change conformation in the GDP- and GTP-bound states) and any additional residues contacting the nucleotide (Figure 1C and Figure S3).

The distribution of STOP codon scores (Figure 1D, E) fell into two groups: STOP codon mutations before Gsp1 sequence position 175 had narrowly distributed fitness scores no lower than −2.90 (scores are $\log_2$-transformed changes in variant abundance relative to wild-type). In contrast, STOP mutations after position 175 had substantially lower fitness scores (down to −10.5). Positions 1–174 comprise a standard GTPase fold, whereas positions 175–219 comprise a C-terminal extension specific to the Ran subfamily (Figure 1C and Figure S3A). Thus, the first set of STOP codon mutants (positions 1–174) likely represent the growth defect of a null Gsp1 mutant, as internal truncations in the GTPase fold likely result in nonfunctional proteins. Because loss of the wild-type plasmid and protein can occur throughout the growth phase of the experiment, the meaning of absolute fitness measurements is complex. For example, over the six generations of wild-type growth between sampling we would expect a null allele growing purely in the absence of wild-type to have a fitness score of −6 corresponding to a 2^6-fold depletion compared to wild-type. However, we observe an average fitness score of around −2 for null alleles, indicating an effective selection without wild-type of two generations.

Mutations with worse scores than null alleles must have a functional effect more detrimental than loss-of-function, and we termed these mutations "toxic gain-of-function", or toxic/GOF. Using a conservative definition of scores worse than the mean STOP codon mutation score at positions 1–174 by more than three standard deviations, more than half of all deleterious mutations (58.4%, and 28.4% of all mutations) were toxic/GOF. Toxic/GOF mutations were not exclusive to the active site regions defined above, but were broadly distributed across the Gsp1 structure, including in interfaces with Gsp1 partner proteins, in parts of the Gsp1 buried core, and at surface positions outside of the interaction interfaces (Figure 2A).

Structural locations of toxic/GOF mutations are not exclusive to the active site or the protein core.

Both the prevalence of toxic/GOF mutations and their locations across the GTPase fold were unexpected. To identify potential mechanisms underlying these findings, we defined sequence positions that were enriched in toxic/GOF mutations. We counted the number of toxic/GOF mutations at each position and compared this observed distribution to a hypergeometric null distribution parameterized according to the total number of toxic/GOF mutations in the dataset (Figure S1C, Methods). Positions with 10 or more toxic/GOF mutations showed significant enrichment and were labeled as toxic/GOF positions. In total, 60 out of 219 Gsp1 sequence positions were toxic/GOF; 57 of these positions were identical in amino acid identity between *S. cerevisiae* Gsp1 and human Ran. While many other positions (91/219) showed 1–9 toxic/GOF mutations, and many others showed a small number of STOP-like or intermediate mutations, we chose to focus on the set of positions most enriched in toxic/GOF mutations as they most likely indicated regions at which energetic perturbations can impact cellular function of Gsp1.

Given most substantial fitness effects observed in mutational perturbation studies are typically from mutations at residues in active sites required for function, or at residues in the protein core critical for stability, we asked whether the locations of toxic/GOF positions overlapped with the active site or the core. Only half (30/60) of the toxic/GOF positions are in the active site (Figure 2B, blue) and an additional 10 positions are in the C-terminal extension. Thus, 20/60 toxic/GOF positions are situated in the GTPase fold but distal to the active site (Figure 2B, **red**). 16 out of the 46 active site positions are not toxic/GOF. Conversely, only 19 out of the 60 toxic/GOF positions are in the buried protein core (Figure 2C, red), and 43 out of the 62 core positions are not toxic/GOF (Figure 2C, orange). Moreover, mutations at active site positions would typically be expected to ablate function and therefore lead to a loss-of-function phenotype (similar to STOP). However, we observe 517/848 (61%) of mutations at active site positions are toxic/GOF, compared to only 93 (11%) STOP-like mutations (Figure 2A). Similarly, mutations at positions in the protein core that destabilize Gsp1 would be expected to exhibit a fitness cost similar to that observed for STOP codons in the GTPase fold, but not be toxic/GOF. In addition, computational stability calculations (Methods) showed little correlation between predicted destabilization and decreased fitness when including toxic/GOF mutations, and only a modest correlation for mutations in the buried core when excluding toxic/GOF mutations (Figure S4). Thus, the mechanism of Gsp1 toxic/GOF mutations is not satisfactorily explained by either simply the location in the active site or by destabilization of the protein.

Distal toxic/GOF mutations allosterically alter the balance of the switch states.

The prevalence of toxic/GOF mutations in the C-terminal extension (Figure 1C, E) provided the first evidence that the toxicity of the mutants stems from perturbed regulation: Deleting the C-terminus of Ran/Gsp1 is known to alter the balance between the switch states by stabilizing the GTP-bound form,[23] which may explain the enrichment of cancer mutations in the C-terminus of Ran.[24] We therefore asked whether all toxic/GOF mutations (Figure 4A) perturbed Gsp1 GTPase switch function. This model would account for the toxic/GOF effects of mutations at the 40 positions in the GTPase active site or C-terminus. Of the

remaining 20 toxic/GOF distal sites within the GTPase fold (Figure 2B), 13 are located in the interfaces with key regulators of the GTPase switch Rna1 (GAP), Srm1 (GEF), and Yrb1; Y157 is an allosteric site previously identified to be coupled to the Gsp1 active site,[3] consistent with the proposed effect of mutations on regulated switching; and S155 is a known phosphorylation site[25] neighboring the conserved G5 SAK motif in the active site (Figure 1C). Four of the remaining five toxic/GOF positions are clustered in the Gsp1 structure outside of the active site, and along with the final position (H50) and two other toxic/GOF positions (N156 and F159) form distal interaction networks in crystal structures of Ran/Gsp1 that extend up to 16Å away from the nucleotide ligand to the Switch I and the C-terminal extension in the GDP-bound state[26] (Figure 3A). We verified that toxic/GOF mutations at these positions indeed had severe fitness defects compared to WT or an internal STOP-codon mutant when co-expressed with WT using a yeast spotting assay (Figure 3B), and that a C-terminal deletion variant was as toxic as the toxic/GOF mutations at these positions.

To examine whether toxic/GOF mutations also perturbed switch function in this unexplained set of mutants (as described above for perturbations to regulator interfaces and the C-terminal tail), we purified and characterized pairs of toxic/GOF (F28V, F54A, F159L, and F163L) and WT-like mutants (F28Y, F54W, F159W, and F163Y) at the four Phenylalanine residues that are clustered in the structure but distal from the active site. All purified mutants were well-folded and stable (Figure S5A, B). We then assessed switching by measuring the rate of GEF-mediated nucleotide exchange to either GTP or GDP using recombinantly expressed and purified *S. cerevisiae* Srm1, the GEF of Gsp1 (Figure 3C, Figure S5C–E, and Methods). All mutants except F159L had reduced or similar GEF-catalyzed nucleotide exchange rates compared to WT (Figure S5D). However, the exchange was dependent on the nucleotide: toxic/GOF mutants had a faster rate of exchange to GTP than to GDP while the WT-like counterparts had a preference for GDP over GTP, identical to WT (Figure 3D, Figure S5C–E). Hence, toxic/GOF mutations reversed the nucleotide preference of the switch but WT-like mutations did not. We also measured GAP-catalyzed GTP hydrolysis and found that toxic/GOF mutations did not have reduced GTP hydrolysis (Figure S5F). We conclude that toxic/GOF mutations distal to the active site can indeed allosterically perturb the molecular function of the switch by disfavoring the GDP-bound state, while WT-like mutations at the same positions do not.

An allosteric map of a GTPase switch.

Our analyses assign functional roles to all 60 toxic/GOF positions in our dataset, mapping the functionally essential residues in a GTPase molecular switch (Figure 4A, B). While the active site (nucleotide recognition sites and the GTPase switch loops) is the most common location for toxic/GOF positions, 33% of toxic/GOF positions (20/60) are outside of the active site (Figure 4A). We consider these sites as allosteric given that they are at least 5 and up to 30Å away from the nucleotide (Figure 4C). Overall, our method quantifying perturbations to cellular function in the native network identifies many new sites of allosteric regulation, even surpassing a recent study of allostery quantifying effects on biophysical function in peptide binding domains.[5]

We identify several mechanisms for how perturbations at regions outside of the active site allosterically affect GTPase switching: First, 13 sites are in interaction interfaces with the key regulators Rna1 (GAP) and Srm1 (GEF), which accelerate interconversion between the GTP- and GDP-bound states, and Yrb1, the *S. cerevisiae* homolog of human RanBP1, which stabilizes the GTP-bound state of Gsp1 and increases interaction with the GAP.[23] Second, distal positions in protein-protein interaction interfaces are in addition directly coupled to the switch by modulating the efficiency of GTP hydrolysis.[3] Third, we show here that a previously unknown allosteric cluster in the structure core (Figure 4A, C, red) is coupled to switch regulation by altering the nucleotide preference (Figure 3). Finally, the toxic/GOF positions also include 4 sites of posttranslational modifications (PTMs).[25,27,28] Relatively small perturbations at all identified sites resulted in cellular defects consistently worse than a null mutant, which suggests that the effect on the rates of regulated switching between GTPase states is the key quantitative parameter dominating the functional effects of any Gsp1 mutation.

Comparison to regulatory sites in other GTPases.

