Systems biology of allosteric GTPase mutations

by

Christopher Mathy

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
Tanja Kortemme
74ABEB009F65402...

Tanja Kortemme

Chair

DocuSigned by:
Martin Kampmann
DocuSigned by:4AF...

Martin Kampmann

DocuSigned by:
John Dueber
DocuSigned by:483...

John Dueber

DocuSigned by:
Sourav Bandyopadhyay
F1E602E27884493...

Sourav Bandyopadhyay

Committee Members

Dedication

This work is dedicated to my grandparents, Dédée and Maurice. While they always asked after my studies when we phoned, during visits my grandmother would go a step further, grabbing my hand to gently insist, « *Il faut que tu travailles bien, pour que tu puisses avoir une belle situation, après que je suis partie* ». May she rest in peace, while I keep those words and her memory close.

This work must also be dedicated to my niece Vera and my nephews Pascal, Leo, and Remy. In the first year of my PhD, I was chatting with someone who prompted me to close my eyes and visualize what was most important to me, as a mindfulness exercise. Instantly, I saw Vera, at the time the sole trailblazer of our family's next generation. My heart belongs to you all, in hopes of repaying all of the joy you have brought into our lives. Any good I manage to do here on this planet is ultimately for you and your futures.

Acknowledgments

I have overwhelming gratitude to the friends, family, and colleagues who made my time in graduate school one of not only scholarly, but also great personal growth. My roots grew deep in San Francisco and the Bay Area because of the bonds we forged together, and the care with which these folks have treated me.

Tanja, my thesis advisor and mentor, guided me with enthusiasm, wisdom, and patience toward a simple goal: doing good science. She has been a true role model, showing how collaboration, respect, and kindness in addition to scientific curiosity and rigor form the basis of a fulfilling scholarly career.

Tina was my chum, mentor, and Gsp1 co-conspirator without whose skill and ingenuity as a scientist none of this work could have been done. We had countless laughs working together, and I learned from her each and every day.

Cody took on so many roles in my grad school life, as classmate, labmate, housemate, and, most importantly, enthusiastic participant in countless goofs and bits.

Anum is one of the most devoted scientists and humans I know, and I'm grateful that I could always count on her to give a shit about the right things.

Numerous faculty generously mentored me during my research, including Mark Kelly who took me under his wing to teach me NMR, Martin Kampmann, John Dueber, and Sourav Bandyopadhyay who provided thoughtful advice as members of my thesis committee, and Nevan Krogan and Dan Bolon who were supportive and genial collaborators.

My labmates, including Sunny, Hilary, Kale, Kyle, Amanda, Samuel, XingJie, James, Simon, Ziyue, Amy, Connor, Rob, Lu, Ben J, Dom, Ben O, and Deniz, kept things lively and made it a joy to come to work.

Many friends joined me on the afternoon walks, dinner parties, soccer games, and weekend hikes that form so many happy memories and kept me grounded during the ups and downs of grad school: Ivana, Jon, Annie, Ben D, Cody, Laurel, Karen, Jordi, Emily, Nate, Zeynep, Jordan, and Andoni from the Bioengineering program; Snow, Matvei, Yessica, Kelly, and Maru from UCSF; Sarah J, Dev, Sean M, Daniel, Josh, Marc, Maria, Caitlin, Priya, Cristina, Zach C, and Marco from Stanford; Evan, Sean T, Erin, Daria, Leo, Monica, Sarah K, and Russell from 59 Carr and the Harmonics; and folks from near and far like Eleni, Romain, and Zach KG. Ashley and Jason deserve special thanks, for they have become friends so dear to me that it's hard to imagine what the past decade would have looked like, or who I would be, without them around.

Finally, my family have given me the most precious gift, a sense of home, when I am with them in Illinois, California, Maryland, or France, and their love has shaped me into the person I am today: my dad Jean-Philippe, my siblings Patricia and Gabe, their spouses Kino, Rafaela, and Eileen, my niece and nephews Vera, Pascal, Leo, and Remy, and all of the extended McCluskey and Mathy clans (or *le tribu*, as my dad says), especially those in the Bay Area who have welcomed me into their lives these past years. And I still see my mom, Rosemary, every time I look out at the Pacific Ocean; my soul and hers are tied, here in California.

Contributions

Some of the material in this dissertation has been published in the references below. The co-authors listed in these publications assisted, directed, or supervised the research that forms the basis for this dissertation.

Chapter 2:

Perica, T.*, Mathy, C. J. P.*, Xu, J., Jang, G. M., Zhang, Y., Kaake, R., Ollikainen, N., Braberg, H., Swaney, D. L., Lambright, D. G., Kelly, M. J. S., Krogan, N. J., & Kortemme, T. (2021). systems level effects of allosteric perturbations to a model molecular switch. *Nature* 599, 152–157. https://doi.org/10.1038/s41586-021-03982-6

* denotes equal contribution

T.P., C.J.P.M., N.J.K. and T.K. identified and developed the core questions. T.P. and C.J.P.M. performed the majority of the experiments and data analysis. J.X. and T.P. performed the E-MAP screens. G.M.J. performed the pull-down experiments. D.L.S. and R.K. performed the mass spectrometry experiments and together with T.P. analyzed the data. N.O. contributed to the design of Gsp1 mutants. H.B. contributed to E-MAP analysis. M.J.S.K. suggested the NMR studies. C.J.P.M. and M.J.S.K. performed the NMR experiments and analyzed the data. T.P. performed the kinetics experiments. D.G.L. contributed to the analysis of the kinetics data. T.P., C.J.P.M. and Y.Z. purified the proteins. Y.Z. performed the western blot experiments. T.P., C.J.P.M. and T.K. wrote the manuscript with contributions from the other authors. N.J.K. and T.K. oversaw the project.

Chapter 3:

Mathy, C. J. P., Mishra, P., Flynn, J. M., Perica, T., Mavor, D., Bolon, D. N. A.†, & Kortemme, T.† (2022). Complete mutational mapping of a GTPase switch in vivo reveals novel allosteric regulation. In *bioRxiv* 2022.04.13.488230. https://doi.org/10.1101/2022.04.13.488230

† denotes co-corresponding author

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.

Systems biology of allosteric GTPase mutations

Christopher John Pascal Mathy

Abstract

Living systems operate at many scales, from biochemical reactions of individual atoms and molecules to complex behaviors of cells and organisms, and even evolutionary adaptation of entire ecosystems. Understanding the relationships between these processes, namely how changes at one scale propagate to other scales, is a fundamental pursuit of biology. One such complex propagation is called a genotype-phenotype map, defined here as how a protein mutation impacts its function in the context of its molecular interaction network to ultimately alter cellular fitness. Our generally poor understanding of this propagation limits our prediction of the effects of disease mutations and our ability to rationally engineer mutations for precisely tuning protein function in the dynamic cellular environment. In this dissertation, I present two studies of the small GTPase switch Gsp1, the *S. cerevisiae* homolog of human Ran, which uncover novel allosteric mechanisms governing how the effects of point mutations propagate from the molecular to the cellular scale.

In Chapter 1, I outline the systems biology approach to studying molecular interaction networks, introduce the components of the network of Ran/Gsp1, and motivate the use of mutagenesis in the study of protein structure and cellular function. In Chapter 2, I describe the genetic and physical interaction profiling of point mutations in Gsp1 partner interfaces, which led to the discovery of novel allosteric sites coupled to the GTPase switch, as confirmed by enzyme kinetics and ^{31}P nuclear magnetic resonance. Analysis of the genetic interaction profiles showed that distinct cellular processes were sensitive to changes in either the rates of GTPase hydrolysis

or nucleotide exchange, prompting a model for a single GTPase selectively and independently controlling different downstream pathways by regulated tuning of its switching. In Chapter 3, I describe a mutational scanning study which quantitatively measured the fitness effect of all possible point mutations in Gsp1. The scan revealed an unexpected widespread toxic/gain-of-function response, in which mutations were more deleterious than loss of gene function by truncation of Gsp1 via internal STOP codon. Sites enriched for toxic/gain-of-function mutations included a novel allosteric cluster of residues which stabilize the GDP-bound state of Gsp1, confirmed by enzyme kinetics. The study defined a functional map of allosteric regulatory sites in Gsp1 which generalizes to other GTPases and confirmed that perturbation of the switch mechanism is the dominant factor in the effect Gsp1 mutations exert at the cellular level. Finally, in Chapter 4, I discuss the implications of these findings for future studies of molecular switches and their interaction networks, as well as for the use of high-throughput genome-wide measurements to guide the engineering of protein function.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP-MS – *affinity purification mass spectrometry*

CD – *circular dichroism*

DMS – *deep mutational scanning*

E-MAP– *epistatic mini-array profile*

EMPIRIC – *extremely methodical and parallel investigation of randomized individual codons*

GAP – *GTPase activating protein*

GEF – *guanosine nucleotide exchange factor*

GI – *genetic interaction*

GOF – *gain-of-function*

GTP/GDP – *guanine 5' triphosphate / diphosphate*

MUT – *mutant*

NMR – *nuclear magnetic resonance*

PDB – *Protein Data Bank*

PTM – *posttranslational modification*

ΔrASA – *change in relative accessible surface area*

SCA – *statistical coupling analysis*

SGA – *synthetic genetic array*

WT – *wild type*

Chapter 1. Introduction

Systems biology: uncovering the mechanisms governing molecular interaction networks

Life depends on the integrated and coordinated activity of biochemical processes performed by many distinct molecular species sequestered together in cells, such as proteins, nucleic acids, lipids, sugars, organic small molecules, and inorganic ions. This intricate dance of molecules can be abstracted as graphs in which each node represents a species and each edge represents some form of molecular interaction. So-called molecular interaction networks are widely used to represent metabolic or signaling pathways, protein-protein interaction networks, and gene regulatory circuits. The fields of *systems biology* or *network biology* encompass the collection and integration of experimental data on molecular interaction networks followed by the use of computational tools to model, inspect, and predict systems level behaviors. By providing frameworks for understanding and predicting the complex processes of life, systems biology can advance biological discovery, inform therapeutic interventions, and guide the bioengineering of new living systems.

Models of interaction networks rely on *molecular mechanisms*, or the rules for how various interactions change the quantity, identity, or activity of the molecules involved. Biologists uncover, test, and update molecular mechanisms through the iterative process of experimentation and model building, which includes both conceptual and quantitative models, the latter of which can be mathematical, statistical, or both. By focusing on the emergent properties of molecular interaction networks, systems biologists seek to uncover new or expand existing mechanisms to capture natural processes more accurately.

An example can illustrate what molecular mechanisms are, and how systems biologists approach the task of defining or refining them. One of the most well-known and foundational molecular mechanisms is the central dogma of molecular biology, which specifies the allowed (and disallowed) paths of residue-by-residue transfer of genetic information in biological systems, with the most common being gene expression through transcription followed by translation (DNA → RNA → Protein).[1] Recent advances in molecular biology techniques have uncovered more nuanced details of gene expression, such as the phenomenon of discontinuous temporal regimes of transcriptional activity known as *bursting.*[2] While bursting can be recapitulated *in vitro* using a small number of molecular parts,[3] single molecule RNA probes and single-cell RNA sequencing have shown its *in vivo* relevance is more complex and widespread than previously appreciated, with some studies going so far as to suggest that burst frequency controlled by enhancers could be the major factor leading to transcriptional changes during development.[4,5] The studies that advanced the bursting field followed a "systems" approach: rigorous, quantitative experimental measurements of molecular species in their full biological context produced datasets for parameterizing, evaluating, and discriminating between many possible models, with the ultimate goal of expanding on how a powerful molecular mechanism (the central dogma) truly operates in a complex biological process (gene expression).

Generally, systems biologists posit that iterative integration of systems level experimentation with computational modeling can lead to more complete and actionable understanding of complex biology,[6] with the model providing insight that is "greater than the sum of its parts". This approach contrasts with reductionism, which asserts that an understanding of each biochemical component individually is sufficient to describe the behavior of an entire biological system.[7] This does not mean reductionist models cannot explain much of biology; rather, systems level models are needed

to extend reductionist models so as to explain *the rest of* biology. And just as the advent of molecular biology revealed that much of biology was encoded in the sequence and structure of biomolecules, the high-throughput genome-wide technologies of the last two decades have shown that an additional level of complexity emerges from the structure and dynamics of molecular interaction networks. Thus, a systems biology approach to understanding a biochemical phenomenon truly relies on both molecular-scale and systems-scale investigation.

A non-exhaustive list of other important (but in many cases, incompletely understood) mechanisms operating at the molecular and systems scales are shown in **Table 1.1** and **Table 1.2**. Many of the molecular mechanisms have been observed to act in concert, even within single molecular interaction events, to produce the emergent systems level behaviors of cellular pathways. For instance, allosteric stabilization of a functional conformation is essential to the cooperativity observed in the classic example of hemoglobin.[8] Likewise, systems level phenomena often occur together: feedback and ultrasensitivity can drive oscillations and bistability,[9] and some (but not all) bistable systems are known to exhibit hysteresis.[10,11] The specific local connectivity of groups of molecular interactions, known as a "network motif,"[12] is also key in defining which systems level properties are accessible to a network.[10] Much work has been done outlining modular motifs that can operate predictably in cells,[13–16] such as circuits driven by chimeric antigen receptors (CARs) functionalized on T cells to respond to sets of signals unique to tumor microenvironments. More generally, defining the relationships between these and other molecular and systems level mechanisms is necessary for better understanding cellular processes in both healthy and disease states, for predicting how therapeutic interventions will act, and for guiding the engineering of new cellular behaviors.

The studies in this thesis aim to identify the molecular and systems level mechanisms at play in the essential and conserved eukaryotic interaction network of the small GTPase switch Ran/Gsp1, introduced in detail in the next section. Ran/Gsp1 was chosen for its notable systems level property of multifunctionality, as it regulates several distinct processes. The studies ultimately present evidence for differential sensing of the kinetics of switching by distinct processes as the systems level mechanism underlying Ran/Gsp1 multifunctionality, and for significant and underappreciated allostery in the GTPase fold as an essential molecular feature allowing for interaction partners and posttranslational modifications to regulate the switch.

The molecular interaction network of the small GTPase Ran/Gsp1, a model protein switch

Gsp1 (the *S. cerevisiae* homolog of human Ran) is a member of the Ras-superfamily of small GTPases,[17] themselves a part of a larger class of proteins known as switches. Protein switches undergo regulated cycling between two states, one of which is the principally active on-state. In GTPases, the on- and off-states correspond to binding to GTP or GDP, respectively, with the former being recognized preferentially by downstream effectors. Activation occurs through exchange of a bound GDP for a GTP from solution while deactivation occurs via hydrolysis of a bound GTP to GDP. These rates of switching are intrinsically slow but accelerated via interaction with opposing regulators, the GTPase activating proteins (GAPs) and the guanine nucleotide exchange factors (GEFs). While , while other GTPases have been observed to interact with several distinct GAPs and GEFs, Ran/Gsp1 is known to have only one GAP (human RanGAP / yeast Rna1) and one GEF (human RCC1 / yeast Srm1). The key structural features of small GTPases are a set of evolutionarily conserved regions that confer specificity for binding guanine

nucleotides, as well as the well-known Switch I and Switch II loops which undergo large conformational changes depending on the nucleotide bound and resulting in the state-dependent differences in interaction affinities with partners.[18] In addition, Ran/Gsp1 has a unique C-terminal extension which stabilizes the GDP-bound state and facilitates binding of the key partner RanBP1/Yrb1, in the GTP-bound state.

Due to their ability to activate or inhibit signal transduction in a regulated fashion, GTPases and other switch proteins are key components of larger signaling pathways that need to selectively turn *on* or *off*. [19] This centrality means switches are often misregulated in disease,[20] and also makes them attractive targets for engineering responsive behavior and external control into both existing and novel cellular pathways.[21] Our understanding of protein switches is extensive enough that simpler types of protein switches, such as sensors of a single input signal, can be readily designed *de novo*.[22,23] Additionally, design rules have been mapped for some naturally-occurring protein switches with more complex molecular or systems level functions, such as those which rely on allostery between distinct protein domains[24] or those which show ultrasensitivity.[25]

However, Ran/Gsp1 has a notable characteristic that differentiates it from many other switches and for which no design rules are readily available: it is multifunctional, both at the molecular level in that it binds many partners from several fold families (also known as multi-specificity),[26] and at the systems level in that it regulates many distinct cellular processes. These processes include nucleocytoplasmic transport of both proteins and RNA,[27,28] cell cycle progression,[29,30] spindle assembly,[31,32] nuclear envelope assembly,[33] chromatin structure at the telomeres,[34,35] and even ciliogenesis in humans.[36] Ran/Gsp1 regulation of these pathways is mediated by interactions with an extensive network of effector proteins, many of which have been characterized through high-resolution crystal structures. These interaction partners include: the karyopherin transporter

family;[37,38] the essential partner NTF2, which binds the GDP-bound form of Ran/Gsp1 to shuttle it into the nucleus; a class of proteins containing the Ran-binding domain (RanBD, the monomeric version of which is known as Yrb1 in yeast and RanBP1 in humans);[39,40] nucleoporins containing RanBDs as well as zinc finger domains that bind Ran/Gsp1;[41] and many other proteins from various fold families and cellular pathways such as the exosome subunit Rrp44/Dis3,[42,43] the GID/CLTH complex components RanBP9 and RanBP10,[44–46] and the conserved regulator RanGRF/Mog1 which stimulates nucleotide release of Ran/Gsp1 and sequesters the switch in a stable nucleotide-free complex,[47] a process thought to be important for removing active Ran/Gsp1 from the GTPase cycle[48] and/or enabling faster nucleotide exchange in concert with RanBP1/Yrb1.[49,50] These many interactions place Ran/Gsp1 as the primary regulator at the center of a dense molecular interaction network. Ran has no homologs in humans, and although *S. cerevisiae* Gsp1 has one highly similar (>95% amino acid identity) paralog named Gsp2 arising from an ancient whole genome duplication,[51,52] Gsp2 has been observed at 10-fold lower expression levels than Gsp1, and unlike Gsp1 exhibits carbon source dependency and is not required for cell viability.[51]

The central role of Ran/Gsp1 in key cellular processes, its high degree of connectivity in a large interaction network, and its relative lack of functional homologs evoke a puzzle: how can a single protein *simultaneously* but *independently* regulate so many unrelated processes? That is, how can Ran/Gsp1 dynamically alter the activity of one pathway (e.g., nuclear export) in response to a stimulus without also transducing the same change to all other pathways (e.g., mitotic progression). The studies in Chapters 2 and 3 seek to uncover the mechanisms behind this multifunctionality by perturbing *S. cerevisiae* Gsp1 with point mutations and examining the resulting changes in structure, biochemical function, network interactions, and cellular fitness.

Point mutations as high-resolution, targeted perturbations of gene function

The use of mutations to explore gene function is a foundational idea in genetics. Mutations introduce variations in genes, and gene function can be inferred by examining the resulting changes to measurable phenotypes. While the human *RAN* allele was identified through a cDNA library screen for homology to Ras proteins,[53] the *S. cerevisiae GSP1* and *GSP2* alleles were identified using a random mutagenesis approach coupled to a screen for genetic interactions with the GEF Srm1, identifying mutations which suppressed a temperature-sensitive allele.[51]

As opposed to discovery of gene function using random mutations and targeted screens, molecular biology techniques have also enabled precise engineering of point mutations to explore the contribution of individual residues to the structure, stability, and biochemical function of proteins.[54,55] Scaling up of mutagenesis to multiple sites through experimental or computational alanine scanning allowed for mapping the regions of proteins contributing to a function and identifying "hotspot" residues whose relative contribution to function is greatest.[56–58] Alanine was chosen for its relatively neutral sidechain (a single methyl group), allowing for mutations to probe the role of the original sidechain functional group without introducing additional effects. In recent years, approaches collectively known as *deep mutational scanning* (DMS) have expanded on this scaled up approach to measure the effects that all possible single amino acid substitutions at each site have on a molecular function or cellular fitness.[59–61] The study presented in Chapter 3 uses DMS to construct an exhaustive functional map for Gsp1.

In parallel to improvements in mutagenesis techniques, genome-wide techniques were developed in functional genetics to better define a given gene's function in the full cellular context. In yeast, a genetic interaction profiling method known as synthetic genetic arrays (SGA) was developed to measure functional relationships between genes in a medium-throughput, array-

based format.[62] Double mutant strains are generated by crossing a series of mutants (*query* strains) against a genome-wide set of single gene deletions (*array* or *library* strains), and growth rates are measured for the double mutant and single mutant strains. A quantitative genetic interaction score is computed as the deviation of the observed double mutant growth rate from its expected rate when assuming the two genes do not have a functional relationship, calculated as the product of each single mutant growth rate. If the double mutant grows slower than expected, the genetic interaction is called *negative, synthetic sick,* or *synergistic*, and suggests that the two genes in question buffer one another and/or participate in complementary ways in the same biological process. A double mutant that grows less poorly than expected is called *positive, alleviating,* or *antagonistic*, and suggests that the two genes function directly together (potentially even in the same complex), since deletion of one alters the shared cellular pathway enough that a second deletion is less consequential.[63]

When many genetic interactions are measured for the same query strain, a *genetic interaction profile* can be constructed, serving as a vector representation of the phenotype of the query mutation in a large number of genetic backgrounds. Correlation between profiles of two genes is highly indicative of related function. Early technical limitations in the number of gene pairs that could be screened by SGA prompted the development of a related approach termed epistatic miniarray profiles (E-MAPs). The developers of E-MAPs reasoned that since genetic interactions were rare, it would be useful to select query genes within a smaller set of known colocalized or coregulated genes, so that more high confidence functional interactions could be included in profiles, even while screening a smaller fraction of all possible gene pairs.[64] Both SGA and E-MAP have been used widely by yeast genetics labs throughout the world, and genetic interaction profiling has proven powerful for many applications such as the genome-wide mapping of

functional relationships between genes[65] or the prediction and even structure determination of genes physically interacting in complexes.[66,67]

While genetic interaction profiles based on gene deletions are powerful, they cannot be used to examine the cellular roles of individual functions in multifunctional proteins like Ran/Gsp1, since the deletion of a gene removes all its functions at once. For this reason, a variation on the E-MAP approach using point mutations as queries, termed pE-MAP, was developed and tested on RNA polymerase II (RNAPII).[68] Analysis of the pE-MAP enabled functional assignment of individual RNAPII subdomains, and grouped point mutations based on not only their biophysical properties but also their cellular effects, as indicated by high correlations with genetic profiles of multiple protein complex members and verified experimentally. In this way, the pE-MAP approach combines the structural resolution afforded by targeted point mutations with the rich cellular phenotypes of genetic interaction profiles to enable a functional analysis of proteins that spans the molecular and systems scales. The study presented in Chapter 2 uses the pE-MAP approach in concert with physical interaction profiling to construct a genotype-phenotype map for several point mutations in Gsp1.