While there are no experimental studies probing the function of other GTPases under normal cellular conditions at the residue level, our functional map of Gsp1 is predictive of many activating mutations recently reported for the human H-Ras protein in mouse-derived Ba/F3 cells[29] (Figure 4D). 19/30 positions with activating mutations in H-Ras are also toxic/GOF positions in Gsp1 (Figure 4E). Those positions are enriched in the active site (Figure S6), whereas our Gsp1 perturbation analysis revealed additional allosteric sites including many in regulatory partner interfaces. The additional sites may be specific to Gsp1 or may not be detectable using the overactivation phenotype screened for in the H-Ras assay. Conversely, of the 11/30 activating positions in H-Ras not classified by our stringent cutoff as toxic/GOF in Gsp1, six have at least five toxic/GOF mutations in Gsp1, and all have at least one (Figure S6). Effects on protein folding and stability also add complexity to the fitness defects. The Ras mutational scanning experiments highlighted the role of this interplay between stability and activation for cancer hotspots. In particular, it was shown that mutations at cancer hotspots activate Ras regardless of construct stability, while the effects of mutations at other positions were dependent on stability: mildly destabilizing mutations can be activating due to an increase of nucleotide exchange, resulting in a switch balance perturbed towards the active GTP-bound state, while strongly greater destabilizing mutations can result in loss of cellular function.[29]

We also compared our data to structural and computational analyses of allostery in Ras. Motions of the α4 and α5 helices and loops have been shown to be coupled to those of the active site switch regions in molecular dynamics simulations.[14] Furthermore, a crystal structure of Ras with calcium acetate bound at the α3 and α4 helices and loop 7 shows altered positioning of α3 resulting in ordering of the switch II loop and placement of the catalytic glutamine Q61 closer to the nucleotide. This structure suggested potential allosteric mechanisms underlying calcium dependence of the Ras/Raf/MEK/ERK pathway, although both activation and inactivation in the presence of calcium has been observed.[13] In our assay as well as the Ras cellular activation assay, however, the α3 and α4 helices are largely insensitive to mutation (Figure S7), and in all crystal structures of Gsp1 (both

GTP- or GDP-bound) the α3 helix conformation more closely matches that of H-Ras allosterically activated by calcium acetate, with an ordered switch II. Subtle sequence differences may explain these differences, as also suggested by observations of differential dynamics and allosteric coupling in H-Ras, K-Ras, and N-Ras isoforms.[12,30] The potential variety of structural mechanisms underlying GTPase allosteric activation highlights the need for unbiased, systematic approaches to interrogate allostery.

To extend our analysis beyond Ras, we also compared our data to a computational analysis of GTPases based on residue-residue co-variation in multiple sequence alignments of the GTPase superfamily[31] (Figure 4F). Key "sector" positions identified computationally show more overlap with the Gsp1 toxic/GOF positions than the H-Ras activation data (26/49 of the alignable positions, versus 19/49), again primarily by capturing more residues in the GTPase active site regions (Figure S6). Of the additional 30 positions suggested by the sector analysis, 12 have at least five toxic/GOF mutations in the Gsp1 data, and only four have no toxic/GOF mutations (Figure S6). However, the computational sector analysis misses 23/49 toxic/GOF positions in Gsp1. This finding could indicate a lack of sensitivity or the potential for key regulatory differences between highly conserved GTPases that may be difficult to discern from sequence analysis alone but which are enabled by quantitative perturbations in the native cellular context using our approach.

Discussion

A key finding of our work is the broad sensitivity of a critical molecular switch to perturbations at many allosteric regulatory sites outside the typically studied active site "switch" regions (Figure 4A, C). We propose a model where this sensitivity of the switch facilitates both its responsiveness to many biological inputs and its output signaling specificity.[3] We identify an altered switch balance as the common mechanism by which toxic/GOF mutations affect the cellular function of Gsp1, with the mechanism of each mutation falling into at least one of four classes: (1) perturbation of interaction interfaces with key GTPase cycle regulators,[3,10] (2) allosteric modulation of GTP hydrolysis,[3] (3) allosteric modulation of the relative nucleotide preference to disfavor GDP binding (Figure 3), and (4) alteration of PTM sites.[32] Our findings support a model in which the GTPase switch balance is finely tuned and the sensitivity of this balance to mutations at many residues might explain why GTPases are so highly conserved even outside the active site regions.

We further show that relatively small perturbations to the switch balance of a GTPase can have severely deleterious functional consequences. This finding is consistent with results from kinetic models of ultrasensitivity, where for switches controlled by opposing regulators (Figure 1A) small changes in the concentration or activity of regulators can result in sharp changes in the fraction of the switch "on" state.[33] Our study provides a link between allosteric regulation of the switch balance at the molecular level, the ultrasensitivity of switches,[33] and functional consequences for cellular regulation at the systems level.[3]

The prevalence of allosteric sites in Gsp1 poses interesting questions of structural mechanisms. For example, are allosteric sites at distinct surfaces coupled through common

central residues, such as the "spines" found in kinases.[34] Or, alternatively, are there groups of residues that are coupled within, but not between, distinct networks. These questions could for example be addressed with systematic quantitative mutant cycle analyses using technological advances in high-throughput quantitative enzymology.[35] These analyses may also shed light on the role of subtle sequence differences between GTPase subfamilies of the Ras superfamily, including the distinct roles of residues described in Ras. It will be important to integrate these structural mechanisms with functional roles in different biological contexts. Our residue-level functional map of a GTPase molecular switch and the discovery of new regulatory sites suggests starting points for these analyses and opens avenues to interrogate and target GTPases controlling many essential biological processes including intracellular transport, cell growth, differentiation, and cell survival.

STAR Methods

Resource availability

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tanja Kortemme (tanjakortemme@gmail.com).

Materials availability—Plasmids generated in this study will be deposited to Addgene.

Data and code availability—All sequencing read count data and kinetics data have been deposited on GitHub at https://github.com/cjmathy/Gsp1_DMS_Manuscript and on Zenodo at https://doi.org/10.5281/zenodo.7293738 and are publicly available as of the date of publication. The DOI is listed in the key resources table. All other data reported in this paper are available in the Supplementary Information and Data Tables S1–S3.

All original code for reproducing the EMPIRIC and kinetic data analysis has been deposited on GitHub at https://github.com/cjmathy/Gsp1_DMS_Manuscript and on Zenodo at https://doi.org/10.5281/zenodo.7293738 and is publicly available as of the date of publication. The DOI is listed in the key resources table. The scripts for the Rosetta calculations are available as Supplementary Information Methods S1.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

Microbial strains and media—Gsp1 fitness competition measurements were conducted in the *S. cerevisiae* plasmid swap strain DBY681, generated from the parental strain BY4743 Gsp1∷KanMX. Synthetic dextrose (SD) media lacking amino acids for auxotrophic selection and containing antibiotics were used for library generation and fitness competition. The cells were grown on an orbital shaker at 30°C. For western blots, *S. cerevisiae* cells were grown in either rich (YPD) or synthetic (SD-ura) media at 30°C. For spotting assays, DBY681 *S. cerevisiae* cells were grown in SD-ura at 30°C and OmniMAX *E. coli* cells (Invitrogen, CAT # C854003) were grown in LB at 37°C. Recombinant protein expressions were performed in *E. coli* strain BL21 (DE3). Gsp1 variants were expressed in autoinduction

EZ medium (described below) for 60 hours at 20 °C. GEF and GAP proteins were expressed in 2xYT medium overnight at 25 °C.

Method details

Plasmid and strain construction—To facilitate rapid Fluorescence Activated Cell Sorting (FACS)-based isolation of yeast harboring mutant Gsp1 variants, we generated plasmids marked with GFP or mCherry along with auxotrophic markers. To mimic endogenous expression of Gsp1, we cloned the Gsp1 coding sequence along with its natural promoter sequence (420 bases upstream of the start codon) and its natural 3' region (220 bases downstream from the stop codon). We used centromeric plasmids to approximate genomic copy level. To generate a strong fluorescent signal, we used the Tef1 promoter to drive either GFP or mCherry. We cloned this Gsp1 construct into a URA-marked plasmid with GFP (pRS416Gsp1GFP), and a HIS-marked plasmid with mCherry (pRS413Gsp1mCherry).

We engineered a systematic library including all possible single amino acid changes in Gsp1 as previously described.[36] Briefly, we cloned the Gsp1 open reading frame into pRNDM and created a set of constructs with tiled inverted BsaI restriction sites bracketing 10 amino acid regions of Gsp1. For each amino acid in Gsp1, we used complementary oligonucleotides with single codons randomized as NNN to generate a comprehensive library of variants encoding all possible amino acid changes. We used Gibson assembly to transfer the library into the plasmid swap vector, generating pRS413Gsp1libmCherry. To enable library transfer, this destination vector was modified to harbor a cassette containing an SphI site along with upstream and downstream homologous sequences to Gsp1 promoter and terminator regions respectively. To facilitate short-read estimates of variant frequency we implemented a barcoding strategy as previously described.[36] We used cassette ligation at NotI and AscI restriction sites downstream of Gsp1 gene to introduce an oligonucleotide cassette including an N18 random sequence into the pRS413Gsp1libmCherry variants.

We used paired-end Illumina sequencing to associate the 18 base barcodes with the encoded Gsp1 variants. Amplicons for paired-end sequencing were generated by PCR using a downstream primer that annealed next to the barcode and an upstream primer that annealed either before or within Gsp1. In separate PCR reactions, we used three different upstream primers located immediately before Gsp1, and before nucleotides 200 and 400. The size of each PCR product was less than 1 kb enabling efficient Illumina sequencing. We used asymmetric Illumina paired end sequencing to read 50 bases from the downstream primer and 250 bases from the upstream primer. Reads were organized by barcode sequence and the consensus open reading frame sequence was determined for each barcode.