Mutagenesis studies of proteins in their cellular context are not only useful in predicting the effects of other mutations, but they can also predict the action of other types of perturbations, such as partner binding, posttranslational modifications (PTMs), or small molecule binding. For example, mutations are considered *mimetic* if the mutation results in a new sidechain that is chemically similar to a posttranslationally modified sidechain of the original amino acid. Serine to aspartate mutations are considered phosphomimetic since the carboxyl group of aspartate mimics phospho-serine,[69,70] and lysine to glutamine mutations are considered acetyl-mimetic since the amide group of glutamine somewhat mimics acetylated lysine.[71,72] In this way, study of mutations

that perturb proteins in the same way as a PTM can inform on the molecular and cellular impact of that PTM.

This concept can be extended to mutations at protein interaction sites and small molecule binding sites, as the rearrangement of atoms induced by the mutation can be similar to the rearrangement induced by binding. More formally, residues at any two sites in a protein can be identified as *thermodynamically coupled* if the change in energy of a process due to mutation at one site changes in the background of a mutation at the other site, [73] and this coupling can be exploited by binding interactions or PTMs. This physical similarity between the effects of mutations and other types of perturbations to proteins is core to the value of mutations in understanding cellular regulation of protein function in cells. Cells do not use mutations to reversibly regulate protein function, but instead use PTMs or the binding of other proteins or small molecules. Furthermore, while regulation at known functional sites (such as the active site of an enzyme) can be straightforward to predict from structural information alone, regulation that occurs at unknown allosteric sites is harder to predict. Mutagenesis can serve as a reliable tool to probe for allosterically coupled sites which can then be validated by computational[74] and experimental techniques.[75] This approach is seen in the study presented in Chapter 2: mutations reveal that partner interfaces on Ran/Gsp1 are coupled to the nucleotide binding site, as confirmed by observation of similar changes in the chemical environment of a bound nucleotide's gamma phosphate after distal mutation or partner binding.[76]

Table 1.1 Examples of molecular interaction network principles that operate at the molecular scale.

Mechanism Name	Description
Allosteric Regulation	Modulation of a functional site in a protein by a perturbation at a distal site, as opposed to *orthosteric regulation* via perturbations at the site itself.[77]
Cooperativity	The characteristic of a system in which subsequent molecular events of the same type become more or less favorable based on previous events, often due to allosteric coupling.[78]
Specificity	A metric of how preferred an interaction with one partner is over interactions with proteins related to the partner. It contrasts with promiscuity or molecular multi-specificity, in which multiple similar interactions (of the same or different family of interactors) are similarly favored.[79–81]
Selectivity / Biased Agonism	A phenomenon of different ligand inputs of a receptor resulting in recruitment of distinct sets of downstream signaling proteins.[82,83]
Homomer Asymmetry	Distinct asymmetry of individual subunits in homooligomers, which expands the number of functional conformations sampled by these complexes.[84,85]
Ordered Complex Assembly	The unique pathway(s) by which subunits of larger macromolecular complexes assemble and disassemble.[86,87]
Competition / Occlusion	The existence of overlapping functional sites that prevent two interactions from occurring simultaneously. Differences in interaction strengths allow for thermal ratchets, where sequences of interactions can occur in a unidirectional fashion when coupled to an energy source.[26,37]

Table 1.2 Examples of molecular interaction network principles that operate at the systems scale.

Mechanism Name	Description
Feedback, Autoregulation	A widespread feature of dynamical systems in which the system's output also serves as an input, enabling its current state to dynamically alter its behavior.[88,89]
Multifunctionality or multi-specificity	The observation of a protein performing several unrelated functions, effectively participating in multiple cellular interaction networks that are otherwise insulated from each other.[90–93]
Logic Gates	A general name for network motifs that integrate signals according to Boolean logic operations.[94–96]
Ultrasensitivity	A property of switch-like network motifs, in which a small change in input signal results in a sharp change in downstream signal, often magnified via cascades (multiple switches connected in series) [97,98].
Oscillations	Equilibrium states of a system that nonetheless show time-dependent, periodic differences in activity level.[99,100]
Multistability	The existence of two or more states of a system which are "effectively irreversible", meaning that they are stable and persist until perturbed by a stimulus that can push them into another state.[101–103]
Hysteresis / Memory	The dependence of a signal response function on the previous state of the system, i.e. in which high previous levels of an input signal[10]
Fold-change Detection	A property of systems which respond chiefly to fold-change differences of input signal, but are robust to other parameters, such as the absolute levels of internal components.[104,105]
Crosstalk	A very broad term for how multiple different input signals can affect a common output, indicating that two or more cellular pathways are connected directly (e.g. via shared pathway component) or indirectly (e.g. one pathway activates the expression of another pathway).[106]
Temporal / Kinetic Filtering	The ability for a system to respond differently to an input signal depending on the duration of time it persists.[107]
Phase Separation / Biomolecular Condensates	The compartmentalization of a nonstoichiometric assemblies of biomolecules without lipid membranes, allowing for selective inclusion of certain species and local changes in effective concentration.[108]

Chapter 2. Systems-level effects of allosteric perturbations to a model molecular switch

Summary

Molecular switch proteins whose cycling between states is controlled by opposing regulators[109,110] are central to biological signal transduction. As switch proteins function within highly connected interaction networks,[111] the fundamental question arises how functional specificity is achieved when different processes share common regulators. Here we show that functional specificity of the small GTPase switch protein Ran/Gsp1[112] is linked to differential sensitivity of biological processes to different kinetics of the Ran/Gsp1 switch cycle. We make 55 targeted point mutations to individual protein interaction interfaces of Ran/Gsp1 and show through quantitative genetic[63] and physical interaction mapping that Ran/Gsp1 interface perturbations have widespread cellular consequences. Unexpectedly, the cellular effects of the interface mutations group by their biophysical effects on kinetic parameters of the GTPase switch cycle and not by the targeted interfaces. Instead, we show that interface mutations allosterically tune the GTPase cycle kinetics. These results suggest a model where protein partner binding, or posttranslational modifications at distal sites, could act as allosteric regulators of GTPase switching. Similar mechanisms may underlie regulation by other GTPases, and other biological switches. Furthermore, our integrative platform to determine the quantitative consequences of molecular perturbations may help explain the effects of disease mutations targeting central molecular switches.

Introduction

Proteins perform their cellular functions within networks of interactions with many partners.[111] The interconnectivity of these networks raises the fundamental question of how different individual functions can be controlled with the required specificity, especially when distinct cellular processes share common regulators. Moreover, in highly interconnected networks even small perturbations could have widespread cellular effects.[67,68]

To determine the mechanism and extent by which molecular perturbations affect interconnected biological processes, we targeted a central molecular switch, a GTPase. GTPases are two-state switches controlled by regulators with opposing functions.[110] The two states of GTPase switches are defined by the conformation of their GTP- or GDP-bound forms, and the interconversion between states is catalyzed by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) (**Figure 2.1A**). Other, similar biological switch motifs involve covalent modifications controlled by opposing kinase/phosphatase or acetylase/deacetylase regulators. One striking feature of such motifs is their potential for ultrasensitive response to regulation, where small changes in the activity of the regulators can lead to sharp changes in the state of the switch.[109,110] Moreover, switch motifs such as GTPases are often multi-specific, defined here as regulating several distinct processes.[113]

Here we focus on the multi-specific small GTPase Gsp1 (the S. cerevisiae homolog of human Ran) as a model system. Gsp1/Ran is a highly conserved molecular switch with one main GEF and one main GAP[114] that regulates nucleocytoplasmic transport of proteins[115] and RNA,[116] cell cycle progression,[117] and RNA processing.[118] Crystal structures of Gsp1/Ran in complex with 16 different binding partners are known (**Figure 2.2**, **Figure 2.3**, **Table 2.1**). We reasoned that by placing point mutations in Gsp1 interfaces with these partners, we would differentially perturb

subsets of biological processes regulated by Gsp1. We then determined the functional consequences of these Gsp1 interface mutations on cellular processes in S. cerevisiae using quantitative genetic interaction mapping (**Figure 2.1A**), measured changes to the physical interaction network using affinity purification mass spectrometry (AP-MS), and quantified molecular effects on the Gsp1 switch using biophysical studies in vitro.

Results

Genetic interactions of Gsp1 mutants.

We designed 55 genomically integrated point mutant alleles of S. cerevisiae Gsp1 to target each of its 16 known interactions (**Figure 2.1B**, **Figure 2.2**, **Table 2.2**, **Table 2.3**), avoiding mutations in the Gsp1 nucleotide binding site and the switch I and II regions. We confirmed by Western blot that the mutant Gsp1 protein levels were close to the endogenous wild-type levels (**Figure 2.4**).

To determine the cellular effects of the Gsp1 interface mutations, we performed a genetic interaction (GI) screen in S. cerevisiae using the epistatic mini-array profile (E-MAP) approach.[63,68] We measured growth of each *GSP1* point mutant in the context of an array of 1444 single gene knockouts, resulting in a quantitative functional profile of up to 1444 GI values for each *GSP1* point mutant (**Supplementary Data 1**). The 55 *GSP1* point mutants fell into two clusters, 23 'strong' mutants with rich GI profiles containing 9-373 significant interactions (**Figure 2.1C**), and 32 'weak' mutants with 0-8 significant interactions (**Figure 2.5**, **Figure 2.6**, Methods). The strong mutants covered eleven Gsp1 sequence positions and all 16 structurally characterized Gsp1 protein interaction interfaces (**Figure 2.7A**). Twelve of the *GSP1* interface point mutants had a greater number of significant GIs than an average deletion of a non-essential S. cerevisiae

gene, and six *GSP1* point mutants had more GIs than an average temperature sensitive mutant of an essential gene in a published S. cerevisiae GI map[65] (**Figure 2.1D**). Hierarchical clustering of S. cerevisiae genes based on their GIs with the Gsp1 interface mutations grouped genes by their cellular functions, including mRNA transport, tRNA modification, and spindle assembly regulation (**Figure 2.1C**, **Figure 2.7B**). Taken together, the GI analysis reveals expansive functional consequences of Gsp1 interface point mutations - similar in magnitude to effects typically observed for deleting entire genes - that illuminated many of the biological functions of *GSP1*.

In contrast to their clustering of biological processes, the GI profiles of the Gsp1 point mutants did not group based on their location in the Gsp1 partner interfaces. For example, strains with *GSP1* mutations at residues T34 (T34E/Q) and D79 (D79S/A) have similar GI profiles (**Figure 2.1C**) but these mutations are in different interfaces (**Figure 2.7A**) on opposite sides of the Gsp1 structure (**Figure 2.1B**). This observation was contrary to our initial expectation that Gsp1 achieves its functional specificity by interacting with different partners. To analyze this finding further, we compared the *GSP1* mutant GI profiles to profiles from 3370 S. cerevisiae alleles[65] using Pearson correlations. Significant positive correlations of GI profiles indicate functional relationships[63] (**Supplementary Data 2**, **Table 2.4**, **Figure 2.7C**). Strikingly, GI profiles of *GSP1* mutants and of Gsp1 physical interaction partners were on average no more similar to each other in instances where the Gsp1 mutation was located in the partner interface than when the mutation was not (**Figure 2.1E**, **Figure 2.7D**). This result suggests that the rich functional profiles of *GSP1* mutants cannot simply be explained by considering only the partner interaction targeted by the interface point mutation.

Physical interactions of Gsp1 mutants.

To investigate further why the GI profiles of Gsp1 mutations did not group based on targeted physical interactions of Gsp1, we sought to determine how interface point mutations affected the physical protein interaction network of Gsp1. We tagged wild-type Gsp1 and 28 mutants covering all interface residues shown in **Figure 2.7A**with an amino- or carboxy-terminal 3xFLAG tag and quantified the abundance of 316 high-confidence 'prey' partner proteins in complex with Gsp1 by AP-MS (**Figure 2.8**, **Figure 2.9**, **Supplementary Data 3**). We refer to the prey partner protein abundance in the pulled-down Gsp1 complexes simply as "abundance" below. Six of the 16 Gsp1 binding partners for which we had structural information were robustly observable in the AP-MS data for both Gsp1 wild type and mutants: the two core regulators Rna1 (GAP) and Srm1 (GEF), as well as four effectors Yrb1, Kap95, Pse1 and Srp1. As expected, the abundance of the prey partner was decreased on average (although not always) when the Gsp1 mutation was in the interface core with the prey partner (**Figure 2.8A**, left distribution). However, instead of expected minimal effects, we also found notable changes in prey abundance in cases where the mutation was not directly in the interface (**Figure 2.8A**, right distribution). A wide spread of abundance changes was apparent for the two main GTPase regulators, GAP (Rna1) and GEF (Srm1), even for mutations at positions that are outside either of the interfaces such as T34 (**Figure 2.8B**, **Figure 2.10**, **Table 2.5**). In summary, the AP-MS experiments show that the point mutations, in addition to affecting the targeted interactions, also introduce extensive changes to the physical interaction network of Gsp1 that cannot simply be explained by the interface location of the mutations.

Effect of mutants on Gsp1 switch kinetics.

The AP-MS experiments showed that most Gsp1 interface mutations significantly altered physical interactions with the two principal GTPase regulators, GAP and GEF. This observation

prompted the question whether the mutations, rather than acting indirectly in the cellular context (i.e., by altering the competition between physical interaction partners in the cell), affected the molecular function of the switch directly. To assess the molecular effects of mutation on switch function, we recombinantly expressed and purified wild-type and 24 Gsp1 mutants and measured their effects on GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange *in vitro* (**Figure 2.11A, B**, **Figure 2.12**, **Figure 2.13**, **Figure 2.14**, **Table 2.6**, **Table 2.7**). Of the 24 Gsp1 point mutants, 17 (of which all except K132H had strong GI profiles) showed 3- to >200-fold change in kcat/Km on either or both of the GAP- or GEF-mediated reactions (**Figure 2.12E**). These results show that Gsp1 interface mutations can modulate the GTPase cycle by affecting GTP hydrolysis and nucleotide exchange catalyzed by the GAP and GEF. Moreover, since nine out of the 17 mutations with larger than 3-fold effects are located outside of the interface cores with either the GAP (**Figure 2.11A**) or the GEF (**Figure 2.11B**) as well as outside the known switch regions, our data suggest considerable, previously unappreciated, allostery in the GTPase switch.

Allosteric effects of mutations.

To probe the mechanism of these allosteric effects, we examined the impact of Gsp1 point mutations on the conformational distribution in the active site of GTP-bound Gsp1 using 1D ^{31}P nuclear magnetic resonance (NMR) spectroscopy. Prior ^{31}P NMR data on human Ran[76] showed two distinct peaks for the γ-phosphate of bound GTP arising from differences in the local chemical environment of the γ-phosphate in each of two distinct conformations (termed γ1 and γ2). Our ^{31}P NMR spectra of S. cerevisiae wild-type Gsp1:GTP showed two distinct peaks for the γ-phosphate of bound GTP with 87% of wild-type Gsp1:GTP in the γ2 state conformation (**Figure 2.11C**, **Figure 2.15**). Strikingly, the relative populations of the γ1 and γ2 states were modulated by our

Gsp1 interface mutations and ranged from close to 0% in the γ2 state for T34E and T34Q, to close to 100% for H141R, Y157A, and K132H (**Figure 2.11C**).

Furthermore, we observed a linear relationship between the effect of the mutation on the equilibrium between the γ1 and γ2 conformations (plotted as the natural log-transformed ratio of the equilibrium constant) and the natural log-transformed ratio of the relative catalytic efficiencies of GAP-mediated GTP hydrolysis (**Figure 2.11D**) and intrinsic GTP hydrolysis (**Figure 2.15B, C**, **Table 2.8**). This relationship suggests that the γ2 state represents the active site conformation of Gsp1:GTP competent for GTP hydrolysis. Exceptions to the linear relationship are K132H, which is in the core of the GAP interface and hence expected to directly affect the interaction with the GAP, and D79S and R78K, which are adjacent to the GTPase switch II region and could lead to different perturbations of the nucleotide binding site.

Remarkably, the mutated residues that tune the population of the γ2 state (positions T34, H141, Q147, and Y157) are all distal, affecting the chemical environment of the Gsp1-bound GTP γ phosphate from at least 18 Å away (**Figure 2.15D, E**). Taken together, our in vitro data support an allosteric mechanism where distal mutations at different surface interaction sites of Gsp1 modulate the GTPase switch by differentially affecting GEF-catalyzed nucleotide exchange and GAP-catalyzed GTP hydrolysis. Moreover, comparison between the in vitro kinetic and our AP-MS data showed that the direction of the GTPase cycle perturbation is a good predictor of altered physical interactions with the two main cycle regulators (**Figure 2.16**), even in the context of other potentially competing partner proteins.

Encoding of Gsp1 multi-specificity.

Finally, we asked whether the allosteric effects of the mutations on the GTPase cycle kinetics explained the functional effects observed in the cellular GI profiles. This analysis also provided

insights into the ability of Gsp1 to distinctly regulate different biological processes (functional multi-specificity). We clustered the GI profiles of the Gsp1 mutants based on correlation with the GI profiles of 3358 S. cerevisiae alleles;[65] 276 alleles had significant correlations to Gsp1 mutants (**Figure 2.17A**). We then compared clustering of these GI profile correlations with the biophysical effects of the Gsp1 mutations. Remarkably, the Gsp1 mutant GI profile clustering mirrored an approximate ordering by the in vitro mutant effects on the GTPase cycle: relative GAP efficiency systematically increased with increasing column number and relative GEF efficiency decreased (**Figure 2.17A**). (The clear outlier, K101R, could be explained by acetylation of this residue.[119] The K101R mutation could affect a critical mechanism by which the cell reduces GEF activity,[72] phenocopying mutants with reduced GTP hydrolysis activity.) Overall, genes in **Figure 2.17A** fall into one of three categories: (i) genes in cluster 1, but also cluster 2, that correlate with mutants primarily perturbed in GTP hydrolysis (**Figure 2.17A**, orange bars), (ii) genes in cluster 7 that correlate with mutants primarily perturbed in nucleotide exchange (teal bars), and (iii) genes that correlate strongly with all or most of the Gsp1 point mutants (cluster 5, but also clusters 3, 4, and 6).

Importantly, genes with shared biological functions (gene sets, **Supplementary Data 4**) all predominantly fall into one of the three categories defined above. For example, genes involved in spindle assembly regulation have significant GI profile correlations primarily with Gsp1 mutant group I (**Figure 2.17B**, red points), genes involved in tRNA modification primarily with Gsp1 mutant group III (blue points), and genes important for nucleocytoplasmic transport with Gsp1 mutants from all three groups (green points). The three groups of Gsp1 mutants show distinct kinetic characteristics: Group I has decreased efficiency of GTP hydrolysis, group III decreased nucleotide exchange, and group II shows intermediate behavior (**Figure 2.17C**). Therefore, our

analysis suggests that distinct cellular processes regulated by Gsp1, such as spindle assembly regulation, tRNA modification, and nuclear transport (**Figure 2.17B, D**), as well as 5' mRNA capping, transcriptional regulation, cytoplasm-to-vacuole targeting, and actin, tubulin and cell polarity (**Figure 2.18**) are differentially sensitive to perturbations of GTPase cycle kinetics. Taken together, our findings lead to a model where Ran/Gsp1 acts by three different modes defined by the sensitivity of different biological processes to perturbations of different characteristics of the Gsp1 GTPase cycle, i.e. the ability to (i) cycle, (ii) turn off by hydrolyzing to Gsp1:GDP, and (iii) turn on by producing Gsp1:GTP (**Figure 2.17D**). While other effects such as changes in interaction affinities or expression levels undoubtedly also play a role in modulating the functional effects of our Gsp1 mutations, our model explains to a remarkable degree how a single molecular switch motif can differentially control subsets of biological processes by using one of the three functional modes.

Discussion

The discovery of several new allosteric sites (positions 34, 141, 147, and 157) in Gsp1 has implications for GTPase regulation. Our finding that mutations in Gsp1 interfaces allosterically modulate the switch cycle identifies thermodynamic coupling between distal interfaces and the active site; partner binding or posttranslational modifications at these distal sites could also regulate the switch.

Our observation of widespread functional effects of point mutations inducing relatively small perturbations in the GTPase switch kinetics is reminiscent of the zero-order ultrasensitivity achievable in biological motifs with opposing regulators.[110] While switch-like ultrasensitivity is typically described for systems controlled by covalent modifications (such as phosphorylation),

our results, as well as the observations that cellular levels of small GTPase regulators require tight control,[120,121] corroborate a model of ultrasensitivity for GTPase conformational switches.[122]

Given the prevalence of biological two-state switch motifs with opposing regulators (kinase/phosphatase, acetylase/deacetylase), we envision our approach to engineer defined molecular perturbations and characterize them with systems-level functional genetics integrated with molecular biophysics to be informative for other studies of cellular regulation. The approach could be extended to mammalian systems using CRISPR-based approaches to yield mechanistic insights into the consequences of disease mutations targeting central molecular switches.

Methods

Point mutations in genomic Gsp1 sequence

We identified all residues in Gsp1 that comprised the interfaces with Gsp1 binding partners for which co-complex crystal structures with Gsp1 were available (**Figure 2.3**, **Figure 2.2**, **Table 2.1**). Residues comprising the interface *core*, the surface exposed *rim* around the core, and more buried *support* residues were defined based on per-residue relative solvent accessible surface area (rASA), as previously described.[123] rASA is compared to the empirical maximum solvent accessible surface area for each of the 20 amino acids.[124] rASA values were calculated for the Gsp1 monomer (rASAmonomer) and for the complex (rASAcomplex) using the bio3d R package.[125] The three types of interface residues were defined as: interface core if rASAmonomer > 25%, rASAcomplex < 25% and ΔrASA (change upon complex formation) > 0; rim residues if rASAcomplex > 25% and ΔrASA > 0; and support residues if rASAmonomer < 25% and ΔrASA > 0. All custom code for interface analysis from co-complex crystal structures is provided in the associated code repository at

https://github.com/tinaperica/Gsp1_manuscript/tree/master/Scripts/complex_structure_analyses. We avoided Gsp1 residues that are within 5 Å of the nucleotide (GDP or GTP) in any of the structures or that are within the canonical small GTPase switch regions[126] (P-loop, switch loop I, and switch loop II). We then mutated residues that are located in interface cores (defined as residues that bury more than 25% of their surface upon complex formation, as previously defined,[123] **Table 2.2**, **Figure 2.2**) into amino acid residues with a range of properties (differing in size, charge and polarity) and attempted to make stable and viable *S. cerevisiae* strains carrying a genomic Gsp1 point mutation coupled to nourseothricin (clonNAT / nourseothricin, Werner BioAgents GmbH, CAS 96736-11-7) resistance (**Figure 2.19**). The list of attempted mutants is provided in **Table 2.3**. The genomic construct was designed to minimally disrupt the non-coding sequences known at the time, including the 5′ UTR and 3′ UTR, as well as the putative regulatory elements in the downstream gene Sec72 (**Figure 2.19**). The *GSP1* genomic region was cloned into a pCR2.1-TOPO vector (Invitrogen) and point mutations in the *GSP1* coding sequence were introduced using the QuikChange™ Site-Directed Mutagenesis (Stratagene, La Jolla) protocol. *S. cerevisiae* strains containing mutant *GSP1* genes were regularly confirmed by sequencing the Gsp1 genomic region.

S. cerevisiae **genetics and genetic interaction mapping**

S. cerevisiae transformation

To generate MAT:α strains with Gsp1 point mutations the entire cassette was amplified by PCR using *S. cerevisiae* transformation forward and reverse primers, and *S. cerevisiae* was transformed into the starting SGA MAT:α his3D1; leu2D0; ura3D0; LYS2þ; can1::STE2pr-

SpHIS5 (SpHIS5 is the S. pombe HIS5 gene); lyp1D::STE3pr-LEU2 strain from[127] as described below. Primers for amplifying the *GSP1* genomic region were:

- FWD: GTATGATCAACTTTTCCTCACCTTTTAAGTTTGTTTCG
- REV: GATTGGAGAAACCAACCCAAATTTTACACCACAA

DNA competent *S. cerevisiae* cells were made using a LiAc protocol. The final transformation mixture contained 10 mM LiAc (Lithium acetate dihydrate, 98%, extra pure, ACROS Organics™, CAS 6108-17-4), 50 µg ssDNA (UltraPure™ Salmon Sperm DNA Solution, Invitrogen, 15632011), 30% sterile-filtered PEG 8000 (Poly(ethylene glycol), BioUltra, 8,000, Sigma-Aldrich, 89510-250G-F). A *S. cerevisiae* pellet of approximately 25 µl was mixed with 15 µl of linear DNA PCR product and 240 µl of the transformation mixture, and heat shocked at 42 ºC for 40 minutes. Transformed cells were grown on YPD (20 g Bacto™ Peptone (CAT # 211820, BD Diagnostic Systems), 10 g Bacto™ Yeast Extract (CAT # 212720 BD), and 20 g Dextrose (CAT # D16-3, Fisher Chemicals) per 1-liter medium) + clonNAT plates and incubated at 30 ºC for 3 to 6 days. Many colonies that appeared after 24-48 hours carried the clonNAT cassette but not the *GSP1* point mutation, or the 3xFLAG tag. Cells were therefore sparsely plated, and plates were incubated for a longer period of time after which colonies of different sizes were picked and the mutant strains were confirmed by sequencing.

Epistatic mini-array profiling (E-MAP) of Gsp1 point mutants

Genetic interactions of all viable *GSP1* point mutant (PM-GSP1-clonNAT) strains were identified by epistatic miniarray profile (E-MAP) screens[64,127] using a previously constructed array library of 1,536 KAN-marked (kanamycin) mutant strains assembled from the *S. cerevisiae* deletion collection[128] and the DAmP (*d*ecreased *a*bundance by *m*RNA *p*erturbation) strain

collection,[66] covering genes involved in a wide variety of cellular processes.[68] The E-MAP screen was conducted as previously described in Collins et al.,[66] using the HT Colony Grid Analyzer Java program[63] (http://sourceforge.net/project/showfiles.php?group_id=163953) and the E-MAP toolbox for MATLAB (http://sourceforge.net/project/showfiles.php?group_id=164376) to extract colony sizes of double mutant strains and a statistical scoring scheme to compute genetic interaction scores. Genetic interaction scores represent the average of 3-5 independent replicate screens. Reproducibility was assessed as previously described[63] by comparing individual scores to the average score for each mutant:gene pair, with the two values showing strong correlation across the dataset (Pearson correlation coefficient = 0.83, **Figure 2.20**).

Hierarchical clustering of E-MAP genetic interaction data

All E-MAP library DAmP strains as well as library strains showing poor reproducibility were discarded, leaving 1444 out of the original 1536 library genes. Averaged S-scores of genetic interactions between wild-type and point mutant Gsp1 and the 1444 *S. cerevisiae* genes are provided in **Supplementary Data 1**. Hierarchical clustering on the GI profiles was performed using the average linkage method and the pairwise Pearson correlation coefficient as a distance metric. To identify clusters of functionally related library genes, the hierarchical clustering tree was cut to produce 1200 clusters, resulting in 43 clusters with 3 or more members. Biological function descriptions for genes in these clusters were extracted from the *Saccharomyces* Genome Database (SGD).[129] Clusters of genes representing common functions (complexes, pathways or biological functions) were selected by manual inspection and represented in the main text **Figure 2.1C** and **Figure 2.7B**. All custom code for E-MAP analysis is provided in https://github.com/tinaperica/Gsp1_manuscript/tree/master/Scripts/E-MAP. Clustered heatmaps were produced using the ComplexHeatmap package.[130]

Scaling of published genetic interaction data to the E-MAP format

To enable comparison of *GSP1* point mutant GI profiles to GI profiles of other *S. cerevisiae* genes, published Synthetic Gene Array (SGA) genetic interaction data[65] from CellMap.org[131] were scaled to the E-MAP format using a published non-linear scaling method.[132] First, 75,314 genetic interaction pairs present in both the SGA and a previously described E-MAP dataset used to study chromatin biology[66] were ordered by genetic interaction score and partitioned into 500 equally sized bins separately for each dataset. Bin size (150 pairs per bin) was chosen to provide enough bins for fitting the scaling spline (described below) while still maintaining a large number of pairs per bin such that the mean could be used as a high confidence estimate of the score values in each bin. Scaling factors were computed that scaled the mean of each SGA bin to match the mean of the corresponding E-MAP bin. A non-linear univariate spline was fit through the scaling factors, providing a scaling function that was subsequently applied to each SGA score. The distribution of scores of shared interactions between the scaled SGA and the E-MAP chromatin library was similar to that between replicates in the E-MAP chromatin library, matching what was seen in the previously published scaling of SGA data to E-MAP format[132] (**Figure 2.21**). The SGA genetic interaction scores are taken from CellMap.org.[131] The scaling code is provided in https://github.com/tinaperica/Gsp1_manuscript/tree/master/Scripts/SGA_Scaling.

Significance of genetic interactions

The S-score metric used in scoring genetic interactions measured by the E-MAP method has been previously characterized in terms of confidence that any given averaged S-score represents a significant interaction.[63] We fit a spline to data points from Figure 4C in Collins et al,[63] allowing us to provide an approximate confidence estimate for each of our measured *GSP1* and scaled *S. cerevisiae* SGA genetic interaction scores. The SGA dataset[65] is accompanied by p-values as well

as its own recommendations for a threshold at which individual interactions are considered significant. We plotted the SGA score scaled to E-MAP format vs. the associated p-value (negative log-transformed, **Figure 2.6A**) and found the distribution to have a similar shape to the confidence function for S-scores (**Figure 2.6B**). For example, a 95% confidence threshold is associated with E-MAP S-scores less than -4 or greater than 5, while the median p-value of scaled SGA scores is less than 0.05 for scores less than -5 or greater than 3. We ultimately elected to use a significance cutoff of absolute S-score greater than 3. This threshold corresponds to an estimated confidence value of 0.83 for S-scores less than -3 and 0.65 for S-scores greater than 3. We compared these values to the intermediate significance threshold recommended for the SGA data from Costanzo et al,[65] which was p-value < 0.05 and absolute SGA score > 0.08. After scaling to E-MAP format, this threshold corresponds to scaled S-scores less than -2.97 or greater than 2.25, below our chosen threshold of -3 and 3.

GI profile correlation measurements

Of the 1444 library genes in the *GSP1* point mutant GI profile map, 1129 were present in the SGA dataset from Costanzo et al.[65] Pairwise Pearson correlation coefficients were computed between all *GSP1* point mutants and SGA gene profiles, and all profiles trimmed to include only genetic interaction measurements with the 1129 shared library genes. Due to the relative sparsity of GI profiles, pairwise comparisons are dominated by high numbers of non-significant interactions. Accordingly, we did not consider correlations with *GSP1* point mutants or SGA gene profiles that did not have significant genetic interactions (absolute scaled S-score greater than 3, see above) with at least 10 of the 1129 library genes. This requirement removed all weak Gsp1 point mutants and one strong mutant (R108A) from the correlation analysis (as they had at most nine genetic interactions with absolute score greater than 3), leaving 22 strong mutants and 3370

S. cerevisiae SGA alleles to be included in the correlation analysis. All Pearson correlations and their p-values between Gsp1 mutants and *S. cerevisiae* genes, including all correlations that did not pass our significance filtering procedures, are provided in **Supplementary Data 2**. The subset of Pearson correlations between Gsp1 point mutants and Gsp1 partners with available co-complex X-ray crystal structures, used to make the point plots in **Figure 2.1E** and **Figure 2.7C, D**, are also available in **Table 2.4**.

Statistical significance of correlations was computed using both two-sided and one-sided (positive) t-tests adjusted for multiple hypothesis testing using both the Bonferroni method and the FDR method, which controls the false discovery rate.[133] All p-values reported in the text and figures are one-sided (positive) and corrected by the FDR method, unless otherwise stated. The FDR method of p-value correction has been shown to account for the positive dependency between test statistics, such as those arising from the underlying functional similarities between *S. cerevisiae* alleles.[134] Custom code for GI profile correlation calculations and filtering is provided in the accompanying repository https://github.com/tinaperica/Gsp1_manuscript/tree/master/Scripts/E-MAP/correlations.

Significance testing was used to filter out *S. cerevisiae* gene SGA profiles that did not show a significant correlation (one-sided positive, Bonferroni-adjusted) with the GI profiles of at least two *GSP1* point mutants. In total, 276 *S. cerevisiae* alleles from the SGA had a significant GI profile correlation (one-sided positive, Bonferroni-adjusted) with at least two *GSP1* point mutants and were therefore included in the correlation analysis shown in **Figure 2.17A**. We required alleles to correlate with at least two mutants because the goal of this analysis was to group mutants by similarity, and an allele that only significantly correlated with one mutant is uninformative for this task. After this filtering step, the one-sided p-values were used to populate a matrix of 22 mutants

vs. 276 alleles, and hierarchical clustering was performed using Ward's method. We used Ward's method rather than the average linkage criterion as we found the latter resulted in a wide variety of group sizes due to a few sparsely populated outliers. Using Ward's methods resulted in rounder clusters, allowing us to identify meaningful functional groups of mutants and alleles. Pearson correlation between correlation vectors was used as a distance metric for the mutant (row) clustering, while Euclidean distance was selected for the gene (column) clustering, due to the column vectors being relatively short (22 mutants per column vs. 276 alleles per row) and thus sensitive to outliers when clustered using Pearson correlations as the distance metric (for additional analysis of E-MAP statistics and clustering see the **Supplementary Note**).

For the gene set analysis, we decreased the stringency of inclusion of *S. cerevisiae* SGA genes to include all alleles with a significant GI profile correlation (one-sided positive, Bonferroni-adjusted) with one or more Gsp1 mutants, which added another 201 alleles, resulting in 477 alleles. We made the gene sets larger to increase our confidence in connecting the patterns of correlations between *S. cerevisiae* genes and Gsp1 mutants to the GTPase cycle parameters represented in **Figure 2.17B, D**. Indeed, while *S. cerevisiae* genes that only correlate significantly with one mutant are not informative for grouping mutants, they are informative for annotating the functional effects of individual mutants. Manually curated gene sets of *S. cerevisiae* genes with significant correlations with Gsp1 mutants are provided in **Supplementary Data 4**.

Protein expression levels by Western Blot

S. cerevisiae strains were grown at 30°C in YPD medium (20 g Bacto™ Peptone (CAT # 211820, BD Diagnostic Systems), 10 g Bacto™ Yeast Extract (CAT # 212720 BD), and 20 g Dextrose (CAT # D16-3, Fisher Chemicals) per 1 L medium) for 1.5 - 2 hours until OD600 reached 0.3. Cell culture aliquots of 1 ml were centrifuged for 3 minutes at ~ 21,000 x g and resuspended

in 30 µl of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH = 7.4) and 10 µl of SDS-PAGE Sample Buffer (CAT # 161-0747, BioRad), to a final SDS concentration of 1%, and ~ 2mM beta-mercaptoethanol. Lysates were run (3 µl for most, and 6 µl for slow growing mutants with lower OD_{600}) on Stain-Free gels (4-20%, CAT #4568096, BioRad, Tris/Glycine SDS Buffer (CAT #161-0732, BioRad)). After electrophoresis, the gel was scanned for total protein quantification and the proteins were subsequently transferred to an Immobilon-FL PVDF membrane (CAT #IPF00010, EMD Millipore). The membrane was probed with Rabbit anti-RAN (CAT # PA 1-5783, ThermoFisher Scientific) primary, and Goat anti-Rabbit-IgG(H+L)-HRP (CAT #31460, ThermoFisher) secondary antibodies. The membrane was developed using Super Signal West Femto substrate (CAT # 34096, ThermoFisher), and scanned and analyzed with Image Lab software on a ChemiDoc MP (BioRad). Each blot had at least one wild-type (WT-GSP1-clonNAT) and at least one MAT:α strain control. The total protein levels (TP^{MUT}) for each Gsp1 point mutant lane were then normalized to the wild-type (WT-GSP1-clonNAT) lane of the corresponding blot (TP^{WT}), providing an adjustment value to account for differences in loading between lanes ($a^{MUT} = TP^{MUT}/TP^{WT}$). To compute the relative expression of a Gsp1 point mutant, the density (D^{MUT}) of the Western blot bands corresponding to the Gsp1 point mutant was divided by the total protein adjustment and finally normalized against the same value for the wild-type Gsp1, i.e. $rel.expression^{MUT} = \frac{D^{MUT}/a^{MUT}}{D^{WT}/a^{WT}}$. Note that for blots with a single WT lane, $a^{WT} = 1$. For blots with more than one WT lane included, a^{WT} was computed for each WT lane by normalizing to the average *TP* across all WT lanes, and the average adjusted WT density (D^{WT}/a^{WT}) across all WT lanes was used for computing the relative

expression of point mutants. An example Western blot is provided in **Figure 2.22**, and the final protein expression level data for all mutants are shown in **Figure 2.4**.

Physical interaction mapping using affinity purification mass spectrometry (AP-MS)

S. cerevisiae cell lysate preparation

When choosing mutants for AP-MS we sought to cover all Gsp1 sequence positions where mutations had strong GI profiles (**Figure 2.7A**), as well as several 'weak' mutants. We observed that tagging the endogenous Gsp1 with either an amino-terminal or a carboxy-terminal FLAG tag affects the *S. cerevisiae* growth in culture. We therefore attempted to make each of the mutants intended for AP-MS experiments with both tags, and where both tags were viable, we obtained the AP-MS data for both. We could not make a FLAG-tagged R108Q mutant for AP-MS. *S. cerevisiae* strains for AP-MS were grown in YAPD medium (120 mg adenine hemisulfate salt (CAT # A9126, SIGMA), 10 g Bacto yeast extract (CAT # BD 212720), 20 g Bacto peptone (CAT # BD 211820), 20 g dextrose (D-glucose D16-3 Fisher Chemicals) per 1 L of medium). Each strain was grown at 30ºC for 12 to 24 h to OD_{600} of 1-1.5. The cells were harvested by centrifugation at 3000 RCF for 3 minutes and the pellet was washed in 50 ml of ice-cold ddH_2O, followed by a wash in 50 ml of 2x lysis buffer (200 mM HEPES pH 7.5, 200 mM KCl, 2 mM $MgCl_2$, 30 µM GTP (Guanosine 5′-triphosphate sodium salt hydrate, CAT #G8877, Sigma-Aldrich), 1 mM Dithiothreitol (Promega V3151), 0.1% IGEPAL CA-630 (CAT # I8896, Sigma-Aldrich), and 10% glycerol). Each pellet of approximately 500 µl was then resuspended in 500 µl of 2X lysis buffer supplemented with protease inhibitors without EDTA (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail, CAT # 11836170001, Roche) and dripped through a syringe into liquid

nitrogen. The frozen *S. cerevisiae* cell pellets were lysed in liquid nitrogen with a SPEX™ SamplePrep 6870 Freezer/Mill™.

FLAG immunoprecipitation

FLAG immunoprecipitations were performed as previously described.[135,136] Details are as follows. For FLAG immunoprecipitations, frozen samples were initially kept at room temperature for 5 minutes and then placed on ice or at 4°C in all subsequent steps, unless indicated otherwise. Following the addition of 1.5 – 3.0 ml Suspension Buffer (0.1 M HEPES pH 7.5, 0.1 M KCl, 1 mM $MgCl_2$, 15 µM GTP, and 0.5 mM Dithiothreitol) supplemented with cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche), samples were incubated on a rotator for at least 10 minutes and then adjusted to 6.0 ml total volume with additional Suspension Buffer supplemented with inhibitors before centrifugation at 18,000 rpm for 10 minutes. Anti-FLAG M2 Affinity Gel beads (50 µl slurry; Sigma-Aldrich) were washed twice with 1.0 ml Suspension Buffer. After reserving 50 µl, the remaining supernatant and anti-FLAG M2 Affinity Gel beads were combined and incubated for >= 2 hours on a tube rotator. Beads were then collected by centrifugation at 300 rpm for 5 minutes and washed three times. For each wash step, beads were alternately suspended in 1.0 ml Suspension Buffer and collected by centrifugation at 2,000 rpm for 5 minutes. After removing residual wash buffer, proteins were eluted in 42 µl 0.1 mg/ml 3xFLAG peptide, 0.05% RapiGest SF Surfactant (Waters Corporation) in Suspension Buffer by gently agitating beads on a vortex mixer at room temperature for 30 minutes. Immunoprecipitated proteins (~4 µl) were resolved on 4-20% Criterion Tris-HCl Precast gels (BioRad) and visualized by silver stain (Pierce Silver Stain Kit; Thermo Scientific) (**Figure 2.23**) before submitting 10 µl of each sample for mass spectrometry. At least three independent biological replicates were performed for each FLAG-tagged protein and the untagged negative control.

Liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis

To prepare samples for LC-MS/MS analysis, immunoprecipitated protein (10 µl) was denatured and reduced in 2 M urea, 10 mM NH_4HCO_3, and 2 mM Dithiothreitol for 30 minutes at 60°C with constant shaking, alkylated in the dark with 2 mM iodoacetamide for 45 minutes at room temperature and digested overnight at 37°C with 80 ng trypsin (Promega). Following digestion, peptides were acidified with formic acid and desalted using C18 ZipTips (Millipore) according to the manufacturer's specifications. Samples were re-suspended in 4% formic acid, 2% acetonitrile solution, and separated by a 75-minute reversed-phase gradient over a nanoflow C18 column (Dr. Maisch). Peptides were directly injected into a Q-Exactive Plus mass spectrometer (Thermo), with all MS1 and MS2 spectra collected in the orbitrap. Raw MS data were searched against the *S. cerevisiae* proteome (SGD sequences downloaded January 13, 2015) using the default settings in MaxQuant (version 1.5.7.4), with a match-between-runs enabled.[137,138] Peptides and proteins were filtered to 1% false discovery rate in MaxQuant, and identified proteins were then subjected to protein-protein interaction scoring using SAINTexpress.[139] Protein were filtered to only those representing high confidence protein-protein interactions (Bayesian false discovery rate from SAINT (SAINT BFDR) < 0.05). Protein abundance values for this filtered list were then subjected to equalized median normalization, label free quantification and statistical analysis were performed using MSstats,[140] separately for data from amino- or carboxy-terminally tagged baits. Fold change in abundance of preys for 3xFLAG-tagged Gsp1 point mutants was always calculated compared to the wild-type Gsp1 with the corresponding tag. All AP-MS data are available from the PRIDE repository under the PXD016338 identifier. Fold change values between prey abundance between the mutant and wild-type Gsp1 and the corresponding FDR adjusted p-values are provided in **Supplementary Data 3**. The intersection of all prey proteins identified at least

once with both the amino- or carboxy-terminal 3xFLAG tag, and their interquartile ranges (IQR) of log_2-transformed fold change values across all the Gsp1 mutants, are provided in **Table 2.5**. Quality of data and reproducibility between replicates was assessed based on correlations of protein abundance between replicates (**Figure 2.24, Figure 2.25**).