To generate the plasmid swap strain, DBY681, we started with a heterozygous diploid Gsp1 knockout (BY4743 Gsp1::KanMX) ordered from GE Healthcare (now Horizon Discovery). First, we introduced pRS416Gsp1GFP and selected for transformants on synthetic media lacking uracil. Next, we sporulated the diploid transformants in order to generate haploids bearing the URA-marked plasmid. Successful transformation was evident because the selected haploid yeast cells grew on synthetic media lacking uracil, expressed GFP, grew on G418 antibiotic that selects for endogenous Gsp1 knockout, and lacked growth on synthetic

media having 5-FOA which negatively selects yeast cells with URA-marked plasmid. The resulting DBY681 strain was used for all Gsp1 plasmid swap experiments.

Gsp1 fitness competition—The DBY681 strain was made competent using the lithium acetate method[37] and transformed with the barcoded pRS413Gsp1libmCherry plasmids. Transformation efficiency was determined by plating a small fraction of cells on selection media (SD-Ura-His+G418), aiming for five-fold coverage of the library. Sufficient transformations were performed to introduce each barcoded plasmid variant into more than 10 independent yeast cells. Following transformation, the cells were allowed to recover in synthetic dextrose media lacking uracil (SD-Ura) for ~10 hours at room temperature. The cells were then collected by centrifugation at 5000 × g for 5 minutes, washed multiple times to eliminate residual extracellular plasmid and resuspended in synthetic dextrose media lacking uracil and histidine (SD-Ura-His+G418). Sufficient media was used to achieve an optical density of approximately 0.1 at 600 nm. The cells were grown on an orbital shaker at 30 °C in the double selection media for approximately 42 hours, with constant dilution to maintain the cells in log phase. Experimental replicates were not performed.

A sample of these "initial" cells were retained for sequencing and the remainder were collected by centrifugation at 5000 × g for 10 minutes and resuspended in synthetic dextrose media lacking only histidine (SD-His) to enable loss of the URA-marked WT Gsp1 plasmid. Cells were grown in this medium with orbital shaking at 30 °C for 16 hours, which represents 6 doubling times of the parental DBY681 strain under these conditions. At the end of 16 hrs, cells were collected by centrifugation, then washed and diluted in 1x TBS with 1% BSA. For flow cytometry, the non-fluorescent parental strain W303 was treated as a negative control while DBY681 and W303 transformed with pRS413NoinsertmCherry plasmid were considered as GFP and mCherry positive controls. 3 million cells were analyzed by FACS. Cells that had lost the GFP-marked plasmid encoding WT Gsp1 were isolated by FACS. A total of 500,000 GFP–/mCherry+ cells were isolated by FACS as a sorted sample. The cells were isolated by centrifugation.

Deep sequencing was used to estimate the enrichment or depletion of mutants in the 16 hour sorted sample as compared to the initial sample in double selection media. The initial and sorted yeast samples were lysed using zymolyase and PCR amplified to generate samples for 100 bp Illumina sequencing of barcodes as previously described.[36] Briefly, primers were used that added sequences for identifying each sample as well as for compatibility with Illumina sequencing. Reads with low quality (PHRED score < 20) or that did not match in expected constant regions were eliminated from further analyses. The remaining reads were then parsed into initial and sorted bins and the number of reads of each amino acid mutation in each bin was tabulated. The experimental fitness of each variant was estimated as a selection coefficient based on the counts in the initial and sorted samples using WT synonyms for normalization using the following equation:

$$s_{mut} = \left(\frac{N_{mut,sorted}}{N_{mut,initial}}\right) - mean\left[\left(\frac{N_{WT,sorted}}{N_{WT,initial}}\right)\right]$$

Where s_{mut} is the selection coefficient of a mutant, $N_{mut,sorted}$ is the number of reads of the mutant in the sorted sample, $N_{WT,sorted}$ is the number of reads of WT synonyms in the sorted sample, $N_{mut,initial}$ is the number of reads of the mutant in the initial sample, and $N_{WT,initial}$ is the number of reads of a WT synonym in the initial sample. Using this equation, the average WT synonym has a selection coefficient of 0, while deleterious variants have negative s and beneficial variants have positive s. Alleles with low read counts in the initial sample, defined as less than 2% of the average variant's number of reads, were excluded from all downstream analysis.

Fitness scores were then binned according to thresholds set by the mean and standard deviations of the distributions of scores for WT synonyms and STOP mutants. From the latter distribution we excluded mutations at sequence positions after 174, as these C-terminal STOP mutants showed significant deviations from the relatively consistent distribution of scores for STOP mutants up to and including position 174 (Figure 1E) and correspond to C-terminal deletion mutants that are known to encode fully folded proteins with perturbed biochemical function.[23] Scores within two standard deviations of the mean of the WT synonym score distribution were labeled as WT-like, and scores higher than this cutoff were labeled as beneficial. We note that the majority of mutations labeled by this analysis as "beneficial" had low read counts in the initial library before selection (Figure S2). The number of reads for beneficial mutations spanned 29–508, with only 7/19 beneficial mutations having more than 100 reads. For comparison, the average number of reads for all mutations was 1411. We therefore attributed the "beneficial" designation to be likely due to assay noise. For the STOP mutant distribution, scores within two standard deviations above or three standard deviations below the mean were labeled STOP-like, and scores worse than the bottom cutoff were labeled as toxic/GOF. Finally, scores between the WT-like and STOP-like distributions were labeled as intermediate.

We chose a conservative cutoff for defining toxic/GOF mutations to ensure that these mutations were significantly more toxic than a null mutation, even when considering the inherent noise in the assay (quantified as the variances of the WT-synonymous distribution and STOP (1–174) distributions, which were 0.5 and 0.32, respectively). This cutoff provides a binary toxic/GOF vs. non-toxic/GOF classification, with the non-toxic/GOF mutations further split up into beneficial, WT-like, intermediate, and STOP-like. We proceeded with this binary classification because we were principally interested in a metric to identify the most functionally important residues from the fitness assay for follow-up by biochemical and computational validation. Using a count of the number of toxic/GOF mutations at each position (corrected for a parsimonious null model, as described in the section Statistical modeling of the distribution of toxic/GOF mutations, below) successfully identified the conserved G1–5 regions, suggesting that it was a useful metric for identifying functional residues, in addition to the numerous other positions with many large effect size mutations that we focus on in our study. Other metrics for ranking sequence positions were explored, including mean fitness score, median fitness score, and Pearson or Spearman rank correlation of fitness scores, but ultimately these metrics were discarded as they are more sensitive to both the small number of data points per position (21) and any uncertainty in the individual quantitative scores.

Gsp1 expression levels via western blot—Yeast cells were grown to exponential phase in either rich (YPD) or synthetic (SD-ura) media at 30°C. 10^8 yeast cells were collected by centrifugation and frozen as pellets at −80°C. Cells were lysed by vortexing the thawed pellets with glass beads in lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA and 10 mM PMSF), followed by addition of 2% sodium dodecyl sulfate (SDS). Lysed cells were centrifuged at $18,000 \times g$ for 1 min to remove debris, and the protein concentration of the supernatants was determined using a BCA protein assay kit (CAT #23227, Pierce) compared to a Bovine Serum Albumin (BSA) protein standard. 25 μg of total cellular protein was resolved by SDS-PAGE and was either visualized with Coomassie blue stain, or transferred to a PVDF membrane, and probed using Rabbit anti-RAN primary (CAT # PA1–5783, ThermoFisher Scientific) and Donkey anti-Rabbit HRP-linked secondary (CAT # GENA934, Cytiva Life Science) and visualized with ECL-2 substrate (CAT #80196, Pierce).

Yeast spotting assays—Individual variants of Gsp1 were generated by site-directed mutagenesis using overlapping mutagenic PCR primers and confirmed by Sanger sequencing. Variants were cloned in a HIS-marked plasmid (pRS413 with mCherry). For the yeast spotting assays, the plasmids were transformed into DBY681 (Gsp1::kan, pRS416Gsp1 with GFP) using the lithium acetate method.[37] Transformed cells were recovered in SD-ura media for 6 hours and then 5 μL of a 10x dilution series of cells were spotted onto SD-ura-his plates. For the bacterial spotting assays, the same plasmids were transformed into chemically competent *E. coli*, recovered for 1 hour in LB, and 5 μL of a 10× dilution series of cells were spotted on LB-amp plates.

Statistical modeling of the distribution of toxic/GOF mutations—The null distribution of toxic/GOF scores randomly partitioned among the 219 residue positions was modeled statistically by computing the probability mass function (PMF) of the hypergeometric distribution. This approach computes the probability that a certain number of toxic/GOF scores k would occur at the same position, given n positions in Gsp1, K toxic/GOF scores in the dataset, N total scores in the dataset, and 21 possibilities at each position (20 amino acids and STOP). The PMF of the hypergeometric distribution is given by

$$p_X(k) = Pr(X = k) = C(K, k) * C(N - K, n - k)/C(N, n)$$

where X is the random variable representing the number of toxic/GOF scores at a position and the operator $C(a, b)$ refers to the binomial coefficient *a choose b*. The PMF that was used for defining Gsp1 toxic/GOF positions, shown in Figure S1C, was parameterized as follows:

- $k \in \{0, 1, …, 21\}$, since a given position could have as few as 0 and as many as 21 toxic/GOF scores.