Biochemical and biophysical assays

Protein purifications

All proteins were expressed from a pET-28 a (+) vector with a N-terminal 6xHis tag in *E. coli* strain BL21 (DE3) in the presence of 50 mg/L Kanamycin for 2xYT medium, and 100 mg/L of Kanamycin for autoinduction EZ medium. GEF (Srm1 from *S. cerevisiae*, (Uniprot P21827)) was purified as Δ1-27Srm1 and GAP (Rna1 from *S. pombe*, Uniprot P41391) as a full-length protein (for use of *S. pombe* Rna1 see the **Supplementary Note**). ScΔ1-27Srm1 and SpRna1 were expressed in 2xYT medium (10 g NaCl, 10 g yeast extract (BD Bacto™ Yeast Extract #212720), 16 g tryptone (Fisher, BP1421) per 1 L of medium) overnight at 25 ºC upon addition of 300 μmol/L Isopropyl-β-D-thiogalactoside (IPTG). Gsp1 variants were expressed by autoinduction for 60 hours at 20ºC.[141] The autoinduction medium consisted of ZY medium (10 g/L tryptone, 5 g/L yeast extract) supplemented with the following stock mixtures: 20xNPS (1M Na_2HPO_4, 1M KH2PO4, and 0.5 M $(NH_4)_2SO_4$), 50x 5052 (25% glycerol, 2.5% glucose, and 10% α-lactose monohydrate), 1000x trace metal mixture (50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM each of $MnCl_2$ and $ZnSO_4$, 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 imidazole, and 2 mM β-mercaptoethanol using a microfluidizer from Microfluidics. For Gsp1 purifications, the lysis buffer was also supplemented with 10 mM $MgCl_2$. The His-tagged proteins were purified on Ni-NTA resin

(Thermo Scientific #88222) and washed into a buffer containing 50 mM Tris (pH 7.5) and 100 mM NaCl, with 5 mM $MgCl_2$ for Gsp1 proteins. The N-terminal His-tag was digested at room temperature overnight using up to 12 NIH Units per mL of bovine thrombin (Sigma-Aldrich T4648-10KU). 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). Storage buffer for Gsp1 proteins was supplemented with 5 mM $MgCl_2$. Protein concentrations were confirmed by measuring at 10-50x dilution using a Nanodrop (ThermoScientific). The extinction coefficient at 280 nm used for nucleotide (GDP or GTP) bound Gsp1 was 37675 M^{-1} cm^{-1}, as described in.[142] The ratio of absorbance at 260 nm and 280 nm for purified Gsp1 bound to GDP was 0.76. Extinction coefficients for other proteins 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 ºC.

In our hands every attempt to purify the *S. cerevisiae* homologue of GAP (Rna1, Uniprot P11745) from *E. coli* yielded a protein that eluted in the void volume on the Sephadex 200 size exclusion column, indicating that the protein is forming soluble higher-order oligomers. We were, however, successful in purifying the *S. pombe* homologue of GAP (Rna1, Uniprot P41391) as a monomer of high purity as described above, and we used the purified *S. pombe* homolog of Rna1 in all of our GTP hydrolysis kinetic experiments. Although we cannot exclude slight differences between the kinetic parameters of *S. pombe* and *S. cerevisiae* Rna1, we do not believe such differences would significantly affect our conclusions for two main reasons: First, residues in the interface with Gsp1 are highly conserved between *S. pombe* and *S. cerevisiae* GAP Rna1,

suggesting that mechanism of catalysis and kinetic parameters are also likely to be similar. *S. pombe* and *S. cerevisiae* Rna1 proteins have an overall 39% sequence identity and 53% sequence similarity. Importantly, all but one interface core residues are identical in sequence between *S. cerevisiae* and *S. pombe* homologues (**Figure 2.26**). The X-ray crystal structure of Ran GTPase and its GAP used in our analyses is a co-complex structure of the *S. pombe* homolog of Rna1 (PDB: 15kd), human Ran, and human RanBP1 (**Table 2.1**). Second, we rely only on the *relative differences* between GAP kinetic parameters of different Gsp1 mutants to group our mutants into three classes. Even in the case of differences between the absolute kinetic parameters between the *S. pombe* and *S. cerevisiae* GAP Rna1, the order of mutants is less likely to be different, and even in the case of some differences, we expect the grouping to be robust to these changes (see the **Supplementary Note** for more detail).

Circular dichroism (CD) spectroscopy of protein thermostability

Samples for CD analysis were prepared at approximately 2 μM Gsp1 in 2 mM HEPES pH 7.5, 5 mM NaCl, 200 μM $MgCl_2$, and 50 μM Dithiothreitol. CD spectra were recorded at 25 °C using 2 mm cuvettes (Starna, 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 (**Figure 2.27**) and is reported in **Table 2.9**.

GTP loading of Gsp1

Gsp1 variants for GTPase assays as well as for ^{31}P NMR spectroscopy were first loaded with GTP by incubation in the presence of 20-fold excess GTP (Guanosine 5′-Triphosphate, Disodium Salt, CAT # 371701, Calbiochem) in 50 mM Tris HCl pH 7.5, 100 mM NaCl, 5 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 either NMR buffer or the GTPase assay buffer using NAP-5 Sephadex G-25 DNA Grade columns (GE Healthcare # 17085301). We were unable to obtain sufficient material for some mutants (H141E/I, Y148I), for which we collected AP-MS data, since these mutants precipitated during the nucleotide exchange process at the high concentrations required for ^{31}P NMR, possibly because of the limited stability of nucleotide-free Ran/Gsp1 generated during exchange, as noted previously.[143]

Reverse phase high performance liquid chromatography (HPLC)

Analysis of bound nucleotide was performed using reverse-phase chromatography as previously described[142] using a C18 column (HAISIL TS Targa C18, particle size 5 µm, pore size 120 Å, dimensions 150 x 4.6 mm, Higgins Analytical # TS-1546-C185). The column was preceded by a precolumn filter (The Nest Group, Inc, Part # UA318, requires 0.5 µm frits, Part # UA102) and a C18 guard column (HAICart SS Cartridge Column, HAISIL Targa C18, 3.2x20 mm, 5µm, 120 Å Higgins Analytical # TF-0232-C185, requires a Guard Holder Kit, Higgins Analytical # HK-GUARD-FF). To prepare the nucleotide for analysis, a Gsp1 sample was first diluted to a concentration of 25-30 µM and a volume of 40 µl. The protein was denatured by addition of 2.5 µl of 10% perchloric acid ($HClO_4$). The pH was raised by addition of 1.75 µl 4 M sodium acetate (CH_3COONa) pH 4.0. The nucleotide was separated from the precipitated protein before

application to the column by spinning at 20,000 x g for 20 minutes. 30 µl of supernatant was withdrawn and mixed 1:1 with reverse-phase buffer (10 mM tetra-n-butylammonium bromide, 100 mM KH_2PO_4 / K_2HPO_4, pH 6.5, 0.2 mM NaN_3). 20 µl of sample was injected onto the equilibrated column and run isocratically in 92.5% reverse-phase buffer, 7.5% acetonitrile at a flow rate of 1 ml/min for 35 min (~20 column volumes). Nucleotide retention was measured by monitoring absorbance at both 254 nm and 280 nm. Example HPLC reverse phase chromatogram of GTP-loaded wild-type Gsp1 is shown in **Figure 2.28**.

NMR Spectroscopy

Gsp1 samples for ^{31}P NMR spectroscopy were first loaded with GTP as described above, and buffer exchanged into NMR Buffer (D_2O with 50 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$, 2 mM Dithiothreitol). Final sample concentrations were between 250 µM and 2 mM, and 400 µl of sample were loaded into 5 mm Shigemi advanced microtubes matched to D_2O (BMS-005TB; Shigemi Co. Ltd, Tokyo, Japan.). ^{31}P NMR experiments were performed on a Bruker Avance III 600 MHz NMR spectrometer with a 5 mm BBFO Z-gradient Probe. Spectra were acquired and processed with the Bruker TopSpin software (version 4.0.3). Indirect chemical shift referencing for ^{31}P to DSS (2 mM Sucrose, 0.5 mM DSS, 2 mM NaN_3 in 90% H_2O + 10% D_2O; water-suppression standard) was done using the IUPAC-IUB recommended ratios.[144] Spectra were recorded at 25°C using the pulse and acquire program zg (TopSpin 3.6.0), with an acquisition time of 280 milliseconds, a recycle delay of 3.84 seconds, and a 65° hard pulse. *4,096 complex points were acquired over the course of 4,096 scans and a total acquisition time of 4.75 hours. Spectra were zero-filled once and multiplied with an exponential window function (EM) with a line-broadening of 6 Hz (LB = 6) prior to Fourier transformation. Peaks were integrated using the auto-integrate function in TopSpin 4.0.7, and peak areas were referenced to the bound GTP-β peak of

each spectrum. The peak at approximately -7 ppm is defined as γ1 and the peak at approximately -8 ppm is defined as γ2. The percent of γ phosphate in γ2 is defined as a ratio of areas under the curve between the γ2 and the sum of the γ1 and γ2 peaks.

Kinetic measurements of GTP hydrolysis.

Kinetic parameters of the GTP hydrolysis reaction were determined using a protocol similar to one previously described.[145] 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, Thermo Fisher) upon binding of the free phosphate GTP hydrolysis product (excitation at 425 nm, emission at 457 nm). All experiments were performed in GTPase assay buffer (40 mM HEPES pH 7.5, 100 mM NaCl, 4 mM $MgCl_2$, 1 mM Dithiothreitol) at 30°C in 100 µl reaction volume on a Synergy H1 plate reader from BioTek, using Corning 3881 96-well half-area clear-bottom non-binding surface plates. The phosphate sensor at 20 µM and 50 µM concentrations was calibrated with a range of concentrations of K_2HPO_4 using only the data in the linear range to obtain a conversion factor between fluorescence and phosphate concentration. 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, and the rest of the data were linearly fit to estimate intrinsic GTP hydrolysis rate (**Table 2.8**). Although we do estimate the intrinsic hydrolysis rates from the background data, the estimate is only approximate, as well as 10^5 to 10^6 lower than the rate of GAP-mediated GTP hydrolysis, which is why we do not use intrinsic hydrolysis rates when fitting the GAP-mediated hydrolysis data. The affinity of Rna1 for GDP-bound Ran is negligible (K_d of

100 μM for Ran:GDP,[146] which is ~250-fold weaker than the estimated K_m for GAP-mediated GTP hydrolysis) and was not taken into account when fitting the data.

As the estimated K_m for the GAP-mediated hydrolysis for many of the Gsp1 variants was low (in the 0.1-0.4 μM range, resulting in difficulties to reliably measure hydrolysis at low substrate concentrations), we sought to estimate the kinetic parameters (k_{cat} and K_m) by directly analysing the full reaction progress curve with an analytical solution of the integrated Michaelis-Menten equation (see section below for details).

Estimating the k_{cat} and K_m parameters of GAP-mediated hydrolysis using an accurate solution to the integrated Michaelis-Menten equation.

Others (e.g. Goudar *et al*[147]) have shown that both k_{cat} and K_m can be estimated with reasonable accuracy/precision from a single time-course with *initial* [S] > K_m by directly analyzing the *full* reaction progress curve with an analytical solution of the *integrated* Michaelis-Menten equation based on the Lambert ω function. This analysis is possible because the full reaction progress curve is characterized by an initial linear phase for [S] > K_m, a final exponential phase for [S], and a transition phase for [S] ~ K_m. Whereas k_{cat} is sensitive to the slope of the initial linear phase (i.e. the initial velocity), K_m is sensitive to the shape of the progress curve, which will have an extended linear phase if K_m << initial [S] or no linear phase if K_m >> initial [S]. Use of the integrated Michaelis-Menten analysis requires the experiment to be set up with the following conditions: (i) [Gsp1:GTP_0] > K_m, (ii) [GAP_0] <<< [Gsp1:GTP_0], and (iii) the reaction time course F(t) is measured to completion (i.e. until it approaches equilibrium). Our experiments were all set up to fulfill those conditions, which means that the F(t) sampled a concentration range from [Gsp1:GTP] (at t = 0) > K_m to [Gsp1:GTP] (at t = final time) << K_m. The entire F(t) can then be directly analyzed by a non-linear fit with the analytical solution for the integrated Michaelis-Menten equation. As

the initial linear phase of the time course is well measured, k_{cat} can be well determined. As the exponential phase and transition region of the time course are also well measured, the maximum likelihood value of K_m can also be determined.

Specifically, each time course was fitted to an integrated Michaelis Menten equation:

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

where $[E]_t$ 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 (GTP loaded Gsp1), 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, as previously described,[147] where

$$\omega = Lambert\ omega(\frac{[S]_0}{K_m}\ e^{\frac{[S]_0 - k_{cat}[E]_t * time}{K_m}}).$$

The curves were fit with the custom-made software DELA.[148] Examples of full reaction progress curves and their integrated Michaelis-Menten fits are shown in **Figure 2.13**.

We confirmed that the kinetic value parameters we obtained for wild-type Gsp1 using the phosphate sensor and integrated Michaelis-Menten equation were similar to those estimated using intrinsic tryptophan fluorescence.[149] Their values were a K_m of 0.45 μM and k_{cat} of 2.1 s^{-1} at 25˚C for mammalian Ran hydrolysis activated by *S. pombe* GAP, while our values for wild type *S. cerevisiae* Gsp1 and *S. pombe* GAP at 30˚C are K_m of 0.38 μM and k_{cat} of 9.2 s^{-1}.

For most mutants a concentration of 1 nM GAP (SpRna1, Rna1 from *S. pombe*) was used. In order to run the time courses to completion, for mutants with low k_{cat}/K_m enzyme concentrations of 2-5 nM were used. Initially we collected time course data for all Gsp1 variants at approximately

8 μM concentration of loaded Gsp1:GTP with 1 nM GAP and 20 μM phosphate sensor. If the estimated K_m was higher than 1 μM, we repeated the time course kinetic experiments with higher concentration of Gsp1:GTP of approximately tenfold above the K_m.

To quantify the accuracy of parameter (k_{cat}, K_m) estimation for GAP-mediated GTP-hydrolysis by the integrated Michaelis Menten approach over a range of kinetic parameters and substrate concentrations [Gsp1:GTP] we simulated data covering the range of parameters estimated for all of our Gsp1 point mutants, and estimated the accuracy of parameters determined given the Gaussian noise similar to our experimental data. The largest standard deviations were 3%, 17%, and 18% for k_{cat}, K_m, and k_{cat}/K_m, respectively (**Figure 2.29**). In addition, we analyzed how the χ^2 statistic changed as the Michaelis Menten parameters were systematically varied around the estimated maximum likelihood values (**Figure 2.30**). For these analyses, the k_{cat} or K_m values were independently fixed and incremented while the remaining parameters were fit to generate χ^2 surfaces for one degree of freedom. Confidence intervals (CIs) for which χ^2 increased by 4.0 compared to the maximum likelihood minimum were estimated by linear interpolation after iterative bisection. A χ^2 increase of 4.0 corresponds to the 95% confidence limit for a normal distribution. The k_{cat}/K_m ratio and corresponding χ^2 values were derived from the analyses with systematic variation of either k_{cat} or K_m. CIs for k_{cat}/K_m were estimated by linear interpolation without iterative bisection. The χ^2 surfaces approach a parabolic shape with a well-defined minimum at the maximum likelihood value. The CIs are further consistent with the parameter ranges obtained from the simulations. Thus, both the simulations and χ^2 surfaces indicate that k_{cat} and K_m are estimated with reasonable accuracy over the range of parameter values and experimental conditions used in this study.

The Michaelis Menten k_{cat} and K_m parameters and their standard deviations were calculated from at least three technical replicates from two or more independently GTP-loaded Gsp1 samples (**Table 2.6**). For more details on the kinetic analysis see the **Supplementary Note.**

Kinetic measurements of Srm1 mediated nucleotide exchange.

Kinetic parameters of GEF mediated nucleotide exchange were determined using a fluorescence resonance energy transfer (FRET) based protocol.[149] Each Gsp1 variant was purified as a Gsp1:GDP complex, as confirmed by reverse phase chromatography. Nucleotide exchange from GDP to mant-GTP (2'-(or-3')-O-(N-Methylanthraniloyl) Guanosine 5′-Triphosphate, CAT # NU-206L, 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. Each time course was measured in GEF assay buffer (40 mM HEPES pH 7.5, 100 mM NaCl, 4 mM $MgCl_2$, 1 mM Dithiothreitol) with excess of mant-GTP. The affinity of Ran/Gsp1 is estimated to be 7-11-fold lower for GTP than for GDP,[143] and for most variants of Gsp1 we measured time courses at Gsp1:GDP concentrations ranging from 0.25 to 12 μM with an excess mant-GTP concentration of 200 μM. For Gsp1 variants with high K_m values that had to be measured at concentrations of up to 200 μM we used an excess of 1000 μM mant-GTP. In addition, we fit the data using a combination of fits following the approach of Klebe.[149] For concentrations of substrate (Gsp1:GDP) that were much lower than the excess of mant-nucleotide (200 μM) we used a combination of two exponential decays, and for reactions with high concentrations of Gsp1, where the relative excess of mant-nucleotide was lower, we always estimated the initial rates using linear fits to the very beginning of the reaction, when levels of mant-nucleotide-bound Gsp1 are very low and therefore exchange is overwhelmingly from Gsp1-GDP to Gsp1-mant-nucleotide.

All kinetic measurements were done at 30ºC in 100 µl reaction volume using 5 nM GEF (Δ1-27Srm1), except for higher concentrations of the mutants with high K_m values that were measured at 20 nM GEF. Data were collected in a Synergy H1 plate reader from BioTek, using Corning 3686 96-well half-area non-binding surface plates. For low concentrations of Gsp1:GDP the time course data were fit to a combination of two exponential decays:

$$Y = span1 * \exp(k_{exchange} * Time) + span2 * \exp(k_{background} * Time) + fluorescence_{plateau}$$

where $k_{exchange}$ is the rate constant of the GDP to mant-GTP exchange, $k_{background}$ is the rate constant of background decay due to photo-bleaching, and span1 and span2 are the fluorescence amplitudes for the two processes. For high concentrations of substrate, or for mutants with very low rates, the initial velocity was determined by a linear fit to the initial 10-20% of the data. As the intrinsic exchange rate in the absence of GEF is estimated to be more than 10^4 lower[143] we do not use the intrinsic rate for fitting the data. The kinetic parameters of the nucleotide exchange were determined by fitting a Michaelis-Menten equation to an average of 38 data points (ranging from 17 to 91) per Gsp1 point mutant for a range of substrate concentrations from [Gsp1:GDP] = 0.25 µM to [Gsp1:GDP] >> K_m. Michaelis-Menten fits are shown in **Figure 2.14**. Michaelis-Menten k_{cat} and K_m parameters for GEF-mediated nucleotide exchange are provided in **Table 2.7**. The errors of the k_{cat} and the K_m parameters were determined from the standard error of the exponential fit of the Michaelis-Menten equation to the data. The error of the catalytic efficiency (k_{cat}/K_m) was calculated by adding the standard errors of the individual parameters and normalizing it for the values of the parameters $\left({}^{kcat}/_{Km}\sqrt{\left({}^{std.error(kcat)}/_{kcat}\right)^2 + \left({}^{std.error(Km)}/_{Km}\right)^2}\right)$. All custom code for fitting and analysis of kinetics data is provided in the accompanying repository

(https://github.com/tinaperica/Gsp1_manuscript/tree/master/Scripts/kinetics). For more details on the kinetic analysis see the **Supplementary Note.**

Supplementary Note

Linkage criteria used for E-MAP hierarchical clustering analyses

For clustering of Gsp1 mutants and E-MAP library genes (**Figure 2.1C**, **Figure 2.5, Figure 2.7**) we used average linkage to be consistent with how we and others have clustered and represented genetic interaction (GI) data in previously published E-MAP datasets (as detailed in Braberg et al[150]). Even though our data are based on screens of point mutants of a single protein, different from most previous studies that screen knockouts of many different genes, we show that average linkage remained an appropriate criterion for clustering our E-MAP matrix based on the recovery of known groups of functionally related genes within the dendrogram of library genes (**Figure 2.1C**, **Figure 2.7B**).

The clustering analysis in **Figure 2.17A** had the goal of assigning the Gsp1 mutants by functional similarity to classes in an unbiased manner, and to assess whether the classes of mutants matched the grouping defined by the *in vitro* kinetics and NMR data. To quantify functional similarity, we adopted the widespread approach of computing correlation coefficients between GI profiles. Most studies have represented these data as networks and used existing annotations (typically Gene Ontology categories) to assert functional groupings[65] but we sought to use an unsupervised clustering approach instead. To do so, we used Ward's linkage criterion, since it was designed to build hierarchies by selecting joining operations that minimize within-group dispersion[151] to find compact, spherical clusters. Indeed, we found Ward's linkage resulted in rounder clusters reflecting known biological functions, and these clusters were less sensitive to

sparsely populated outliers. In contrast, we found the average linkage criterion to be more sensitive to a few sparsely populated outliers (resulting in a variety of group sizes).

Nonetheless, to show that the linkage method used does not alter our primary conclusions regarding the grouping of mutants, we compared the clustering from average linkage with the clustering from Ward's method in the dendrograms (**Figure 2.31**). Both methods identify the three main classes of mutants (I, II, and III below, I, III, and IV above), but average linkage is more sensitive to the sparsely populated vectors and outliers, resulting in a wider variety of cluster sizes.