- $n = 219$, the number of sequence positions in Gsp1.

- $K = 1188$, the total number of toxic/GOF scores in the dataset.

- $N = 4519$, the total number of mutations in the dataset after removing mutations with low reads.

The PMF was computed using the *dhyper* function in the *stats* package of the programming language *R*.

Protein purifications—Gsp1 variants were expressed from a pET-28 a (+) vector with an N-terminal 6×His tag in *E. coli* strain BL21 (DE3) in the presence of 50 mg/L Kanamycin in autoinduction EZ medium for 60 hours at 20°C.[38] The autoinduction medium consisted of ZY medium (10 g/L tryptone, 5 g/L yeast extract) supplemented with the following stock mixtures: 20xNPS (1M Na2HPO4, 1M KH2PO4, and 0.5 M (NH4)2SO4), 50× 5052 (25% glycerol, 2.5% glucose, and 10% α-lactose monohydrate), 1000× trace metal mixture (50 mM FeCl3, 20 mM CaCl2, 10 mM each of MnCl2 and ZnSO4, and 2 mM each of $CoCl_2$, $CuCl_2$, $NiCl_2$, Na_2MoO_4, Na_2SeO_3, and H_3BO_3 in ~60 mM HCl). Cells were lysed in 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM imidazole, and 2 mM β-mercaptoethanol using a microfluidizer from Microfluidics. The His-tagged proteins were purified on Ni-NTA resin (Thermo Scientific #88222) and washed into a buffer of 50 mM Tris (pH 7.5), 100 mM NaCl, and 4 mM $MgCl_2$. The N-terminal His-tag was digested at room temperature overnight using 12 NIH Units per mL of bovine thrombin (Sigma-Aldrich T4648–10KU). Proteins were then bound to an additional 1 mL of Ni-NTA resin to remove non-specific binders and passed through a 0.22 uM filter. Purity was confirmed to be at least 90% by SDS polyacrylamide gel electrophoresis. Samples were concentrated on 10 kDa spin filter columns (Amicon Catalog # UFC901024) into a storage buffer of 50 mM Tris pH 7.5, 150 mM NaCl, 4 mM $MgCl_2$, and 1 mM Dithiothreitol. Using this protocol, Gsp1 variants are purified bound to GDP (as any bound GTP is likely hydrolyzed completely during the lengthy incubation steps beginning with thrombin cleavage). The complete hydrolysis to GDP was confirmed for this protocol previously[3] using reverse phase high performance liquid chromatography on a C18 column. Protein concentrations were confirmed by measuring at 10–50× dilution using a Nanodrop (ThermoScientific). The extinction coefficient at 280 nm used for wild-type Gsp1 was 37675 M^{-1} cm^{-1}, based on the value calculated from the primary protein sequence using the ProtParam tool (https://web.expasy.org/protparam/) accounting for the cleaved N-terminal residues, and augmented by 7765 M^{-1} cm^{-1} to account for the bound nucleotide, as described previously (see Note 4.13 in Ref.[39]). Extinction coefficients were calculated for each Gsp1 mutant by the same method. The ratio of absorbance at 260 nm and 280 nm for purified Gsp1 bound to GDP was 0.76 for all mutants except for N156W, for which it was 1.34. Concentrated proteins were flash-frozen and stored at −80 °C.

S. cerevisiae Srm1 (GEF, Uniprot P21827) and *S. pombe* Rna1 (GAP, Uniprot P41391) were also expressed from a pET-28 a (+) vector with a N-terminal 6xHis tag in *E. coli* strain BL21 (DE3). For discussion on the appropriateness of using *S. pombe* GAP for kinetics studies of *S. cerevisiae* Gsp1, see the Supplementary Discussion of Ref.[3]. Srm1 was purified as 1–27Srm1 and GAP as a full-length protein. Sc 1–27Srm1 and SpRna1 were expressed in 2xYT medium (10 g NaCl, 10 g yeast extract (BD BactoTMYeast Extract #212720), 16 g tryptone (Fisher, BP1421) per 1 L of medium) in the presence of 50 mg/L Kanamycin overnight at 25 °C. Expression was induced by addition of 300

μmol/L Isopropyl-β-D-thiogalactoside (IPTG). Cells were lysed in 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, and 2 mM β-mercaptoethanol using a microfluidizer from Microfluidics. The His-tagged proteins were purified on Ni-NTA resin (Thermo Scientific #88222) and washed into a buffer of 50 mM Tris (pH 7.5) and 100 mM NaCl. The N-terminal His-tag was digested at room temperature overnight using 12 NIH Units per mL of bovine thrombin (Sigma-Aldrich T4648–10KU). Proteins were then bound to an additional 1 mL of Ni-NTA resin to remove non-specific binders and passed through a 0.22 uM filter. Proteins were then purified using size exclusion chromatography (HiLoad 26/600 Superdex 200 pg column from GE Healthcare), and purity was confirmed to be at least 90% by SDS polyacrylamide gel electrophoresis. Samples were concentrated on 10 kDa spin filter columns (Amicon Catalog # UFC901024) into storage buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM Dithiothreitol). Protein concentrations were confirmed by measuring at 10–50x dilution using a Nanodrop (ThermoScientific). Extinction coefficients were estimated based on their primary protein sequence using the ProtParam tool (https://web.expasy.org/protparam/). Concentrated proteins were flash-frozen and stored at −80 °C.

Circular dichroism (CD) spectroscopy—Samples for CD analysis were prepared to a concentration of 1 – 2.5 μM Gsp1 in 2.5 mM Tris pH 7.5, 5 mM NaCl, 200 μM MgCl2, and 50 μM Dithiothreitol. CD spectra were recorded at 25 °C using 1- or 2-mm cuvettes (Starna, 21-Q-1 or 21-Q-2) in a JASCO J-710 CD-spectrometer (Serial #9079119). The bandwidth was 2 nm, rate of scanning 20 nm/min, data pitch 0.2 nm, and response time 8 s. Each CD spectrum represents the accumulation of 5 scans. Buffer spectra were subtracted from the sample spectra using the Spectra Manager software Version 1.53.01 from JASCO Corporation. Temperature melts were performed from 25 °C - 95 °C, monitoring at 210 nm, using a data pitch of 0.5°C and a temperature slope of 1°C per minute. As all thermal melts of wild-type and mutant Gsp1 proteins were irreversible, only apparent T_m was estimated by fitting melts to a two-state unfolding equation:

$$fraction\ unbound = (y_f + m_f * T) + (y_u + m_u * T) * \left(\frac{exp\left(m * \left(\frac{1}{T_m} - \frac{1}{T}\right)\right)}{1 + exp\left(m * \left(\frac{1}{T_m} - \frac{1}{T}\right)\right)} \right)$$

with T corresponding to the temperature in degrees Celsius, y_u and y_f corresponding to the molar ellipticity signal at the unfolded and folded states, and m_u, m_f, and m corresponding to the slopes of signal change at the unfolded state, the folded state, and the state transition.

Kinetic measurements of GEF-mediated nucleotide exchange—Kinetic parameters of GEF mediated nucleotide exchange were determined using a fluorescence resonance energy transfer (FRET) based protocol as previously described.[3] Gsp1 variants were purified as a Gsp1:GDP complex, as described above and verified previously.[3] Nucleotide exchange from GDP to either mant-dGDP (3' - O - (N - Methyl - anthraniloyl) - 2' - deoxyguanosine - 5' - diphosphate, CAT # NU-205L, Jena Biosciences) or mant-dGTP (3' - O - (N - Methyl - anthraniloyl) - 2' - deoxyguanosine 5' triphosphate, CAT # NU-212L, Jena Biosciences) was monitored by measuring a decrease in intrinsic Gsp1 tryptophan fluorescence (295 nm excitation, 335 nm detection) due to FRET upon binding of the mant

group. Experiments were performed in 100 μl reaction volumes containing GTPase assay buffer (40 mM HEPES pH 7.5, 100 mM NaCl, 4 mM MgCl$_2$, 1 mM Dithiothreitol) using 5 μM Gsp1, 2.5 nM Srm1 (GEF), and 100 μM mant-labeled nucleotide. Time courses were collected for 20 min at 30°C in a Synergy H1 plate reader from BioTek, using Corning 3686 96-well half-area non-binding surface plates. Initial rates *v0* of nucleotide exchange were estimated using linear fits to the very beginning of reactions for all variants except F28V. Due to the especially slow exchange rate of F28V, the reactions maintained linearity over the entire time course, and so the true exchange rate was estimated by subtracting the rate of background fluorescence decay (obtained from a control without GEF in a separate well on the same plate) from a linear fit of the full time course. At least four replicates were performed for each variant, allowing for calculation of the standard deviation of v_0 values (*sd*). The preference for GTP over GDP was calculated as $p = \frac{v_0^{GTP}}{v_0^{GDP}}$, with the error of preference (*e*) being computed using error propagation over the division operator:

$$e = \left| \frac{v_0^{GTP}}{v_0^{GDP}} \right| * \sqrt{\left(\frac{sd^{GTP}}{v_0^{GTP}} \right)^2 + \left(\frac{sd^{GDP}}{v_0^{GDP}} \right)^2}$$

Finally, the relative change in preference $\frac{p^{MUT}}{p^{WT}}$ was calculated for each mutant, with the error once again propagated across the division operator. All relative changes in preference were computed using WT rates fit on the same day using the same aliquot of GEF, to normalize for any errors in enzyme concentration measurements. Furthermore, experiments for pairs of toxic/GOF and WT-like mutants were always performed on the same day using the same aliquots of GEF.