Potential dependencies between alleles when computing genetic interaction (GI) profile correlations

We use Pearson correlations in three cases: (1) as a distance metric for clustering the Gsp1 E-MAP matrix (**Figure 2.1C**, **Figure 2.5**, **Figure 2.7B**), (2) for quantifying the functional similarity of GI profiles of Gsp1 point mutants and *S. cerevisiae* alleles (**Figure 2.1E**, **Figure 2.17ABD**, **Figure 2.18**), and (3) as a distance metric for clustering the vectors of Gsp1 mutant correlations (**Figure 2.17A**). In all three cases, there are certainly dependencies between some *S. cerevisiae* alleles, as evidenced by their own clustering into groups according to their biological function (**Figure 2.1C**): mRNA export genes cluster together, meaning that if one mRNA export gene has a large negative S-score with a Gsp1 mutant, other mRNA export genes are likely to as well (relevant for case 1). Likewise, if the GI profile of a gene is significantly correlated with a Gsp1 mutant, other genes in the same pathway are likely to have correlated profiles as well (relevant for cases 2 and 3). This dependency is expected and is indeed a main benefit of a GI profiling approach, as the S-scores allow us to infer functional relationships between genes and ascribe likely functions to unknown genes.

With regards to Case 1, we note that assessing similarity of GI profiles using Pearson correlations without further correction for dependencies between alleles is a standard analysis.[65,66]

Case 2 is the only analysis for which we compute statistical significance when using correlations. We accounted for the dependencies between alleles by adjusting our p-values to control the False Discovery Rate, which has been shown to be valid when this form of dependency (positive regression dependency) exists between test statistics.[134] In plots showing correlations between genetic interaction profiles of Gsp1 mutants and *S. cerevisiae* alleles, we use these corrected p-values instead of correlation values for simplicity, because, as can be seen in **Figure 2.32**, only positive correlations of above 0.1 have significant p-values.

With regards to Case 3, most GI studies have used the correlations between GI profiles to define edge attributes for graphical representations of GI networks. We elected to keep the data in matrix form and cluster it to identify functionally similar groups of mutants and *S. cerevisiae* alleles in an unbiased fashion. To cluster the Gsp1 vectors of p-values (columns), we used Pearson correlations as a distance metric. To cluster the *S. cerevisiae* alleles (rows), we used the Euclidean distance instead of the Pearson correlation (as stated in the *GI profile correlation measurements* section of the Methods) because the vectors were only 22 entries long and many were sparse, making them especially sensitive to outliers when using Pearson correlation as the distance metric. To test whether the use of Pearson correlations for the clustering of mutant vectors significantly changes our clustering, we re-clustered the matrix in **Figure 2.17A** using the Spearman correlation or the Euclidean distance as distance metrics instead. While there are slight differences in the ordering of mutants using these different distance metrics, the grouping of mutants is very similar to the original heatmap in **Figure 2.17A** in that it identifies a GAP-perturbed group of mutants, a GEF-perturbed group of mutants, and an intermediate group (**Figure 2.33)**. Thus, we believe this

analysis robustly identifies three functional classes of Gsp1 mutants regardless of any effect that dependencies between the *S. cerevisiae* alleles may have on the Pearson correlations.

Robustness of the analysis to leaving out data

Sub-sampling EMAP data.

We randomly subsampled the library genes in the Gsp1 E-MAP (**Figure 2.1C**) and found that similar groupings of mutants were maintained down to 60% of the library (**Figure 2.34**). We also sub-sampled the *S. cerevisiae* alleles in the correlation p-value matrix (**Fig. 2.17A**) and found that the groupings of mutants were maintained down to 50% (**Figure 2.35**).

Withholding mutants

We performed a computational analysis where we withheld each of our mutants from the analysis one at a time, perform the clustering of genetic interaction profiles for the remaining data as in **Figure 2.17A**, and then assign the withheld mutant to the group whose centroid is most correlated with the mutant. In 21/22 cases, the withheld mutant had the highest correlation with the centroid of its original group (**Figure 2.36**, dark bars). This analysis confirms the robustness of our analysis and addresses the question whether our model would be capable of placing a new mutant not included in the analysis into the correct category.

Potential caveats associated with using the GAP (Rna1) from S. pombe

Our GAP-mediated GTP hydrolysis kinetics experiments used the wild type and mutant Gsp1 from *S. cerevisiae*, but Rna1 GAP from *S. pombe*. We chose to use the Rna1 ortholog from *S. pombe* as *S. cerevisiae* Rna1 formed soluble aggregates after purification, and *S. pombe* Rna1 was the only RanGAP for which there was a structure in complex with Ran (PDB IDs: 1K5D and 1K5G). While there could be slight differences between the kinetic parameters of *S. pombe* and *S.*

cerevisiae GAP Rna1 acting on Gsp1, we do not believe these differences would significantly affect our conclusions, based on the following considerations:

Sequence conservation between S. cerevisiae and S. pombe Rna1.

A sequence alignment between *S. cerevisiae*, *S. pombe,* and human GAP proteins shows that all but one interface core residue in the PDB file 1K5D is conserved in sequence between *S. cerevisiae* and *S. pombe* (**Figure 2.26**). Overall, out of the 1290 $Å^2$ buried by *S. pombe* Rna1 upon interface formation with Ran (PDB 1K5D), 997 $Å^2$ (77%) are buried by residues that are conserved in sequence between *S. pombe* and *S. cerevisiae*, and the sequence identity of the Rna1 interface with Ran/Gsp1 (including all residues that change solvent accessible surface area upon complex formation) overall is 71% (**Table 2.1**).

Comparable kinetic parameters to the human Ran/RanGAP1 pair.

The kinetic parameters for our *S. cerevisiae* Gsp1 and *S. pombe* Rna1 GAP are comparable to the kinetic parameters for the human Ran and human RanGAP1 reported by Klebe et al.[149] They estimate a K_m of 0.45 μM and k_{cat} of 2.1 s^{-1} for Ran/RanGAP1 at 25˚C, while our values for the wild type *S. cerevisiae* Gsp1 and *S. pombe* Rna1 at 30˚C are a K_m of 0.38 μM and k_{cat} of 9.2 s^{-1}. In addition, it was shown that Rna1 from *S. pombe* can activate the hydrolysis in both human and *S. cerevisiae* Ran/Gsp1 with very similar observed rates of hydrolysis (Figure 4A in Becker et al[152]).

Conclusions are based on relative values between the wild-type Gsp1 and its point mutants.

Although we report the absolute values of the kinetics parameters, when we compare the kinetics parameters with the results from genetic interaction profiling and AP-MS, we always use the relative parameters as compared to the wild type. Based on the sequence conservation and comparable kinetics described above, we expect the relative ordering of mutants to be similar as

well. Importantly, we use the relative kinetic data to group our mutants into three classes. Even in the case of small quantitative differences caused by using the *S. pombe* instead of the *S. cerevisiae* Rna1 GAP, we make the assumption that these differences would not significantly affect this grouping.

Validity of the Michaelis-Menten formalism for GTPases

Michaelis-Menten formalisms have been used for multiple GTPases including Ran,[143] Ras,[153] or Rap.[154] Historically there have been many attempts to formalize the conditions under which the Michaelis-Menten equation to describe enzyme kinetics are valid (as reviewed by Schnell[155]). These conditions have converged on the steady-state approximation or more generally, on the reactant stationary assumption. The formal condition for steady-state approximation is that $t_{[ES]}$ (the time it takes for the steady-state levels of [ES] complex to accumulate) is substantially shorter than $t_{[S]}$ (the time where [S] changes significantly). The formal condition for reactant stationary assumption is that $[S] \approx [S_0]$ during initial build-up of [ES].

The formal condition for validity of the Michaelis-Menten equation can be expressed as:

$$\frac{[E_0]}{K_m + [S_0]} \ll \left(1 + \frac{K}{K_S}\right)\left(1 + \frac{[S_0]}{K_m}\right),$$

where $K = \frac{k_{cat}}{k_{on}}$ and $K_S = \frac{k_{off}}{k_{on}}$, and k_{off} and k_{on} are the rates of [ES] complex formation.[156]

The measured dissociation constant, $K_S = \frac{k_{off}}{k_{on}}$, for the formation of the Ran:GDP:RCC1 complex from Ran:GDP and RCC1, where RCC1 is the human RanGEF, is 0.9 μM,[149] which is approximately the same as the K_m value obtained for the GEF-mediated nucleotide exchange for both *S. cerevisiae* Gsp1 and human Ran. That means than $K \ll K_S$, and so the condition for validity of the Michaelis-Menten equation can be approximated as $\frac{[E_0]}{K_m + [S_0]} \ll \left(1 + \frac{[S_0]}{K_m}\right)$. Since

in all of our GEF experiments $[E_0]$ = 5-20 nM << K_m and $[E_0]$ << $[S_0]$, the condition holds true for the entire range of $[S_0]$ values, both below and above the K_m.

As $\frac{K}{K_S}$ can also be expressed as $\frac{k_{cat}}{k_{off}}$, and the measured k_{off} of human Ran:GTP and RanGAP from *S. pombe* is estimated to be around 150 s^{-1}, while our measured k_{cat} values range from 1 to 10 s^{-1}, as above, $\frac{K}{K_S} \ll 1$ the assumption of steady-state holds true as long as $[E_0]$ << K_m and $[E_0]$ << $[S_0]$, which is the case as we used 1-5 nM GAP in all of our experiments.

Figures

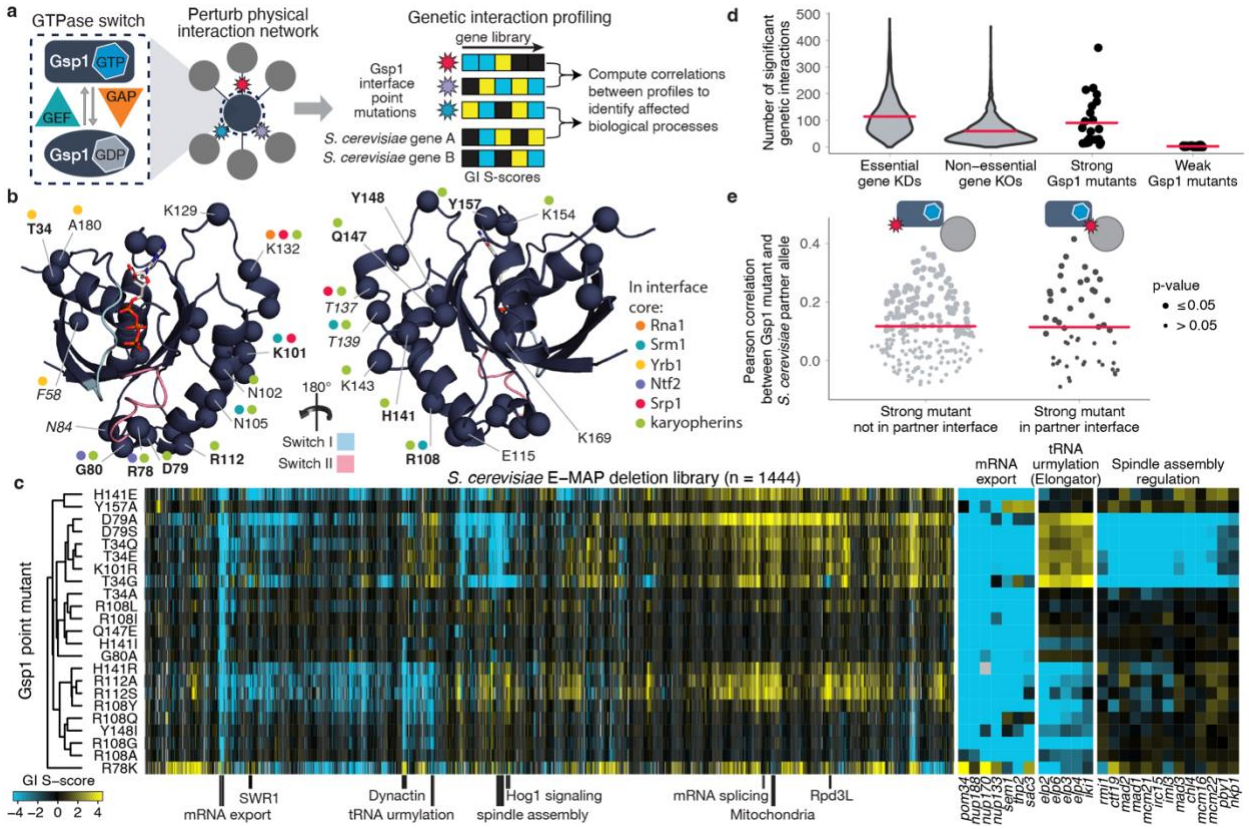


Figure 2.1 GI profiles of Gsp1 interface point mutants cluster by biological processes but not by targeted interfaces.

(a) Interface point mutations enable the probing of biological functions of the multi-specific GTPase switch Gsp1. **(b)** Mutated residue positions shown as Cα atom spheres on the structure of GTP-bound Gsp1. Bold font indicates the positions of mutations with strong GI profiles; italic font indicates the positions not conserved in the sequence between S. cerevisiae and human. Colored dots, interaction partners for which the residue is in the interface core; blue and pink: switch I and switch II regions, respectively. **(c)** GI profiles of 23 GSP1 mutants with nine or more significant GIs, hierarchically clustered by Pearson correlation. A negative S-score (blue) indicates synthetic sick or synthetic lethal GIs; a positive S-score (yellow) indicates suppressive or epistatic GIs. **(d)** Distributions of significant GIs of GSP1 point mutants compared to GIs of mutant alleles of essential and non-essential genes. **(e)** Distributions of Pearson correlations between the GI profiles of Gsp1 interaction partners and GSP1 mutants if mutation is (right, black) or is not (left, grey) in the interface with that partner. Point size corresponds to the false discovery rate (FDR)-adjusted one-sided (positive) P value of the Pearson correlation. Pink bars (d, e), mean.

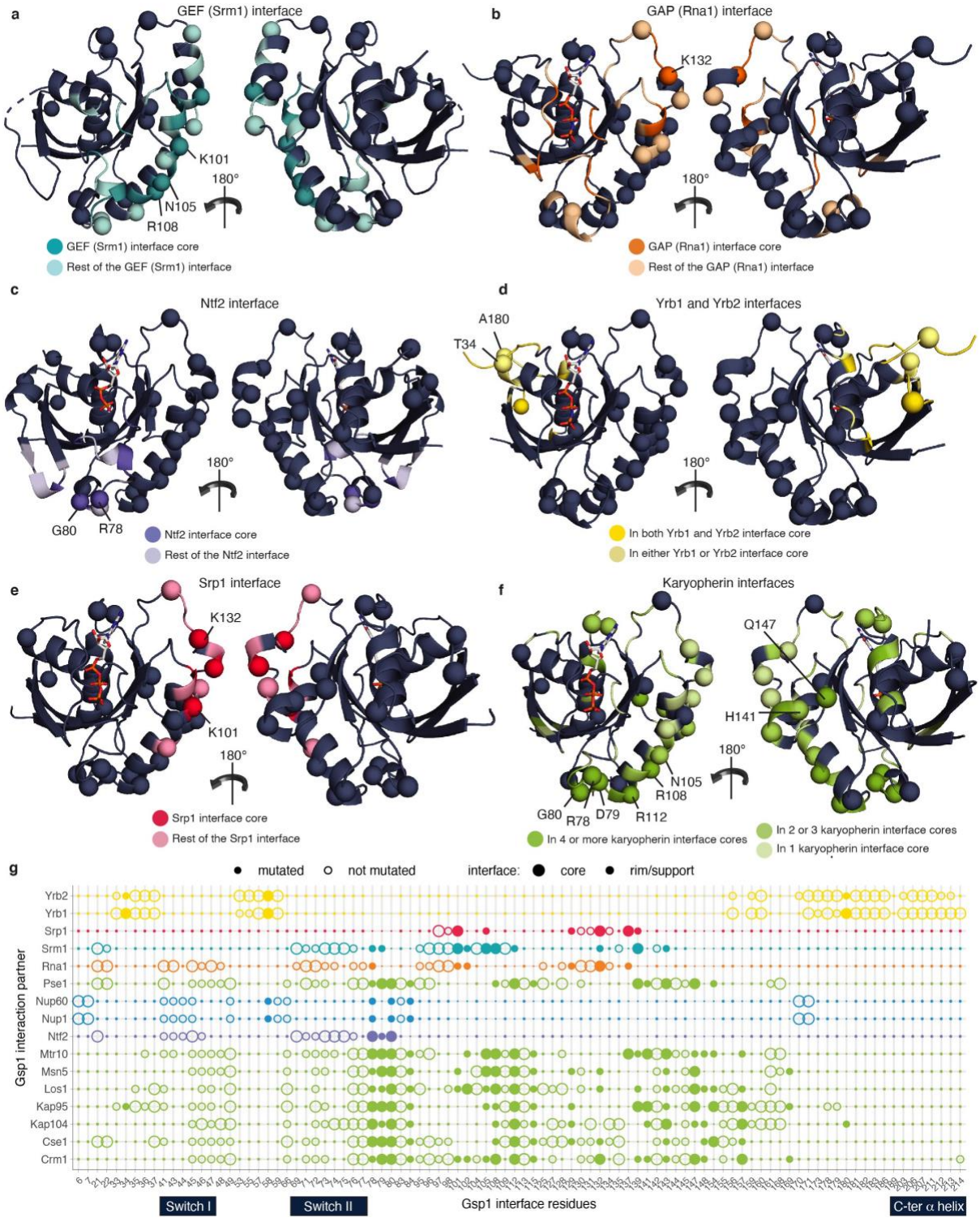


Figure 2.2 Design of interface point mutations in *S. cerevisiae* Gsp1.
Interface residues are categorized as interface core, rim, and support positions (see Supplementary Methods) and provided in Supplementary Table 2. **(a–f)** Structures of RAN (Gsp1) in partner-bound

conformations with interface residues coloured by partner protein. All mutated Gsp1 residues are shown as spheres. **(a)** Srm1 (GEF) interface core (dark teal) and interface rim and support (light teal) PDB 1I2M; **(b)** Rna1 (GAP) interface core (dark orange) and interface rim and support (light orange) PDB 1K5D; **(c)** Ntf2 interface core (dark purple) and interface rim and support (light purple) PDB 1A2K; **(d)** Residues that are in both the core of the Yrb1 and Yrb2 interfaces (dark yellow), and in only one of the two interfaces (light yellow) PDB 1K5D; **(e)** Srp1 interface core (dark pink) and interface rim and support (light pink) PDB 1WA5; **(f)** Residues that are in the core of four or more (dark green), two to three (green) and one (light green) karyopherin interface. Karyopherins are: Kap95, Crm1, Los1, Kap104, Msn5, Cse1, Mtr10. PDB 2BKU. **(g)** Location of Gsp1 residues in partner interfaces. Residues within 5 Å of the nucleotide, in the canonical P-loop, or in the switch I or II regions[126] were not mutated. Residues belonging to the switch I, switch II, and C-terminal α helix are indicated by dark navy bars. Chosen Gsp1 point mutation substitutions are provided in **Table 2.3**.

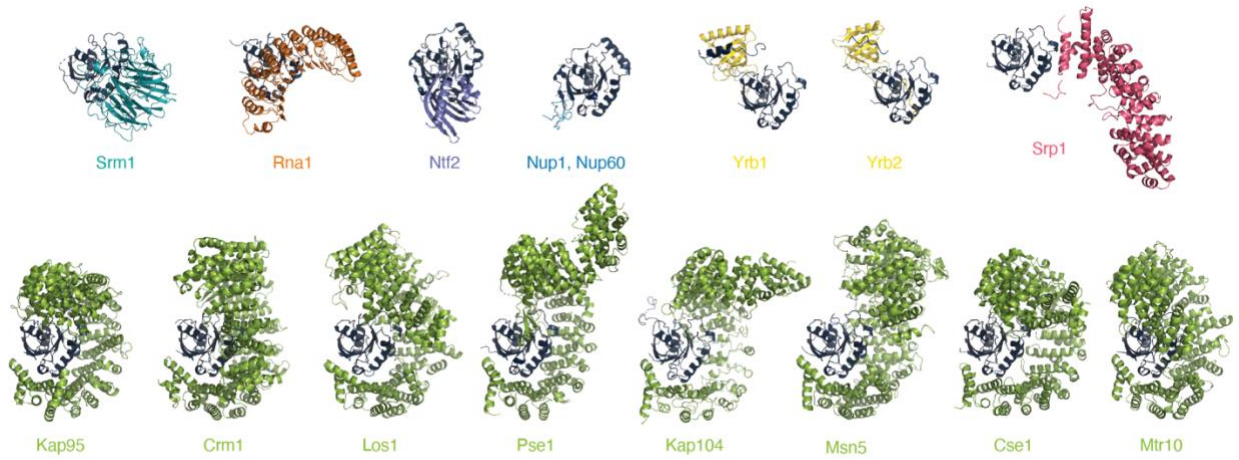


Figure 2.3 Cartoon representation of co-complex structures of *S. cerevisiae* Gsp1 (dark navy) with indicated partners (or homologs).
Srm1 (PDB 1I2M), Rna1 (PDB 1K5D), Ntf2 (PDB 1A2K), Nup1/Nup60 (PDB 3CH5), Yrb1 (PDB 3M1I), Yrb2 (PDB 3WYF), Srp1 (PDB 1WA5), Kap95 (PDB 2BKU), Crm1 (PDB 3M1I), Los1 (PDB 3ICQ), Pse1(PDB 3W3Z), Kap104 (PDB 1QBK), Msn5 (PDB 3A6P), Cse1 (PDB 1WA5), Mtr10 (PDB 4OL0). Species and sequence identity to *S. cerevisiae* homologs for these structures are provided in **Table 2.1**.

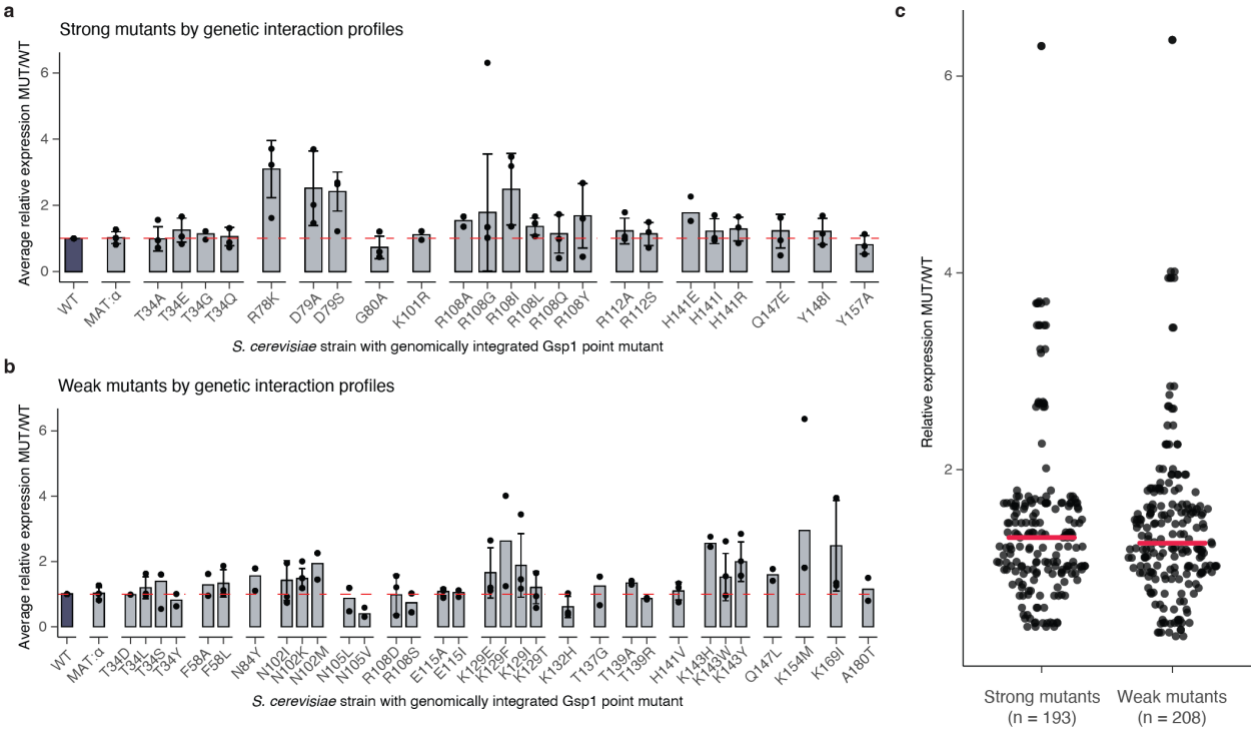


Figure 2.4 Endogenous expression levels of Gsp1 in *S. cerevisiae* strains with genomically integrated GSP1 point mutations profiled by western blot.

(a) Expression data for strong mutants, defined as mutants with nine or more significant GIs. **(b)** Expression data for weak mutants, defined as mutants with fewer than nine significant GIs. In (a) and (b), bar heights indicate averages over two or more biological replicates (n) grown on separate days (except for T34D which has only one biological replicate), with error bars indicating one standard deviation for n >= 3. Overlaid points indicate individual biological replicates (each an average over at least 12 technical replicates per biological replicate for wild-type and MAT:α strains, and between one and six technical replicates per biological replicate for mutant strains). Expression levels are relative to the expression levels of wild-type Gsp1 protein with clonNAT resistance marker (WT) shown as red dashed lines (relative expression of 1). MAT:α is the starting *S. cerevisiae* strain (see Methods). **(c)** Distributions of average relative expression levels for strong and weak mutants. Each point is as in (a) and (b). Horizontal pink bars indicate the mean of the point distributions.

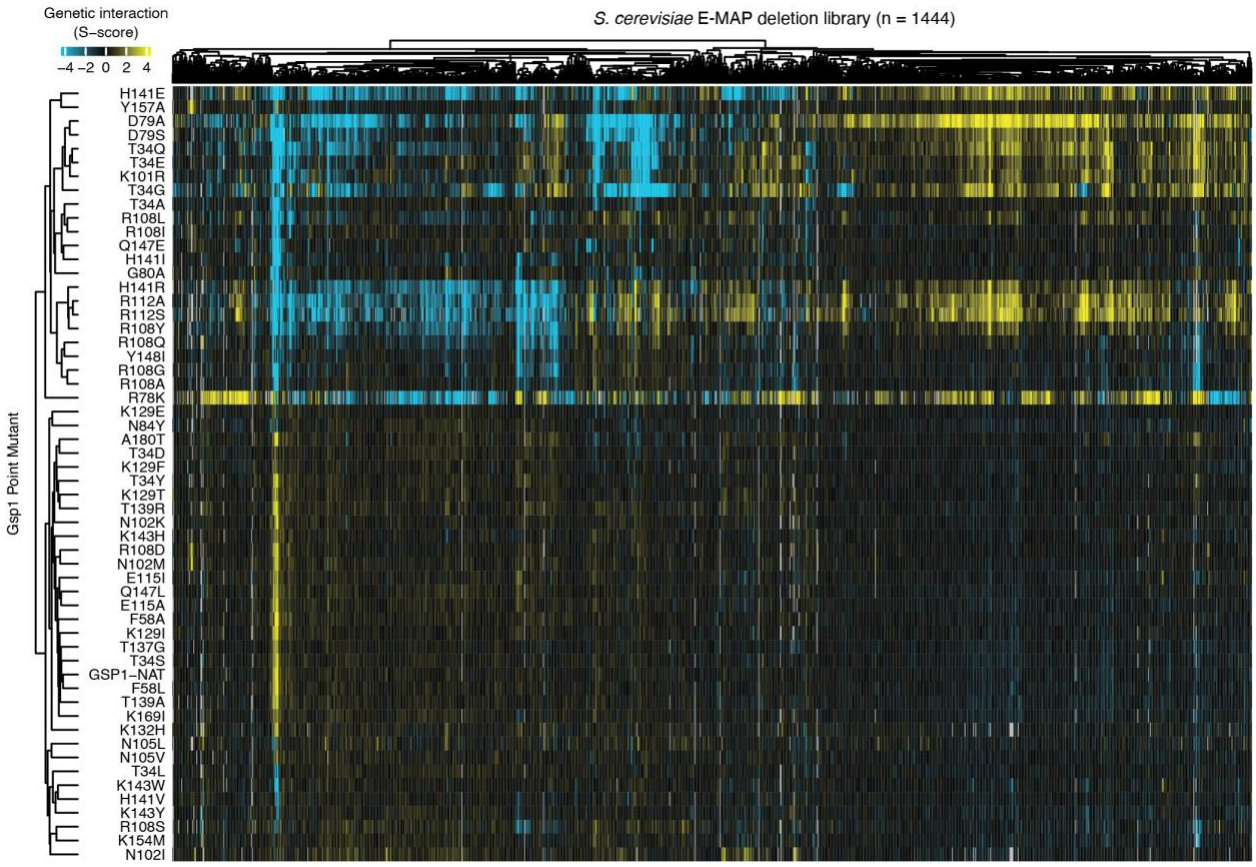


Figure 2.5 GI profiles of the 56 GSP1 strains (wild-type GSP1 with clonNAT cassette and 55 point mutants).

Negative S-score (blue) represents synthetic sick or synthetic lethal GIs, positive S-score (yellow) represents suppressive or epistatic GIs; neutral S-scores (no significant GI) are shown in black. Gsp1 point mutants and *S. cerevisiae* genes are hierarchically clustered by Pearson correlation. GSP1 mutants fall into two clusters: a cluster of 23 strong mutants with nine or more significant GIs and 32 weak mutants with fewer than nine significant GIs.

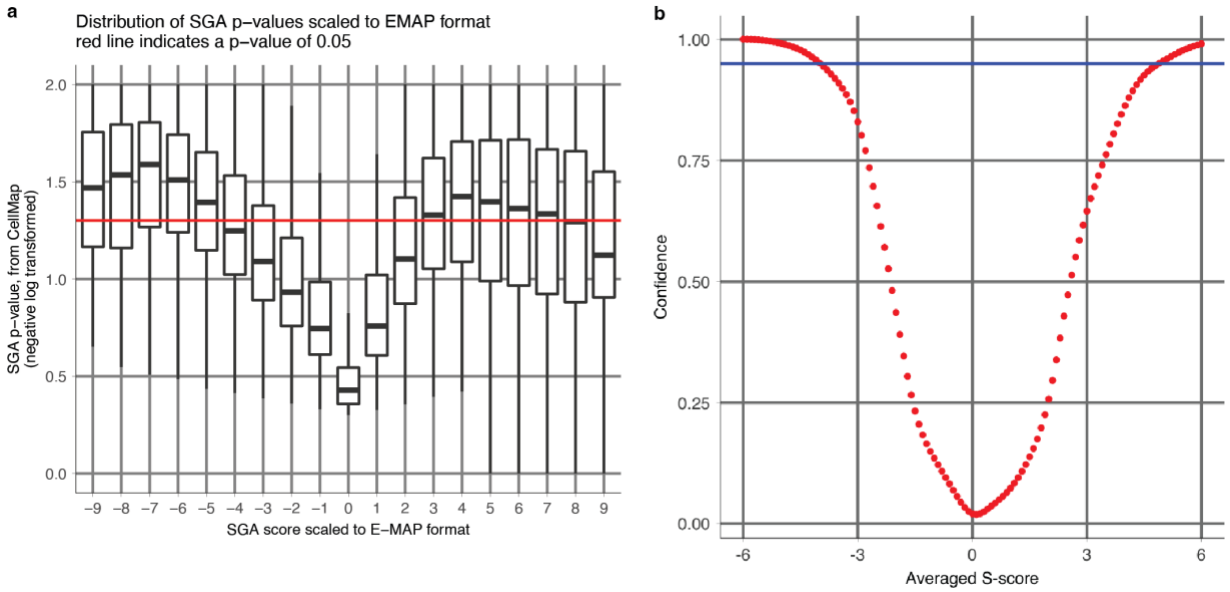


Figure 2.6 Comparison of definitions of high confidence S-scores used in our analysis.
(a) Distribution of the SGA scores scaled to the E-MAP S-scores versus their corresponding published p-values from the CellMap.[65] **(b)** Distribution of the E-MAP S-score averaged from all the individual replicates versus the confidence of the functional genetic interaction reproduced from Collins et al.[63]

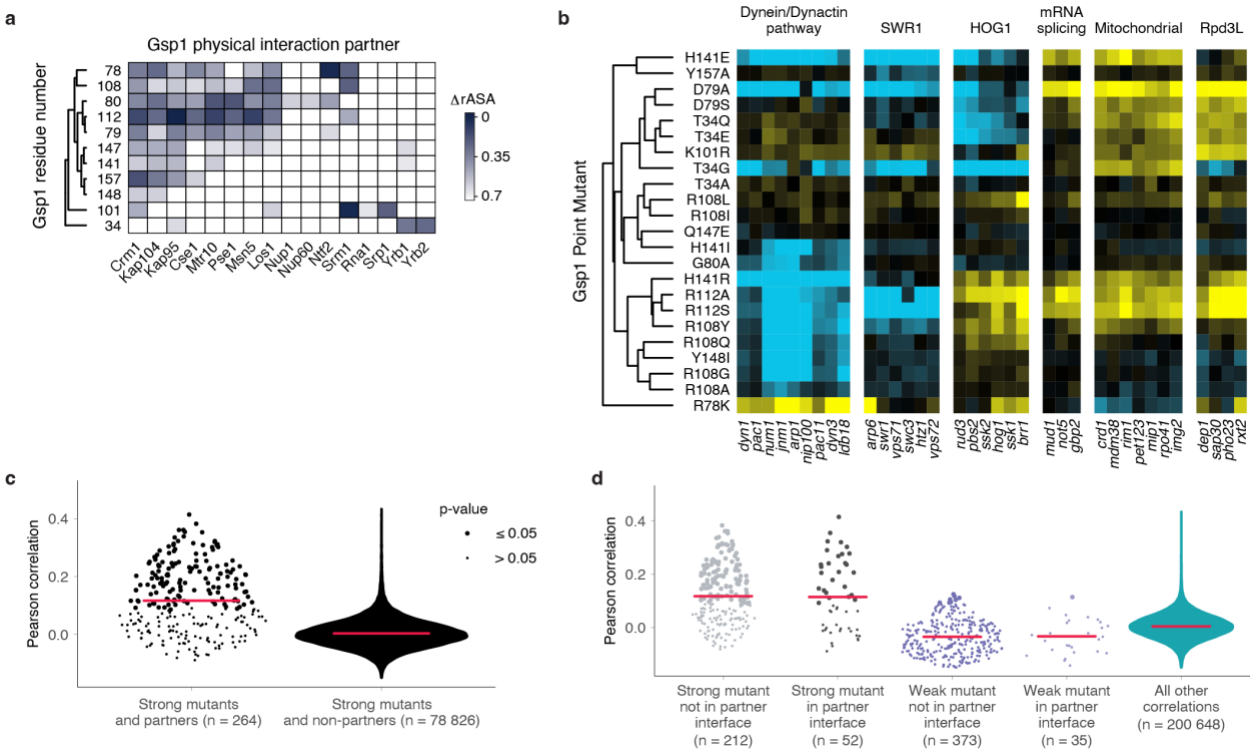


Figure 2.7 Functional profiles of GSP1 mutants cannot be explained solely by the positions of mutations in interfaces.

(a) Locations of mutated residues in structurally characterized interfaces. ΔrASA is the difference in accessible surface area of a residue upon binding, relative to an empirical maximum for the solvent accessible surface area of each amino acid residue type (see Methods). **(b)** GI profiles of *GSP1* mutants group *S. cerevisiae* genes by biological processes and complexes, such as the dynein/dynactin pathway, SWR1 complex, the Hog1 signaling pathway, mRNA splicing, mitochondrial proteins, and the Rpd3L histone deacetylase complex. **(c)** Distributions of Pearson correlations between the GI profiles of strong *GSP1* mutants and alleles of Gsp1 direct interaction partners with available co-complex crystal structures (left) and strong *GSP1* mutants and alleles of all other *S. cerevisiae* genes (right). **(d)** Distributions of Pearson correlations between the GI profiles of Gsp1 interaction partners and strong and weak *GSP1* mutants if mutation is (black and light purple) or is not (grey and dark purple) in the interface with that partner. Teal violin plot on the right represents the distribution of all other Pearson correlations between *GSP1* mutants and *S. cerevisiae* genes. In (c) and (d), point size indicates the false discovery rate adjusted one-sided (positive) p-value of Pearson correlation, and pink bars indicate the mean of the point distributions; n denotes the number of *GSP1* point mutant-gene GI profile correlations in each category. Data for strong mutants are also shown in **Figure 2.1E** and included here for comparison.

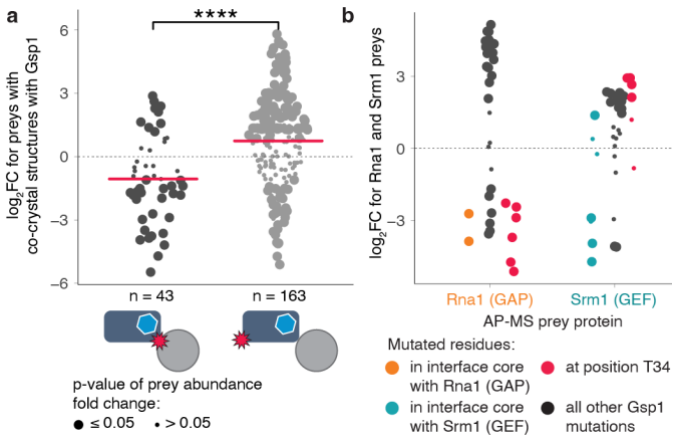


Figure 2.8 Gsp1 interface point mutations rewire the physical interaction network of Gsp1, including interactions with the switch regulators GEF (Srm1) and GAP (Rna1).
Shown is the log_2-transformed fold change ('log_2FC' in the figure) between the abundance of partner proteins pulled down with a Gsp1 mutant versus those pulled down with wild-type Gsp1. **(a)** Change in abundance of partner proteins with crystal structures in complex with Gsp1 (Rna1, Srm1, Yrb1, Kap95, Pse1, Srp1) in which the mutation is (left) or is not (right) in the interface core with the partner. n refers to the number of partner abundance changes in each category. The mean log_2-transformed fold change values (pink bars) are −1 and 0.73, respectively (t-test P value = 1.6 × 10−5). Point size corresponds to the P value of abundance fold change. **(b)** Change in abundance of pulled-down Rna1 and Srm1. Point size as in (a); points colored by interface location.

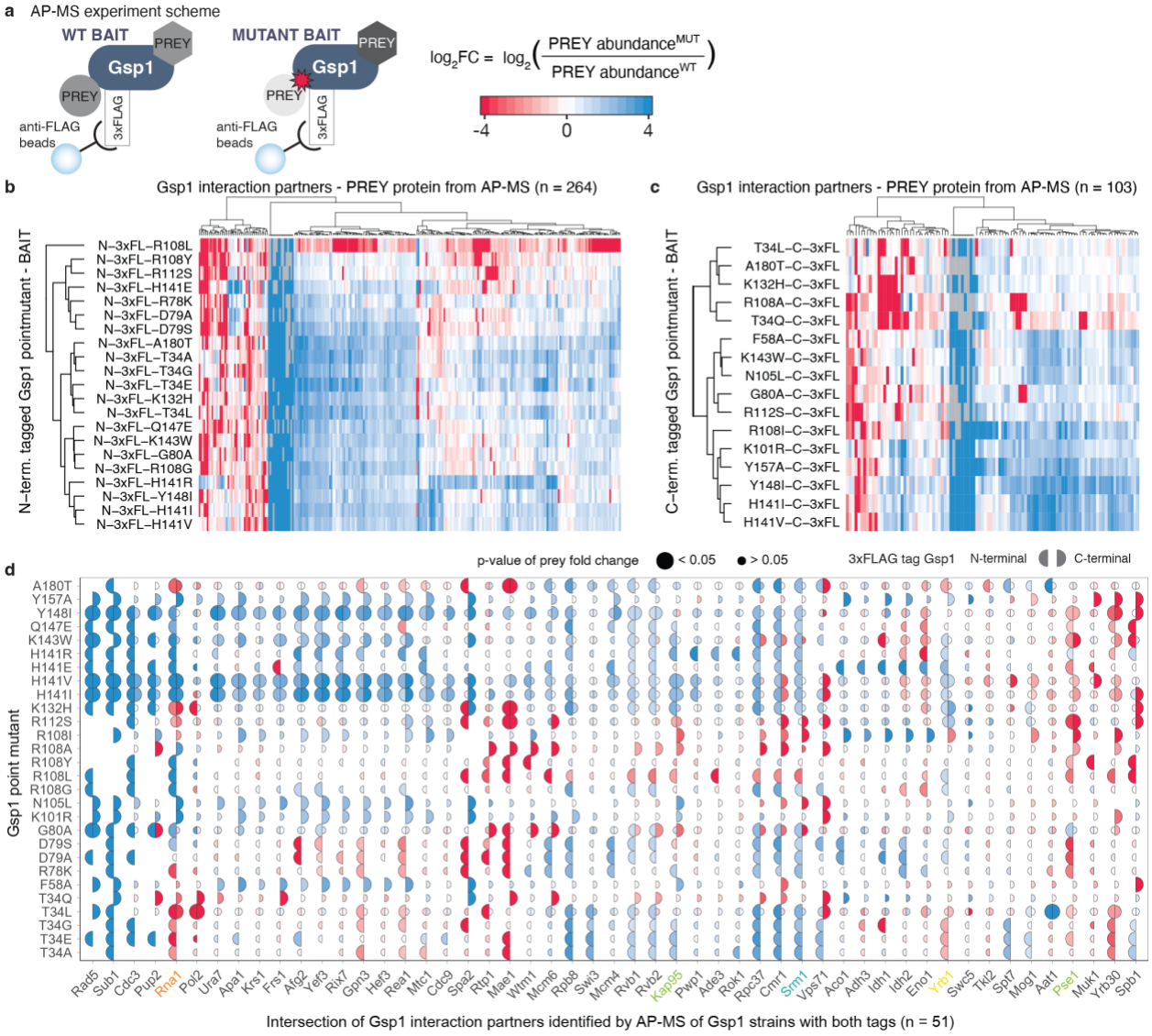


Figure 2.9 Interface point mutations in Gsp1 rewire its physical interaction network.
(a) Schematic representation of the affinity purification mass spectrometry (AP-MS) experiment to determine the abundance of pulled-down protein interaction partners of wild type and mutant Gsp1. The change in abundance of partner proteins pulled down with Gsp1 mutants in (b), (c), and (d) is represented as $\log_2$-transformed fold change (FC) between abundance of a partner pulled-down with a Gsp1 mutant versus pulled-down with wild-type Gsp1 ($\log_2$(abundance(PREY)^{MUT}/abundance(PREY)^{WT}). To account for possible tag effects, the fold change in prey abundance was always computed relative to the wild-type protein with the corresponding tag. Decreased abundance compared to pull-down with wild-type Gsp1 is annotated in red and increased abundance in blue. The $\log_2$-transformed fold change values are capped at +/− 4. **(b)** Amino- and **(c)**, -carboxy terminally 3xFLAG-tagged Gsp1 point mutants (rows) and prey proteins identified by AP-MS (columns) hierarchically clustered by the $\log_2$-transformed fold change in prey abundance. **(d)** Prey proteins pulled down by both amino- and carboxy-terminal tagged constructs. Left semi-circle represents an amino-terminal 3xFLAG-tagged Gsp1 point mutant, and right semi-circle represents carboxy-terminal 3xFLAG-tagged Gsp1 point mutant. Semi-circle size is proportional to the significance of the $\log_2$-transformed fold change (false discovery rate adjusted p-value) of the prey

abundance in pulled-down complexes with a Gsp1 mutant compared to complexes with the wild-type Gsp1. Overall we identified 316 high-confidence prey partner proteins, with the amino- and carboxy-terminally tagged Gsp1 mutants pulling down 264 and 103 preys, respectively, including 51 overlapping preys. The difference in preys identified by experiments with N- or C-terminal tags illustrates the sensitivity of the interaction network to perturbation of Gsp1.