GTP loading of Gsp1 for GAP-activated hydrolysis assay—WT Gsp1 was loaded with GTP by incubation in the presence of 20-fold excess GTP (Guanosine 5′-Triphosphate, Disodium Salt, Calbiochem CAT # 371701) in 50 mM Tris HCl pH 7.5, 100 mM NaCl, 4 mM MgCl$_2$. Exchange of GDP for GTP was initiated by the addition of 10 mM EDTA. Reactions were incubated for 3 hours at 4°C and stopped by addition of 1 M MgCl$_2$ to a final concentration of 20 mM MgCl$_2$ to quench the EDTA. GTP-loaded protein was buffer exchanged into a GTPase assay buffer of 40 mM HEPES pH 7.5, 100 mM NaCl, 4 mM MgCl$_2$, 1 mM DTT using NAP-5 Sephadex G-25 DNA Grade columns (GE Healthcare # 17085301).

Kinetic measurements of GAP-activated GTP hydrolysis—Kinetic parameters of the GTP hydrolysis reaction were determined as previously described.[3] Gsp1 samples for GTP hydrolysis kinetic assays were first loaded with GTP as described above. GTP hydrolysis was monitored by measuring fluorescence of the *E. coli* phosphate-binding protein labeled with 7 - Diethylamino - 3 - [N - (2 - maleimidoethyl) carbamoyl] coumarin (MDCC) (phosphate sensor, CAT # PV4406, ThermoFisher) upon binding of the free phosphate GTP hydrolysis product (excitation at 425 nm, emission at 457 nm). Experiments were performed in 100 μl GTPase assay buffer (40 mM HEPES pH 7.5, 100 mM NaCl, 4 mM MgCl$_2$, 1 mM Dithiothreitol) using 5 μM Gsp1:GTP, 1 nM *Sp*Rna1 (GAP), and 20

µM phosphate sensor. Time courses were collected for 60 min at 30°C in a Synergy H1 plate reader from BioTek, using Corning 3881 96-well half-area clear-bottom non-binding surface plates. A conversion factor between fluorescence and phosphate concentration was calibrated for the 20 µM concentration of the sensor with a range of concentrations of K_2HPO_4, considering only data in the linear range. For each individual GAP-mediated GTP hydrolysis experiment, a control experiment with the same concentration of GTP-loaded Gsp1 and the same concentration of sensor, but without added GAP, was run in parallel. The first 100 s of these data were used to determine the baseline fluorescence. The kinetic parameters (k_{cat} and K_m) were estimated by directly analyzing the full reaction progress curve with an analytical solution of the integrated Michaelis-Menten equation, as done previously[3] using the custom-made software DELA.[40] Specifically, each time course was fitted to an integrated Michaelis Menten equation:

$$fluorescence = B + [E] * \left(C_i + (C_f - C_i) * \left(1 - K_m * \frac{\omega}{[S]_0} \right) \right)$$

Where $[E]$ is the total enzyme (GAP) concentration, C_i is the initial fluorescence, C_f is the final fluorescence, $[S]_0$ is the initial concentration of the substrate (Gsp1:GTP), and B is the baseline slope in fluorescence per second. Exact concentration of loaded Gsp1:GTP $[S]_0$ was estimated based on the plateau fluorescence and the sensor calibration parameters to convert the fluorescence to free phosphate concentration. The ω parameter was solved by using the Lambert ω algorithm,

$$\omega = Lambert\ omega \left(\frac{[S]_0}{K_m} * exp \left(\frac{[S]_0 - k_{cat} * [E] * time}{K_m} \right) \right).$$

Structural bioinformatics—Protein structures were downloaded from the PDB-REDO databank web server.[41] Secondary structure annotation of the GTP-bound (PDB 3M1I, chain A) and GDP-bound (PDB 3GJ0) states were performed using PyMOL (Schrödinger, Inc.) with the command *ss H/S*, followed by manual inspection and comparison to the results of the DSSP algorithm[42] implemented in the PyRosetta interface (version *2020.28+release.8ecab77aa50*) to the Rosetta molecular modeling suite.[43]

Assignments of structural regions (structure core, interface core, and surface) of Gsp1 were previously reported[3] whereby burial of a residue (in either the structure core or interface core) was defined based on per-residue relative solvent accessible surface area (rASA)[44] compared to the empirical maximum solvent accessible surface area for each of the 20 amino acids.[45] Annotations of the canonical Ras superfamily GTPase regions were taken from Ref.[46] as well as studies of Ran structures.[47–50] The key GEF binding region annotations were taken from Ref.[51].

Rosetta ΔΔG calculations—Stability calculations for all 19 possible point mutations were performed using the application *cartesian-ddg*[52,53] in the Rosetta software suite. Calculations were performed for both the GTP-bound (PDB 3M1I, chain A) and GDP-bound (PDB 3GJ0) structures. First, the structures were minimized in cartesian coordinates using the *relax* application, the *ref2015_cart* score function, and constraints to starting

coordinates. The *relax* protocol was run 20 times and the lowest scoring structure was chosen. The GTP-bound structure was truncated after position 183, as the C-terminal extension contains unresolved regions in this crystal structure and adopts a different conformation when bound to Yrb1. The prepared starting structures were then run through the *cartesian-ddg* protocol, which computes energy scores in Rosetta Energy Units (REU) for each mutation by choosing the best scoring rotamer for the mutant amino acid, then minimizing the structure 5 times in cartesian coordinates while only allowing movement of sidechain atoms within a 6Å window around the mutated residue and backbone atoms within a three residue window (1 neighboring residue on each side), and finally taking the average score of the 5 structures. ∆G scores are computed by performing the same protocol at each site while choosing the best WT rotamer at the first step, and then taking the difference between the mutant and WT energies. Finally, the ∆∆G values were scaled down using a scaling factor of 0.298, determined from a benchmark set of stability calculations (performed in parallel with the Gsp1 calculations) for which experimental ∆∆G values are available,[54,55] as recommended by the authors of the *cartesian-ddg* protocol.[53] Position Q71 was excluded from the analysis, as the GTP-bound structure harbors a Q71L mutation at the catalytic glutamine. The full set of command line flags for the *relax* and *cartesian-ddg* protocols are provided in the Supplemental File under Methods S1. The *movemap* file *gsp1.movemap* was not included for *relax* runs on the benchmark set. All associated configuration files as well as the datasets of Gsp1 and benchmark set ∆∆G values are available in full at the code repository at https://github.com/cjmathy/Gsp1_DMS_Manuscript and at https://doi.org/10.5281/zenodo.7293738.

Comparison to H-Ras mutagenesis data—Alignment of sequence positions between Gsp1 and H-Ras was performed with the *bio3d* package[56] using the function *pdbaln* followed by refinement of the alignment upon inspection of the structural superposition using the function *pdbfit*. PDB structures used for the superposition were 3M1I, 1K5D, 1WQ1, and 3L8Z. The sequence alignment is shown in Figure S6A. In total, 156 structurally aligned positions were included in the analysis. Fitness scores from the human H-Ras mutagenesis study[29] were obtained from datasets deposited on GitHub at https://github.com/fhidalgor/ras_cancer_hidalgoetal (commit *0dcb01b* from Dec. 22, 2021, downloaded on January 31, 2022). Receiver operating characteristic (ROC) curves were produced as described in the original study,[29] namely by considering H-Ras mutations with a fitness score greater than 1.5 times the standard deviation in the Ba/F3 dataset as activating (true positives), with the other mutations labeled as true negatives. Then, a variable threshold value of Gsp1 fitness is used, and for each threshold value, mutations with a Gsp1 fitness score less than that threshold (starting with the most deleterious mutations and proceeding to decreasingly deleterious Gsp1 mutations) are considered to predict H-Ras activation.

For the analysis of overlap with Gsp1 toxic/GOF positions (Figures 4E, Figure S6), a threshold of 2 or more activating mutations at a position was chosen for defining H-Ras activation positions, since a large number of positions have only a single activating mutation out of the 21 possible mutations. This threshold was supported by a chi-squared test evaluating the strength of association between the Gsp1 toxic/GOF and H-Ras activating

sets when applying the threshold (P = 7.711 e-5) vs. including all positions with one or more activating mutation (P = 0.0411).