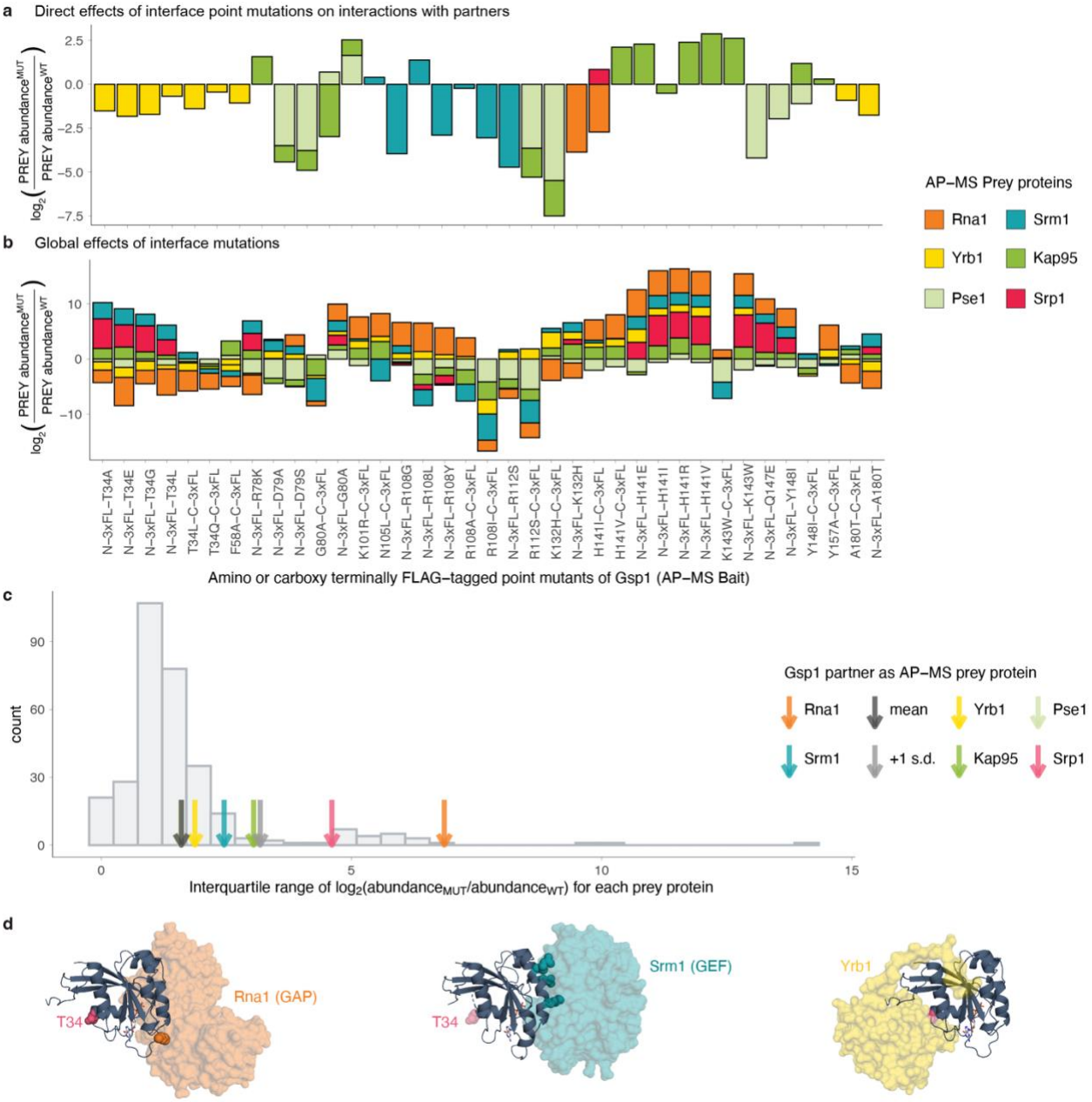


Figure 2.10 Gsp1 interface mutations rewire interactions with the core regulators Srm1 and Rna1. **(a, b)** Protein-protein interactions between interface mutants of Gsp1 and Gsp1 partners for which there are co-complex X-ray crystal structures (core regulators Srm1 and Rna1, and effectors Yrb1, Kap95, Pse1, and Srp1). Change in pulled-down prey partner abundance is expressed as $\log_2$(PREY abundance^{MUT}/PREY abundance^{WT})). N-3xFL and C-3xFL labelled mutants are tagged with an amino- or carboxy-terminal triple FLAG tag, respectively, and partners are colored as indicated. **(a)** Bar plot depicting changes in pulled-down prey partner abundance when the point mutation is in the core of the Gsp1 interface with the prey partner. **(b)** Bar plot depicting all changes in pulled-down prey partner abundance for core regulators Srm1 and Rna1, and effectors Yrb1, Kap95, Pse1, and Srp1, regardless of whether the mutation is directly in the interface core with the partner or not. **(c)** Distribution showing the variation in $\log_2$-transformed fold change in abundance of all prey proteins pulled down with the Gsp1 mutants, as defined by interquartile range

(IQR) across mutants. Values for core partners shown as arrows (Rna1 orange, Srm1 teal, Yrb1 yellow, Kap95 green, Pse1 light green, Srp1 pink). Mean and +1 standard deviation of IQR values are highlighted with a dark grey and a light grey arrow, respectively. The extent to which the abundance of the two cycle regulators Rna1 and Srm1 changed across the Gsp1 point mutants is larger than the change for an average prey protein. All IQR values are provided in **Table 2.5**. **(d)** Position of T34 with respect to the interfaces with Rna1 (GAP, orange surface, PDB 1K5D), Srm1 (GEF, teal surface, PDB 2I1M), and Yrb1 (yellow surface, PDB 1K5D). As the coordinates for T34 are not resolved in the 2I1M structure, in all three structures the pink spheres show the residue location in the aligned 1K5D structure. Gsp1: navy cartoon; GTP nucleotide: stick representation. Residues that were mutated in the Rna1 and Srm1 interfaces are shown in sphere representation and are colored in orange (Rna1, left) or teal (Srm1, middle).

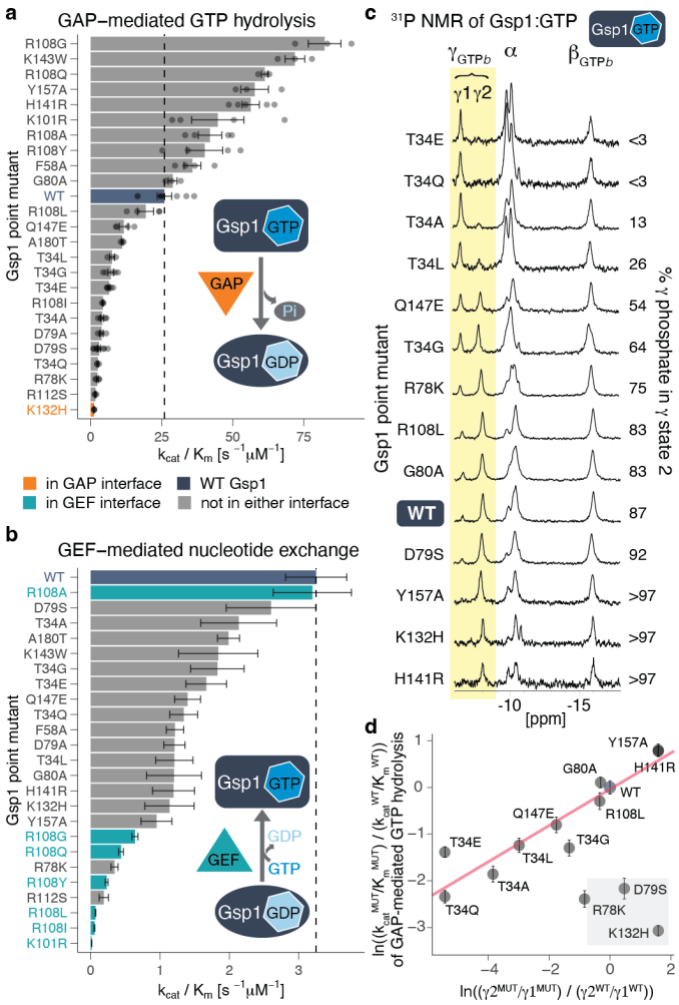


Figure 2.11 Point mutations in Gsp1 interfaces allosterically modulate GTPase cycle parameters by tuning active site conformational distributions.

(a, b) Catalytic efficiency (k_{cat}/K_m) of GAP-mediated GTP hydrolysis (a) or GEF-mediated nucleotide exchange (b) of Gsp1 mutants. Dotted lines indicate the wild-type (WT) efficiency. In (a), points represent k_{cat}/K_m from an individual experiment fit to an integrated Michaelis–Menten equation. Error bars, s.e.m. from n ≥ 3 replicates. In (b), error bars are the s.e.m. of the Michaelis–Menten fit to data from n ≥ 17 measurements at different substrate concentrations. **(c)** ^{31}P NMR of GTP-bound Gsp1 point mutants. NMR peak heights are normalized to the β-peak of bound GTP (β_{GTPb}). The two peaks of the γ-phosphate of bound GTP are highlighted in yellow. **(d)** Natural log (ln)-transformed ratios (mutant to wild-type; MUT/WT) of the exchange equilibrium constants (K_{ex} = population in γ2/population in γ1, assuming a detection limit of 3% for the γ-peak estimation by ^{31}P NMR) plotted against the natural log-transformed ratios (MUT/WT) of the relative catalytic efficiency (k_{cat}/K_m) of GAP-mediated GTP hydrolysis. Error bars, s.e.m. from n ≥ 3 replicates. Pink line, least-squares linear fit, excluding K132H, R78K and D79S (grey box).

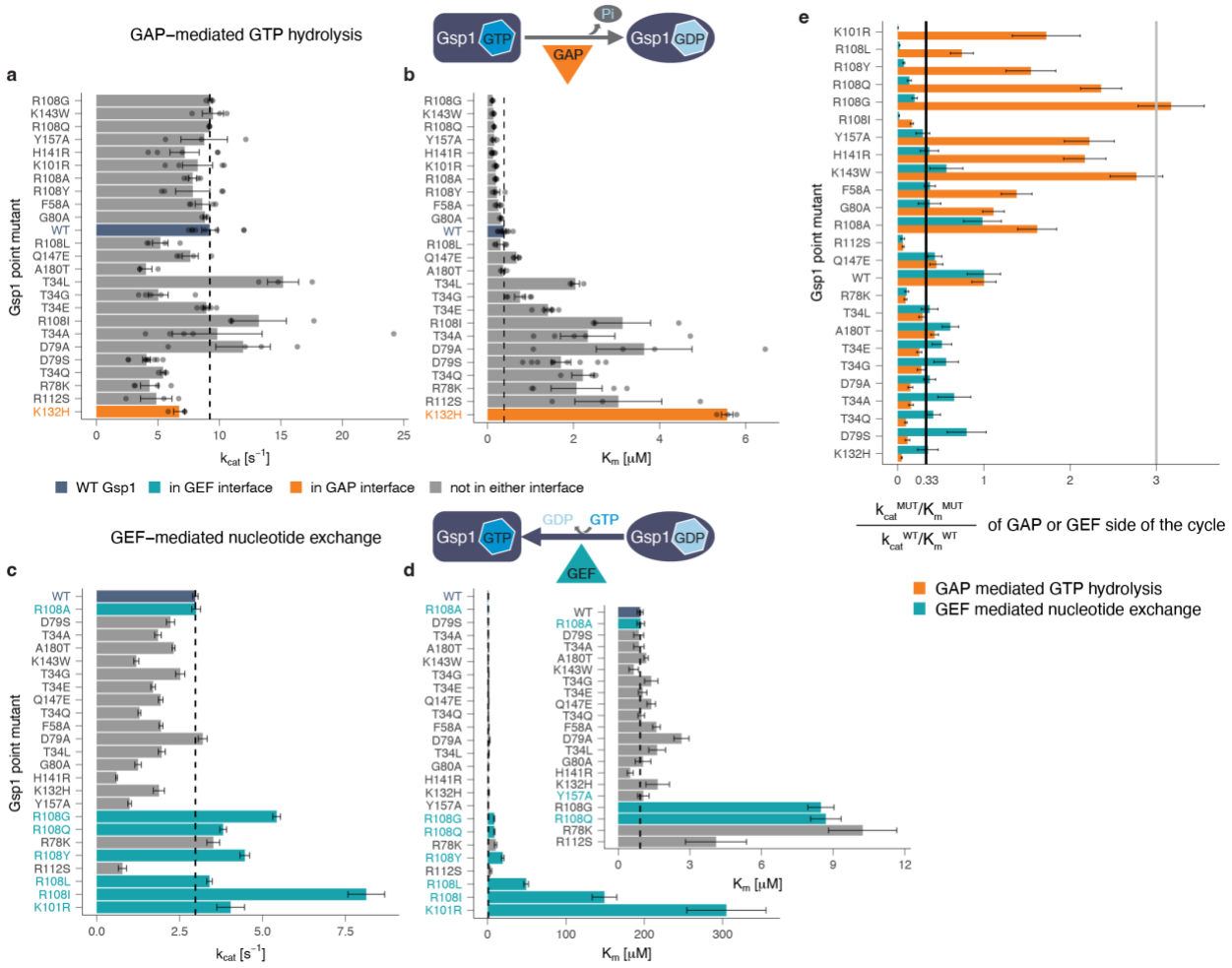
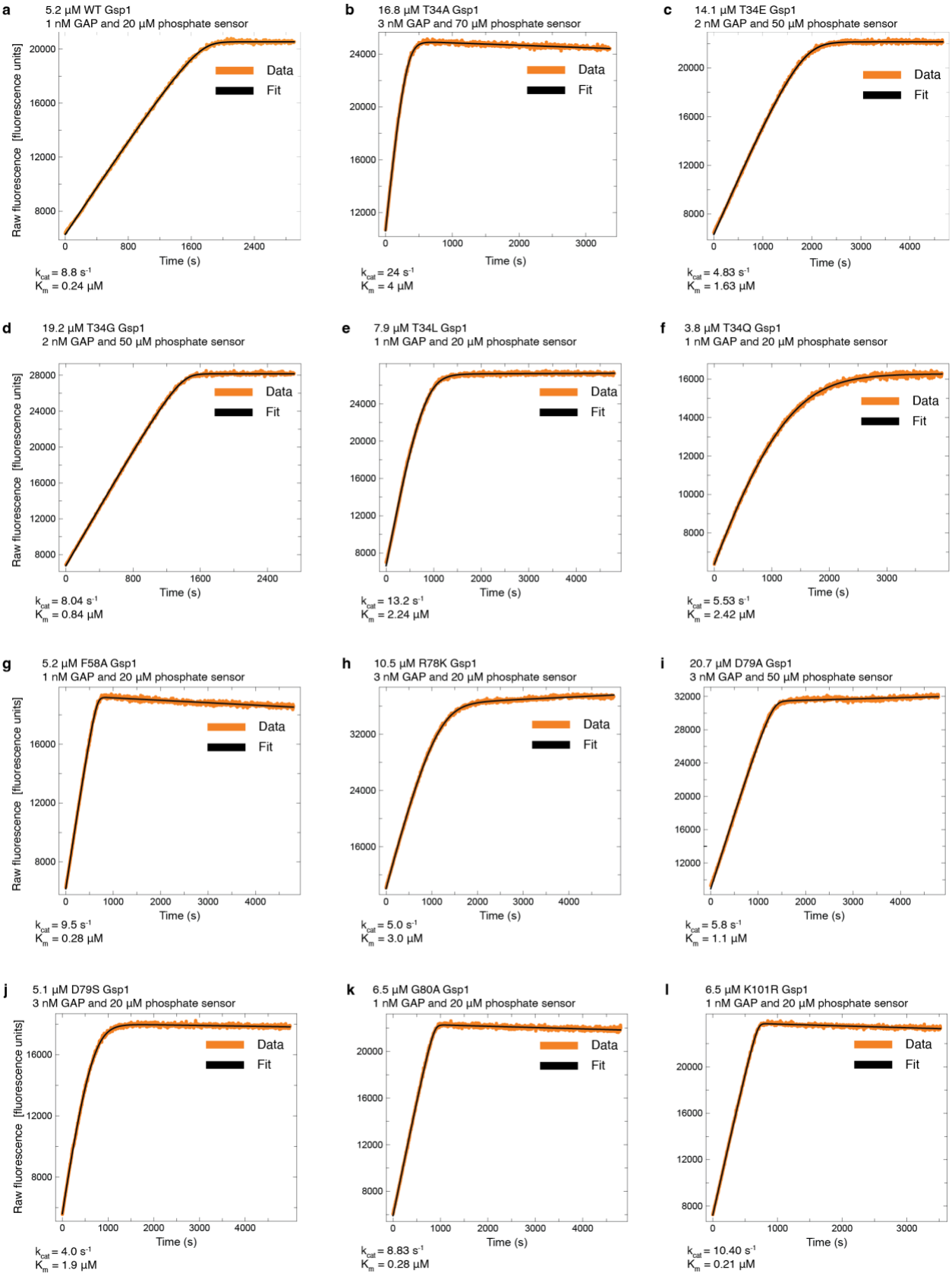
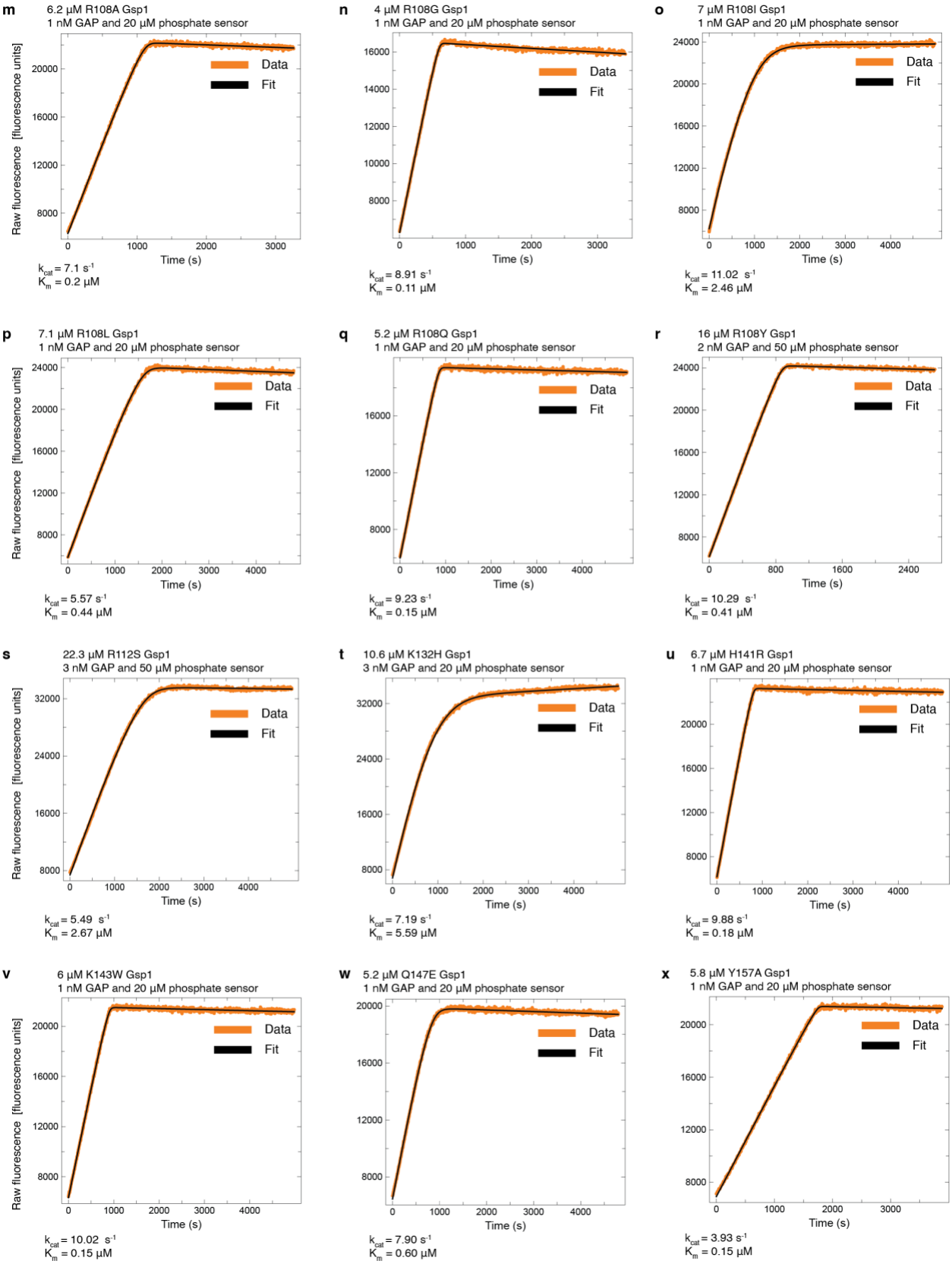


Figure 2.12 Effect of Gsp1 point mutations on the in vitro efficiency of GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange.