Comparison to Statistical Coupling Analysis—H-Ras sector positions identified by statistical coupling analysis[57] were taken from an analysis notebook document by the Ranganathan group publicly available on their Github (https://github.com/ranganathanlab/pySCA/blob/master/notebooks/SCA_G.ipynb, commit *301f874*, downloaded on February 9, 2022) prepared in concert with their study.[31]

Quantification and Statistical Analysis

Quantification and statistical analysis were performed using custom code in Python and R, available on GitHub at https://github.com/cjmathy/Gsp1_DMS_Manuscript and on Zenodo at https://doi.org/10.5281/zenodo.7293738. Details of the statistical analysis can be found in the figures, figure legends, the results section of the text, the Method details section of the STAR Methods, and the Supplemental information. Statistical test and number of samples are indicated whenever appropriate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Motlagh HN, Wrabl JO, Li J, and Hilser VJ (2014). The ensemble nature of allostery. Nature 508, 331–339. [PubMed: 24740064]

2. Piazza I, Kochanowski K, Cappelletti V, Fuhrer T, Noor E, Sauer U, and Picotti P (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell 172, 358–372.e23. [PubMed: 29307493]

3. Perica T, Mathy CJP, Xu J, Jang GM, Zhang Y, Kaake R, Ollikainen N, Braberg H, Swaney DL, Lambright DG, et al. (2021). Systems-level effects of allosteric perturbations to a model molecular switch. Nature 599, 152–157. [PubMed: 34646016]

4. Dokholyan NV (2016). Controlling Allosteric Networks in Proteins. Chem. Rev 116, 6463–6487. [PubMed: 26894745]

5. Faure AJ, Domingo J, Schmiedel JM, Hidalgo-Carcedo C, Diss G, and Lehner B (2022). Mapping the energetic and allosteric landscapes of protein binding domains. Nature 604, 175–183. [PubMed: 35388192]

6. Kuriyan J, and Eisenberg D (2007). The origin of protein interactions and allostery in colocalization. Nature 450, 983–990. [PubMed: 18075577]

7. Kholodenko BN (2006). Cell-signalling dynamics in time and space. Nat. Rev. Mol. Cell Biol 7, 165–176. [PubMed: 16482094]

8. Cherfils J, and Zeghouf M (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. Physiol. Rev 93, 269–309. [PubMed: 23303910]

9. Vetter IR, and Wittinghofer A (2001). The guanine nucleotide-binding switch in three dimensions. Science 294, 1299–1304. [PubMed: 11701921]

10. Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, Gee CL, Chakraborty AK, Kortemme T, Ranganathan R, et al. (2017). Deconstruction of the Ras switching cycle through saturation mutagenesis. Elife 6, 727.

11. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, and Alber T (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl. Acad. Sci. U. S. A 108, 16247–16252. [PubMed: 21918110]

12. Volmar AY, Guterres H, Zhou H, Reid D, Pavlopoulos S, Makowski L, and Mattos C (2022). Mechanisms of isoform-specific residue influence on GTP-bound HRas, KRas, and NRas. Biophys. J 10.1016/j.bpj.2022.07.005.

13. Buhrman G, Holzapfel G, Fetics S, and Mattos C (2010). Allosteric modulation of Ras positions Q61 for a direct role in catalysis. Proc. Natl. Acad. Sci. U. S. A 107, 4931–4936. [PubMed: 20194776]

14. Gorfe AA, Hanzal-Bayer M, Abankwa D, Hancock JF, and McCammon JA (2007). Structure and dynamics of the full-length lipid-modified H-Ras protein in a 1,2dimyristoylglycero-3-phosphocholine bilayer. J. Med. Chem 50, 674–684. [PubMed: 17263520]

15. Ostrem JM, Peters U, Sos ML, Wells JA, and Shokat KM (2013). K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 503, 548–551. [PubMed: 24256730]

16. Eisenberg D, Marcotte EM, Xenarios I, and Yeates TO (2000). Protein function in the post-genomic era. Nature 405, 823–826. [PubMed: 10866208]

17. Dunham AS, and Beltrao P (2021). Exploring amino acid functions in a deep mutational landscape. Mol. Syst. Biol 17, e10305. [PubMed: 34292650]

18. Fowler DM, and Fields S (2014). Deep mutational scanning: a new style of protein science. Nat. Methods 11, 801–807. [PubMed: 25075907]

19. Stiffler MA, Hekstra DR, and Ranganathan R (2015). Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell 160, 882–892. [PubMed: 25723163]

20. Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, and Bolon DNA (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J. Mol. Biol 425, 1363–1377. [PubMed: 23376099]

21. Hietpas RT, Bank C, Jensen JD, and Bolon DNA (2013). Shifting fitness landscapes in response to altered environments. Evolution 67, 3512–3522. [PubMed: 24299404]

22. Hietpas R, Roscoe B, Jiang L, and Bolon DNA (2012). Fitness analyses of all possible point mutations for regions of genes in yeast. Nat. Protoc 7, 1382–1396. [PubMed: 22722372]

23. Richards SA, Lounsbury KM, and Macara IG (1995). The C Terminus of the Nuclear RAN/TC4 GTPase Stabilizes the GDP-bound State and Mediates Interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. J. Biol. Chem 270, 14405–14411. [PubMed: 7782302]

24. Zhou J, Tan Y, Zhang Y, Tong A, Shen X, Sun X, Jia D, and Sun Q (2020). GEF-independent Ran activation shifts a fraction of the protein to the cytoplasm and promotes cell proliferation. Molecular Biomedicine 1, 18. [PubMed: 35006421]

25. Holt LJ, Tuch BB, Villén J, Johnson AD, Gygi SP, and Morgan DO (2009). Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science 325, 1682–1686. [PubMed: 19779198]

26. Neuwald AF, Kannan N, Poleksic A, Hata N, and Liu JS (2003). 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. 13, 673–692. [PubMed: 12671004]

27. Swaney DL, Beltrao P, Starita L, Guo A, Rush J, Fields S, Krogan NJ, and Villén J (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat. Methods 10, 676–682. [PubMed: 23749301]

28. Weinert BT, Schölz C, Wagner SA, Iesmantavicius V, Su D, Daniel JA, and Choudhary C (2013). Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep. 4, 842–851. [PubMed: 23954790]

29. Hidalgo F, Nocka LM, Shah NH, Gorday K, Latorraca NR, Bandaru P, Templeton S, Lee D, Karandur D, Pelton JG, et al. (2022). A saturation-mutagenesis analysis of the interplay between stability and activation in Ras. Elife 11. 10.7554/eLife.76595.

30. Gorfe AA, Grant BJ, and McCammon JA (2008). Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins. Structure 16, 885–896. [PubMed: 18547521]

31. Rivoire O, Reynolds KA, and Ranganathan R (2016). Evolution-Based Functional Decomposition of Proteins. PLoS Comput. Biol 12, e1004817. [PubMed: 27254668]

32. de Boor S, Knyphausen P, Kuhlmann N, Wroblowski S, Brenig J, Scislowski L, Baldus L, Nolte H, Krüger M, and Lammers M (2015). Small GTP-binding protein Ran is regulated by posttranslational lysine acetylation. Proc. Natl. Acad. Sci. U. S. A 112, E3679–88. [PubMed: 26124124]

33. Goldbeter A, and Koshland DE Jr (1981). An amplified sensitivity arising from covalent modification in biological systems. Proc. Natl. Acad. Sci. U. S. A 78, 6840–6844. [PubMed: 6947258]

34. Kornev AP, Haste NM, Taylor SS, and Eyck LFT (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc. Natl. Acad. Sci. U. S. A 103, 17783–17788. [PubMed: 17095602]

35. Markin CJ, Mokhtari DA, Sunden F, Appel MJ, Akiva E, Longwell SA, Sabatti C, Herschlag D, and Fordyce PM (2021). Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. Science 373. 10.1126/science.abf8761.

36. Flynn JM, Rossouw A, Cote-Hammarlof P, Fragata I, Mavor D, Hollins C 3rd, Bank C, and Bolon DN (2020). Comprehensive fitness maps of Hsp90 show widespread environmental dependence. Elife 9. 10.7554/eLife.53810.

37. Gietz RD, and Schiestl RH (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc 2, 31–34. [PubMed: 17401334]

38. Studier FW (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif 41, 207–234. [PubMed: 15915565]

39. Smith SJM, and Rittinger K (2002). Preparation of GTPases for structural and biophysical analysis. Methods Mol. Biol 189, 13–24. [PubMed: 12094582]

40. Malaby AW, Chakravarthy S, Irving TC, Kathuria SV, Bilsel O, and Lambright DG (2015). Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. J. Appl. Crystallogr 48, 1102–1113. [PubMed: 26306089]

41. Joosten RP, Long F, Murshudov GN, and Perrakis A (2014). The PDB_REDO server for macromolecular structure model optimization. IUCrJ 1, 213–220.

42. Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, and Vriend G (2015). A series of PDB-related databanks for everyday needs. Nucleic Acids Research 43, D364–D368. 10.1093/nar/gku1028. [PubMed: 25352545]

43. Chaudhury S, Lyskov S, and Gray JJ (2010). PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26, 689–691. [PubMed: 20061306]

44. Levy ED (2010). A simple definition of structural regions in proteins and its use in analyzing interface evolution. J. Mol. Biol 403, 660–670. [PubMed: 20868694]

45. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, and Wilke CO (2013). Maximum allowed solvent accessibilites of residues in proteins. PLoS One 8, e80635. [PubMed: 24278298]

46. Vetter IR (2014). The Structure of the G Domain of the Ras Superfamily. In Ras Superfamily Small G Proteins: Biology and Mechanisms 1: General Features, Signaling, Wittinghofer A, ed. (Springer Vienna), pp. 25–50.

47. Chook YM, and Blobel G (1999). Structure of the nuclear transport complex karyopherin-beta2-Ran × GppNHp. Nature 399, 230–237. [PubMed: 10353245]

48. Scheffzek K, Klebe C, Fritz-Wolf K, Kabsch W, and Wittinghofer A (1995). Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. Nature 374, 378–381. [PubMed: 7885480]

49. Vetter IR, Arndt A, Kutay U, Görlich D, and Wittinghofer A (1999). Structural view of the Ran-Importin beta interaction at 2.3 A resolution. Cell 97, 635–646. [PubMed: 10367892]

50. Vetter IR, Nowak C, Nishimoto T, Kuhlmann J, and Wittinghofer A (1999). Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. Nature 398, 39–46. [PubMed: 10078529]

51. Renault L, Kuhlmann J, Henkel A, and Wittinghofer A (2001). Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). Cell 105, 245–255. [PubMed: 11336674]

52. Frenz B, Lewis SM, King I, DiMaio F, Park H, and Song Y (2020). Prediction of Protein Mutational Free Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. Front Bioeng Biotechnol 8, 558247. [PubMed: 33134287]

53. Park H, Bradley P, Greisen P Jr, Liu Y, Mulligan VK, Kim DE, Baker D, and DiMaio F (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J. Chem. Theory Comput 12, 6201–6212. [PubMed: 27766851]

54. Kellogg EH, Leaver-Fay A, and Baker D (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 79, 830–838. [PubMed: 21287615]

55. Ó Conchúir S, Barlow KA, Pache RA, Ollikainen N, Kundert K, O'Meara MJ, Smith CA, and Kortemme T (2015). A Web Resource for Standardized Benchmark Datasets, Metrics, and Rosetta Protocols for Macromolecular Modeling and Design. PLoS One 10, e0130433. [PubMed: 26335248]

56. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, and Caves LSD (2006). Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22, 2695–2696. [PubMed: 16940322]

57. Lockless SW, and Ranganathan R (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286, 295–299. [PubMed: 10514373]

Highlights:

- Comprehensive mutation of the conserved GTPase switch Gsp1/Ran in its native network

- Allosteric sites are prevalent and widely distributed throughout the GTPase structure

- Modulation of GTPase switch kinetics is a shared mechanism of diverse regulatory sites

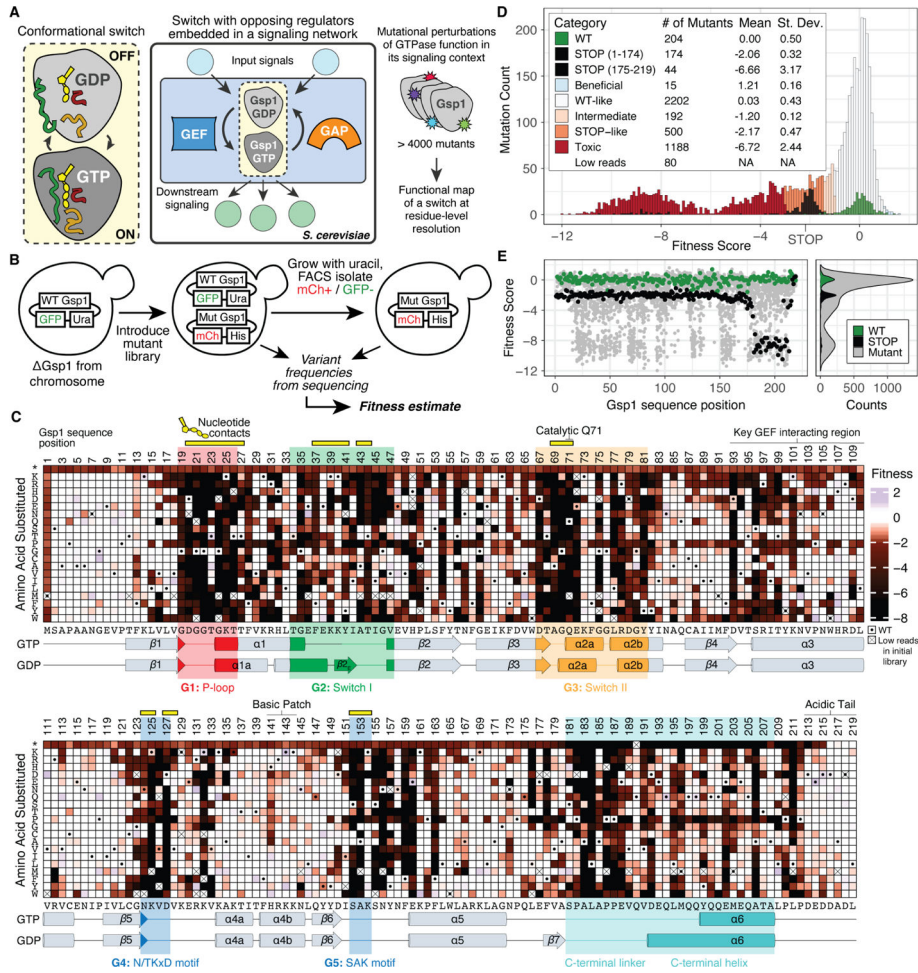- Comparative analysis reveals conserved and subfamily-specific regulatory sites

Figure 1. The cellular function of the GTPase Gsp1 is broadly sensitive to mutational perturbations.

(**A**) All possible single amino acid point mutations are used to exhaustively probe a switch in its native network. (**B**) Generalizable plasmid swap approach to probe essential genes by mutational mapping. (**C**) Heatmap showing quantitative fitness scores ($\log_2$-transformed changes in variant abundance relative to wild-type) for all Gsp1 mutations. Dot indicates WT synonymous codons; X indicates mutants with low reads in the initial library outgrowth. Conserved G1–5 regions are shown in colors corresponding to structural annotations. see Figure S3. Additional annotated functional regions include the catalytic residue Q71, the GEF interacting region, and the basic patch and acidic tail that interact in the GDP-bound structure (Neuwald et al., 2003). Positions of residues contacting the nucleotide or magnesium cofactor are indicated by yellow bars. Secondary structure assignments for each position in the GTP- and GDP-bound states are shown below. (**D**) Histogram of scores colored by bin (Methods). Note that 37 of the STOP mutants are toxic/GOF. (**E**) Distribution of fitness scores ordered by Gsp1 sequence position, colored by mutation type: WT synonymous mutations (green), STOP codon mutations (black), and substitutions (gray).
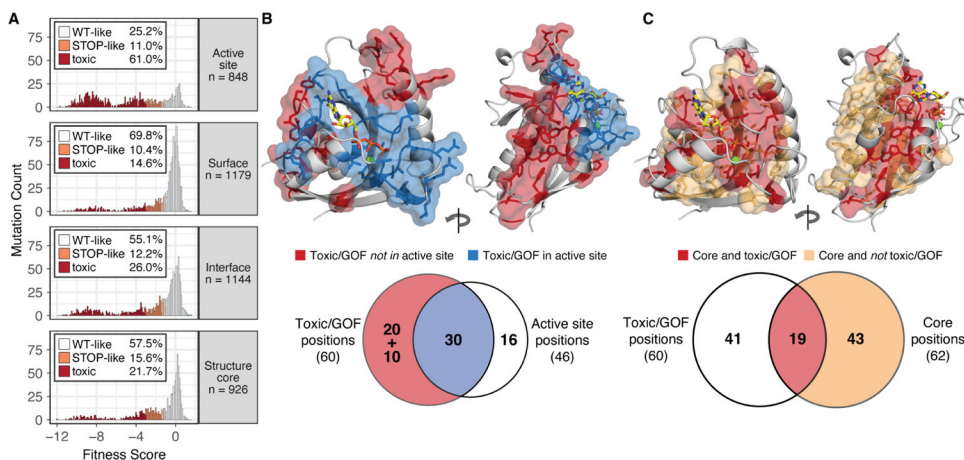
Figure 2. Structural locations of toxic/GOF positions are not exclusive to the active site or the protein core.

(A) Histograms of fitness scores of mutants by structural regions; colors are as in Figure 1D (showing only point mutations, excluding changes that are WT synonymous or to/from STOP; intermediate and beneficial mutations make up the difference to 100%). Fractions are computed within each structural region; n indicates number of mutations. **(B)** and **(C)** Two views rotated by 180 degrees of the Gsp1-GTP structure (PDB ID: 3M1I) showing sidechains of toxic/GOF positions in stick and surface representation (excluding the C-terminal extension). **(B)** Toxic/GOF positions in the GTPase active site shown in blue, other toxic/GOF positions shown in red. Venn diagram below shows overlap of toxic/GOF positions with GTPase active site positions (10 toxic/GOF positions not shown in the structure are in the C-terminal extension). **(C)** Toxic/GOF core positions shown in red, non-toxic/GOF core positions shown in orange. Venn diagram below shows overlap of toxic/GOF positions with structure core positions.
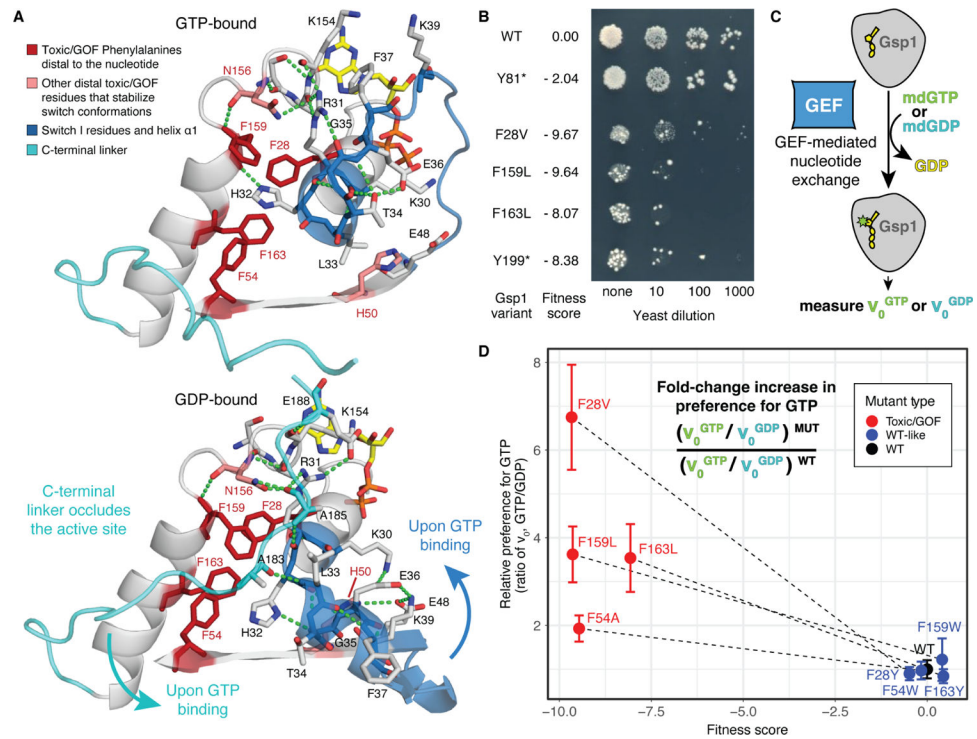
Figure 3. Distal toxic/GOF mutations allosterically alter the balance of the switch states.
(A) Structural depiction of extended networks of interactions in the GTP-bound (top, PDB ID: 3M1I) and GDP-bound states (bottom, PDB ID: 3GJ0). Toxic/GOF mutants characterized in (B) and (D) shown in red. Backbone is colored for the Switch I region (blue) and the C-terminal linker (cyan). The nucleotides are shown in yellow sticks. **(B)** Plate growth assay showing a dilution series of individual Gsp1 variants expressed together with WT in *S. cerevisiae*, with corresponding fitness scores from the EMPIRIC assay. **(C)** FRET-based nucleotide exchange kinetics are measured by adding an excess of mant-labeled fluorescent nucleotide and catalytic amounts of GEF to purified Gsp1 bound to GDP (Methods). **(D)** Relative change in nucleotide preference for pairs of toxic and wild-type like variants at the Phe residues highlighted in (A), calculated as the ratio of initial rate of exchange to GTP divided by the initial rate of exchange to GDP, normalized to the wild-type ratio. At least four replicates were performed for each variant. Error bars represent the standard deviations of v0 measurements propagated across the division operator (Methods).
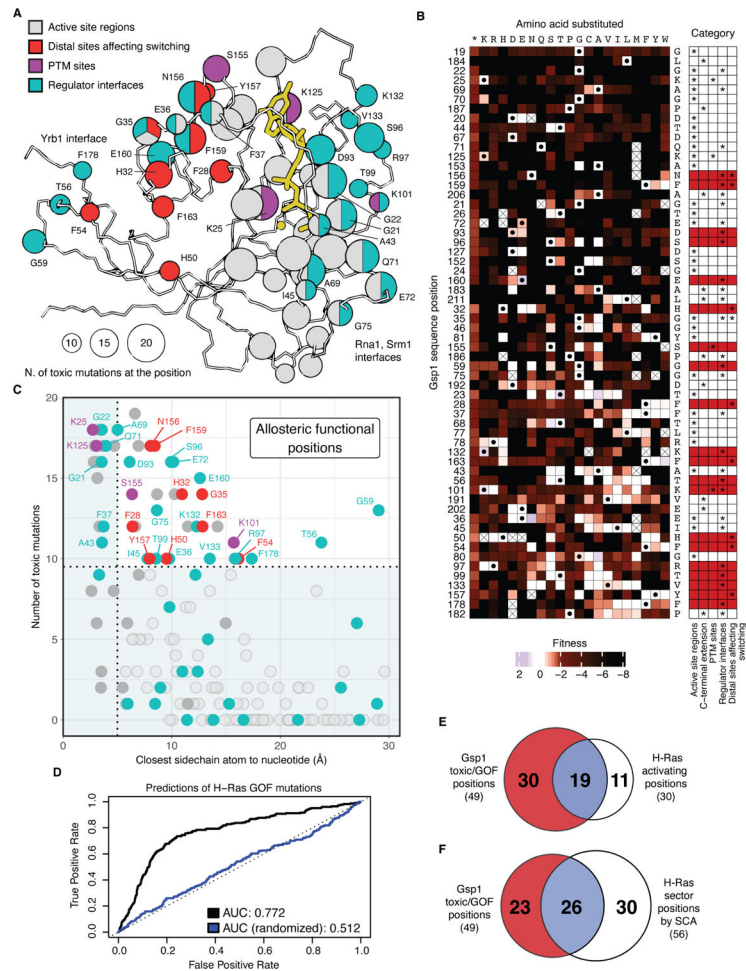
Figure 4. Allosteric map of the Gsp1 GTPase switch.
(A) Wire representation of Gsp1-GTP (PDB ID: 3M1I, positions 1–180). Toxic/GOF positions are shown in sphere representation. Sphere radius represents number of toxic/GOF mutations at each position. Spheres are colored by functional categories, see (B). The nucleotide and Mg^{2+}cofactor are shown in yellow. **(B)** Heatmap showing fitness scores (log$_2$-transformed changes in variant abundance relative to wild-type) at toxic/GOF positions ordered by number of toxic/GOF mutations. WT amino acid residue shown below each column. Functional annotations (stars) are shown below and marked in red for positions outside of the active site. **(C)** Distance of closest sidechain heavy atom at each position to the nucleotide (GTP). Colors are as in (A). Residues not belonging to one of the four categories of functional annotation are indicated by an open circle. **(D)** Receiver operating characteristic (ROC) curves and area under the curve (AUC) showing the statistical power of Gsp1 fitness scores in classifying an *H. sapiens* H-Ras mutant as activating, as defined by Ref.[29]. Datasets were trimmed to the 156 sequence positions alignable for Gsp1 and H-Ras (Figure S6). **(E)** and **(F)**, Overlap of functional sites defined as Gsp1 toxic/GOF and either (E) H-Ras activating or (F) comprising an H-Ras sector defined by statistical coupling analysis (SCA)[31] (Table S1).

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
RAN Polyclonal Antibody	ThermoFisher	CAT# PA1-5783; RRID:AB1087016
Donkey anti-Rabbit HRP-linked Antibody	Cytiva Life Science	CAT# GENA934, RRID:AB2722659
Bacterial and Virus Strains		
BL21 (DE3)	ThermoFisher	CAT# EC0114
Chemicals, Peptides, and Recombinant Proteins		
ECL 2 Western Blotting Substrate	Thermo Scientific	CAT# PI80196
Thrombin from bovine plasma	Sigma-Aldrich	CAT# T4648-10KU
mant-dGDP	Jena Biosciences	CAT# NU-205L
mant-dGTP	Jena Biosciences	CAT# NU-212L
Guanosine 5'-Triphosphate, Disodium Salt	Calbiochem	CAT# 371701
Phosphate Sensor	ThermoFisher	CAT# PV4406
Critical Commercial Assays		
BCA protein assay kit	Thermo Scientific	CAT #23227
Deposited Data		
GitHub repository deposit of raw data and code, archived on Zenodo	This paper	https://doi.org/10.5281/zenodo.7293738
H-Ras fitness scores	Hidalgo et al., 2022	https://doi.org/10.5281/zenodo.6131510
Small GTPase sector positions	Rivoire et al., 2016	PMID: 27254668; https://github.com/reynoldsk/pySCA; https://doi.org/10.1371/journal.pcbi.1004817.s003
Experimental Models: Organisms/Strains		
S. cerevisiae: Gsp1 knockout clone: *BY4743 Gsp1::KanMX*	Horizon Discovery (formerly GE Healthcare)	CAT# YSC1021
S. cerevisiae: DBY681: *BY4743 Gsp1::KanMX-pRS416Gsp1GFP*	This paper	
Recombinant DNA		
pRS416Gsp1GFP	This paper	
pRS413Gsp lmCherry	This paper	
pRS413Gsp IlibmCherry	This paper	
pET28a_6xHis-Gsp1_WT	This paper	
pET28a_6xHis-ScSrm1_1-27	This paper	
pET28a_6xHis-SpRna1	This paper	
Software and Algorithms		
R v4.0.3	R Project for Statistical Computing	https://www.r-project.org

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Python v3.8.4	Python Software Foundation	https://www.python.org
ProtParam	Expasy, Swiss Bioinformatics Resource Portal	https://web.expasy.org/protparam/
Spectra Manager v1.53.01	Jasco	https://jascoinc.com/products/spectroscopy/circular-dichroism/software/spectra-manager/
Gen5 Microplate Reader and Imager Software v3.03	Biotek (now Agilent)	https://www.biotek.com/products/software-robotics-software/gen5-microplate-reader-and-imager-software/
Data Evaluation and Likelihood Analysis (DELA) v1.0	Malaby et al., 2015	PMID: 26306089
PyMOL v2.4.0	Schrodinger, Inc	https://pymol.org/
DSSP (Define Secondary Structure of Proteins) algorithm	Kabsch and Sander, 1983	PMID: 6667333
Rosetta software suite v2020.27.post.dev+93.master.8ecab778ecab77aa50ac1301efe53641e07e09ac91fee3b	Leaver-Fay et al., 2011	PMID: 21187238; https://www.rosettacommons.org/software
cartesian-ddg Rosetta application	Park et al., 2016	PMID: 27766851
PyRosetta v2020.28+release.8ecab77aa50	Chaudhury et al., 2010	PMID: 20061306