(a) k_{cat} and **(b)** K_m values of GAP-mediated GTP hydrolysis of wild-type and point mutant Gsp1. Error bars represent the standard deviation of the k_{cat} and the K_m parameters from the integrated Michaelis-Menten fit for $n \geq 3$ replicates. **(c)** k_{cat} and **(d)**, K_m of GEF-mediated nucleotide exchange of wild-type and point mutant Gsp1. Inset shows the K_m bar plot for all but the four mutants with the highest K_m (K101R, R108L, R108I, and R108Y). Error bars represent the value plus/minus the standard error of the Michaelis-Menten fit to data from $n \geq 17$ measurements at different substrate concentrations. (a-d) Dotted lines indicate the wild-type values. Dark blue bar denotes the wild-type Gsp1, and orange and teal bars highlight the residues that are in the core of the interface with the GAP and GEF, respectively. **(e)** Comparison of relative change in catalytic efficiencies of GAP-mediated GTP hydrolysis (orange bars) and GEF-mediated nucleotide exchange (teal bars) defined as k_{cat}^{MUT}/K_m^{MUT} / k_{cat}^{WT}/K_m^{WT}. Grey line indicates a three-fold increase compared to wild type and black line indicates a three-fold decrease compared to wild type. Error bars represent the added standard error of the mean (for GAP) or standard error of the fit (for GEF) values of the mutant and the wild-type efficiency (k_{cat}/K_m) values. Mutations not in the interface core with the GAP both increased (3-fold, R108G mutant) and decreased (3 to 10-fold, T34E/Q /A/G, R78K, D79S/A, R108I, and R112S mutants) the catalytic efficiency k_{cat}/K_m of GAP-mediated GTP hydrolysis, compared to wild-type Gsp1. As expected, mutations in the interface core with the GEF (K101, and R108) decreased the catalytic efficiency of GEF-mediated nucleotide exchange >40-fold. However, other mutations not in the GEF interface core (R78K, R112S, Y157A) also decreased the efficiency notably (3- to 10-fold).


a
5.2 µM WT Gsp1
1 nM GAP and 20 µM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 8.8 s^{-1}
K_m = 0.24 µM
b
16.8 µM T34A Gsp1
3 nM GAP and 70 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 24 s^{-1}
K_m = 4 µM
c
14.1 µM T34E Gsp1
2 nM GAP and 50 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 4.83 s^{-1}
K_m = 1.63 µM
d
19.2 µM T34G Gsp1
2 nM GAP and 50 µM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 8.04 s^{-1}
K_m = 0.84 µM
e
7.9 µM T34L Gsp1
1 nM GAP and 20 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 13.2 s^{-1}
K_m = 2.24 µM
f
3.8 µM T34Q Gsp1
1 nM GAP and 20 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 5.53 s^{-1}
K_m = 2.42 µM
g
5.2 µM F58A Gsp1
1 nM GAP and 20 µM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 9.5 s^{-1}
K_m = 0.28 µM
h
10.5 µM R78K Gsp1
3 nM GAP and 20 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 5.0 s^{-1}
K_m = 3.0 µM
i
20.7 µM D79A Gsp1
3 nM GAP and 50 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 5.8 s^{-1}
K_m = 1.1 µM
j
5.1 µM D79S Gsp1
3 nM GAP and 20 µM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 4.0 s^{-1}
K_m = 1.9 µM
k
6.5 µM G80A Gsp1
1 nM GAP and 20 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 8.83 s^{-1}
K_m = 0.28 µM
l
6.5 µM K101R Gsp1
1 nM GAP and 20 µM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 10.40 s^{-1}
K_m = 0.21 µM


m
6.2 μM R108A Gsp1
1 nM GAP and 20 μM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 7.1 s^{-1}
K_m = 0.2 μM
n
4 μM R108G Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 8.91 s^{-1}
K_m = 0.11 μM
o
7 μM R108I Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 11.02 s^{-1}
K_m = 2.46 μM
p
7.1 μM R108L Gsp1
1 nM GAP and 20 μM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 5.57 s^{-1}
K_m = 0.44 μM
q
5.2 μM R108Q Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 9.23 s^{-1}
K_m = 0.15 μM
r
16 μM R108Y Gsp1
2 nM GAP and 50 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 10.29 s^{-1}
K_m = 0.41 μM
s
22.3 μM R112S Gsp1
3 nM GAP and 50 μM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 5.49 s^{-1}
K_m = 2.67 μM
t
10.6 μM K132H Gsp1
3 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 7.19 s^{-1}
K_m = 5.59 μM
u
6.7 μM H141R Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 9.88 s^{-1}
K_m = 0.18 μM
v
6 μM K143W Gsp1
1 nM GAP and 20 μM phosphate sensor
Raw fluorescence [fluorescence units]
Time (s)
Data
Fit
k_{cat} = 10.02 s^{-1}
K_m = 0.15 μM
w
5.2 μM Q147E Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 7.90 s^{-1}
K_m = 0.60 μM
x
5.8 μM Y157A Gsp1
1 nM GAP and 20 μM phosphate sensor
Time (s)
Data
Fit
k_{cat} = 3.93 s^{-1}
K_m = 0.15 μM

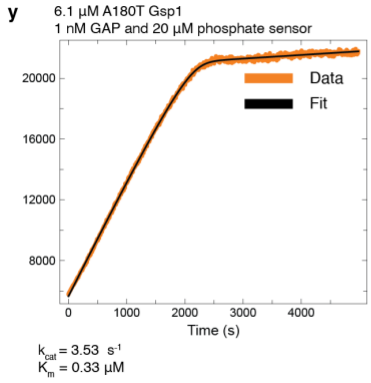
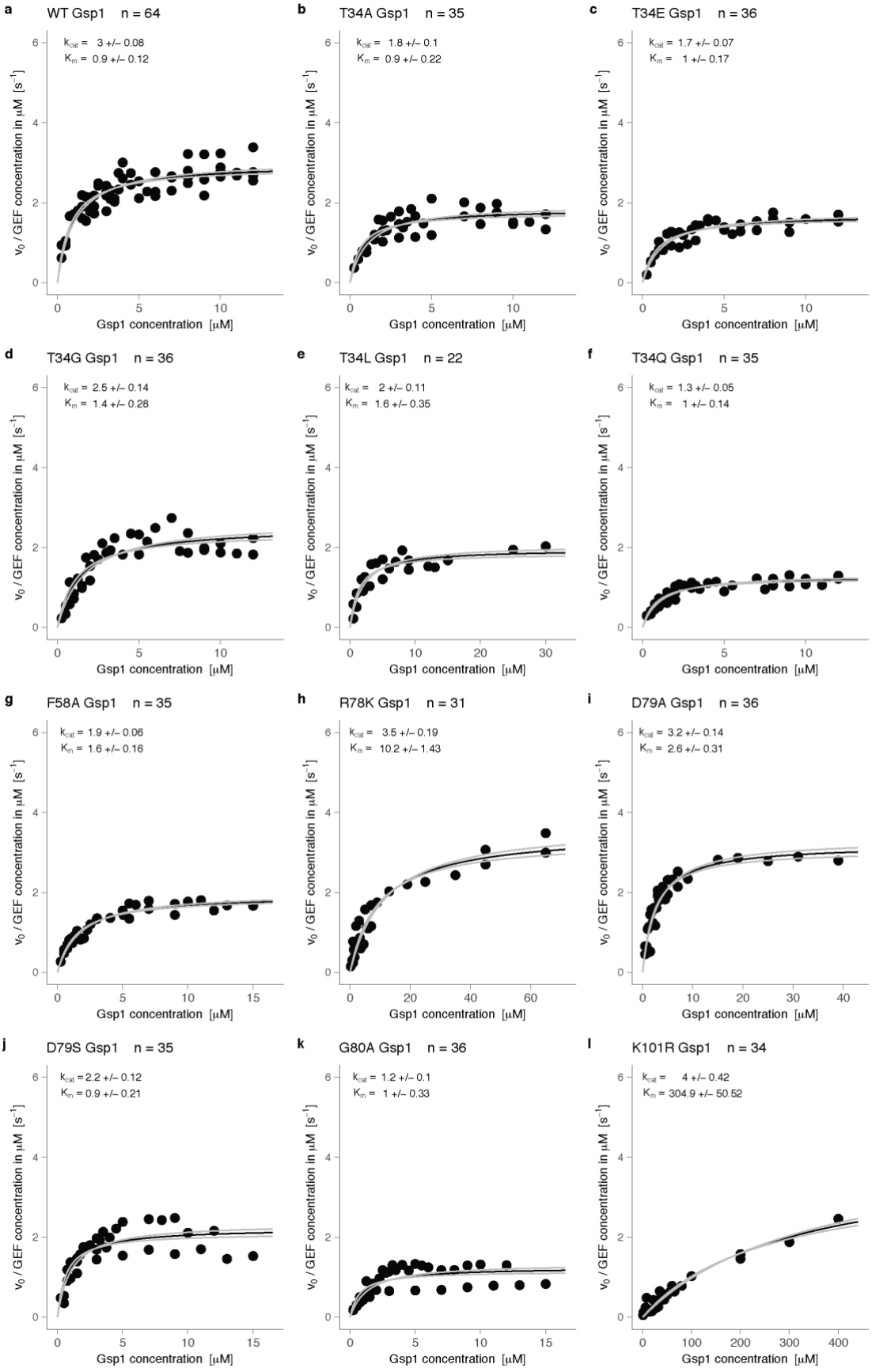
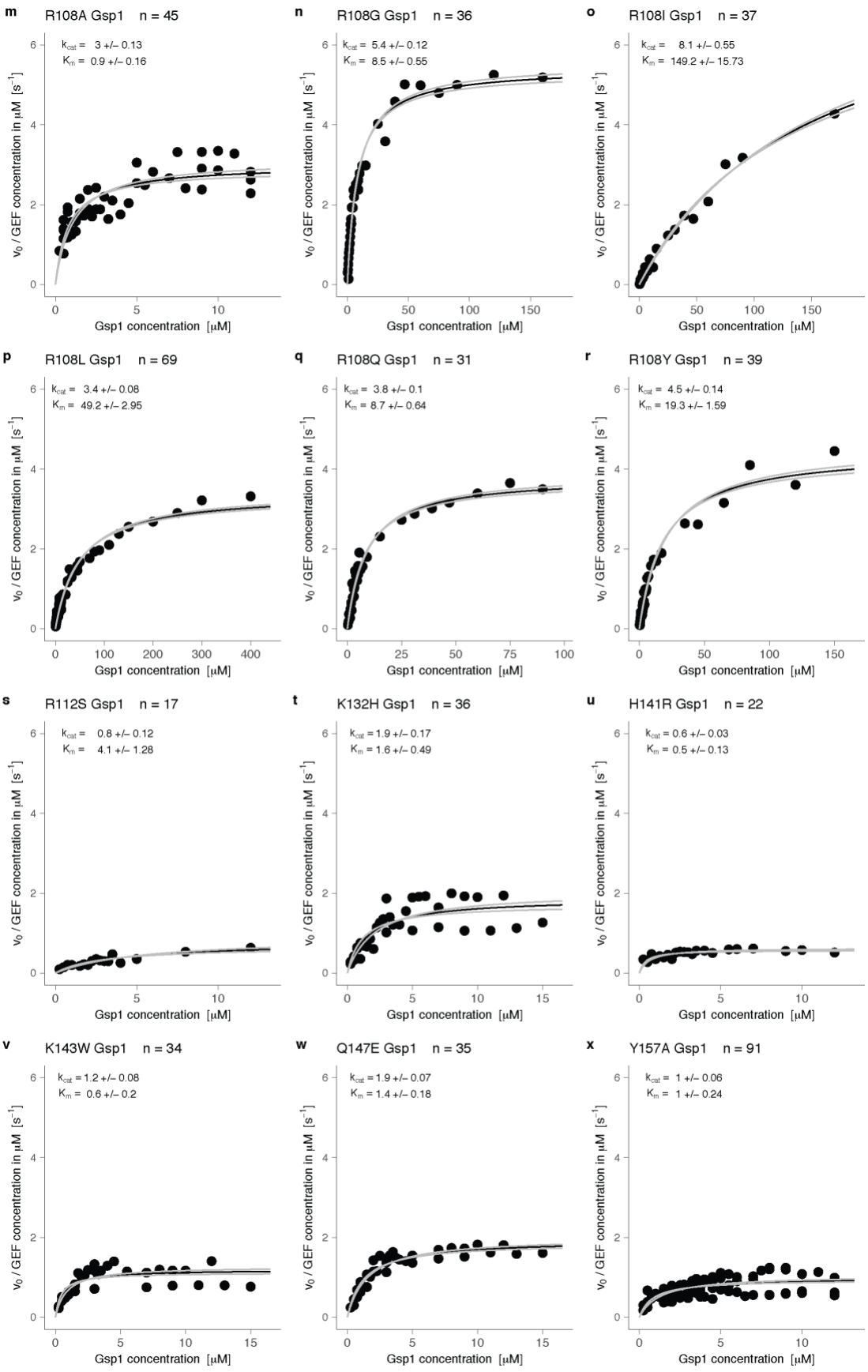


Figure 2.13 GAP-mediated GTP hydrolysis monitored as fluorescence increase upon binding of released free phosphate to a fluorescent phosphate sensor.
Curves were fit with the integrated Michaelis-Menten equation using the DELA software. Final Michaelis-Menten kinetic parameters (k_{cat} and K_m) for each Gsp1 mutant were calculated from three to nine individually fit curves as the ones shown in this figure. **(a)** Wild type Gsp1, **(b-y)**, Gsp1 point mutants.


a
WT Gsp1 n = 64
k_{cat} = 3 +/– 0.08
K_m = 0.9 +/– 0.12
b
T34A Gsp1 n = 35
k_{cat} = 1.8 +/– 0.1
K_m = 0.9 +/– 0.22
c
T34E Gsp1 n = 36
k_{cat} = 1.7 +/– 0.07
K_m = 1 +/– 0.17
d
T34G Gsp1 n = 36
k_{cat} = 2.5 +/– 0.14
K_m = 1.4 +/– 0.28
e
T34L Gsp1 n = 22
k_{cat} = 2 +/– 0.11
K_m = 1.6 +/– 0.35
f
T34Q Gsp1 n = 35
k_{cat} = 1.3 +/– 0.05
K_m = 1 +/– 0.14
g
F58A Gsp1 n = 35
k_{cat} = 1.9 +/– 0.06
K_m = 1.6 +/– 0.16
h
R78K Gsp1 n = 31
k_{cat} = 3.5 +/– 0.19
K_m = 10.2 +/– 1.43
i
D79A Gsp1 n = 36
k_{cat} = 3.2 +/– 0.14
K_m = 2.6 +/– 0.31
j
D79S Gsp1 n = 35
k_{cat} = 2.2 +/– 0.12
K_m = 0.9 +/– 0.21
k
G80A Gsp1 n = 36
k_{cat} = 1.2 +/– 0.1
K_m = 1 +/– 0.33
l
K101R Gsp1 n = 34
k_{cat} = 4 +/– 0.42
K_m = 304.9 +/– 50.52
v_0 / GEF concentration in µM [s^{-1}]
Gsp1 concentration [µM]


m
R108A Gsp1 n = 45
k_{cat} = 3 +/– 0.13
K_m = 0.9 +/– 0.16
n
R108G Gsp1 n = 36
k_{cat} = 5.4 +/– 0.12
K_m = 8.5 +/– 0.55
o
R108I Gsp1 n = 37
k_{cat} = 8.1 +/– 0.55
K_m = 149.2 +/– 15.73
p
R108L Gsp1 n = 69
k_{cat} = 3.4 +/– 0.08
K_m = 49.2 +/– 2.95
q
R108Q Gsp1 n = 31
k_{cat} = 3.8 +/– 0.1
K_m = 8.7 +/– 0.64
r
R108Y Gsp1 n = 39
k_{cat} = 4.5 +/– 0.14
K_m = 19.3 +/– 1.59
s
R112S Gsp1 n = 17
k_{cat} = 0.8 +/– 0.12
K_m = 4.1 +/– 1.28
t
K132H Gsp1 n = 36
k_{cat} = 1.9 +/– 0.17
K_m = 1.6 +/– 0.49
u
H141R Gsp1 n = 22
k_{cat} = 0.6 +/– 0.03
K_m = 0.5 +/– 0.13
v
K143W Gsp1 n = 34
k_{cat} = 1.2 +/– 0.08
K_m = 0.6 +/– 0.2
w
Q147E Gsp1 n = 35
k_{cat} = 1.9 +/– 0.07
K_m = 1.4 +/– 0.18
x
Y157A Gsp1 n = 91
k_{cat} = 1 +/– 0.06
K_m = 1 +/– 0.24
v_0 / GEF concentration in µM [s^{-1}]
Gsp1 concentration [µM]

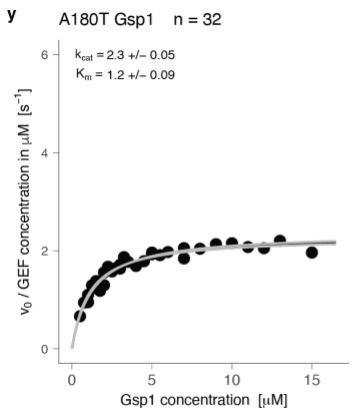


Figure 2.14 Michaelis-Menten plots for GEF-mediated nucleotide exchange.
Black line represents the Michaelis-Menten fit, and the gray lines represent the plus and minus one standard error of the fit. **(a)** Wild type Gsp1. **(b-y)** Gsp1 point mutants.

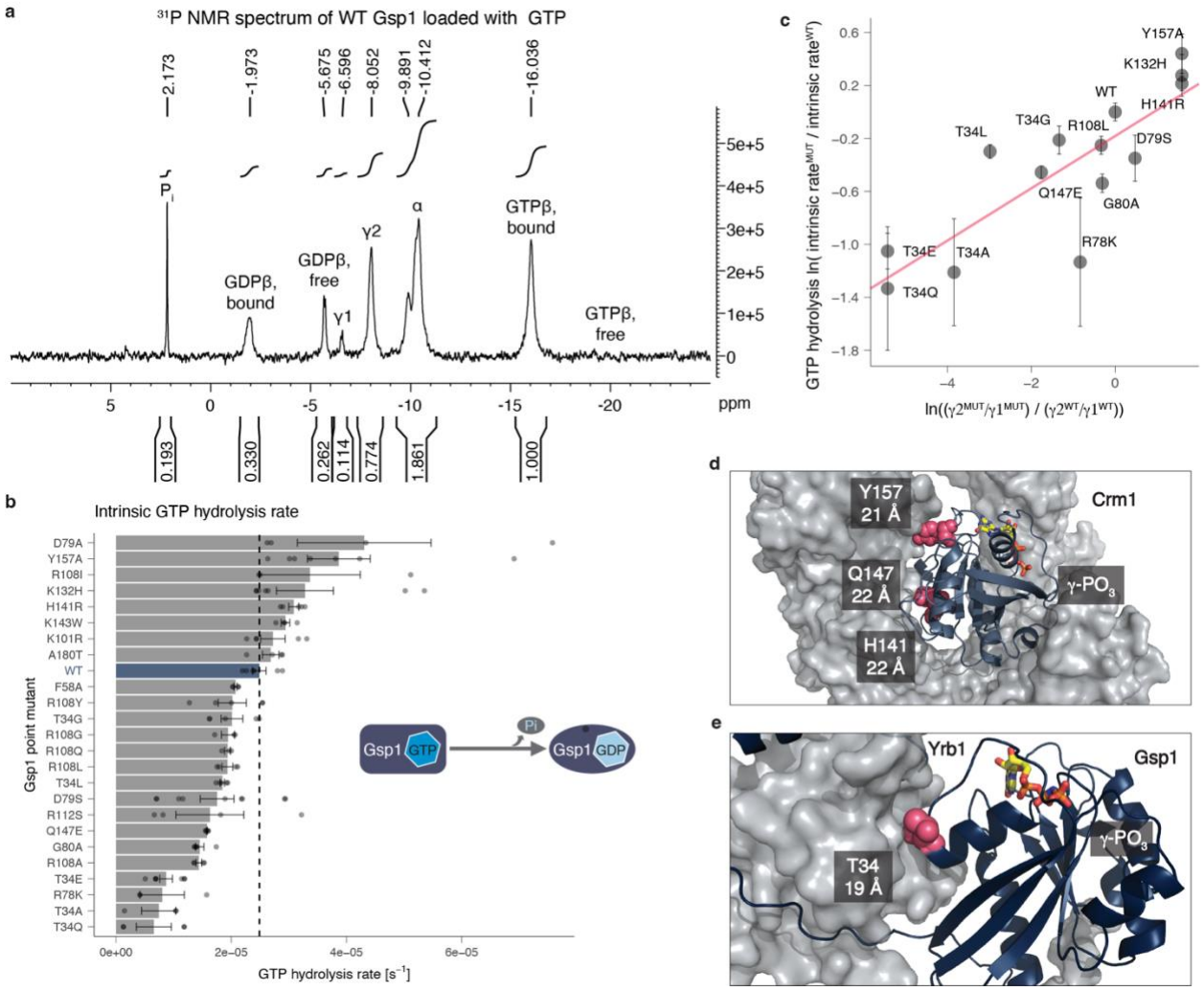


Figure 2.15 Gsp1 interface mutations act allosterically to modulate the rate of GTP hydrolysis.
(a) Annotated 1D ^{31}P NMR spectrum of wild-type Gsp1 loaded with GTP. Peak areas are computed over intervals shown and normalized to the GTPβ bound peak. The peaks from left to right correspond to: free phosphate (Pi), β phosphate of GDP bound to Gsp1 (GDPβbound), β phosphate of free (unbound) GDP (GDPβfree), γ phosphate of GTP bound to Gsp1 in conformation 1 (γ1), γ phosphate of GTP bound to Gsp1 in conformation 2 (γ2), α phosphate of bound or unbound GDP or GTP, β phosphate of GTP bound to Gsp1 (GTPβbound), β phosphate of free (unbound) GTP (GTPβfree). **(b)** Rate of intrinsic GTP hydrolysis of wild-type Gsp1 and mutants. Dotted line indicates wild-type value. Error bars represent the standard deviations from n ≥ 3 replicates (dots). **(c)** Natural log-transformed exchange equilibrium constant between the γ2 and γ1 conformations plotted against the relative rate of intrinsic GTP hydrolysis represented as a natural logarithm of the ratio of the rate for the mutant over the rate of the wild type. The pink line is a linear fit. Error bars represent the standard deviation from n ≥ 3 replicates of intrinsic GTP hydrolysis measurements. **(d)** Location of Y157, H141, and Q147 (pink spheres) in the Crm1 interface (grey surface, PDB 3M1I). Gsp1: navy cartoon; GTP nucleotide: yellow stick representation. **(e)** Location of T34 (pink spheres) in the interface with Yrb1 (grey surface, PDB 1K5D). Distances from the γ phosphate of GTP to the residue α-carbon are indicated below the residue numbers in (d) and (e).

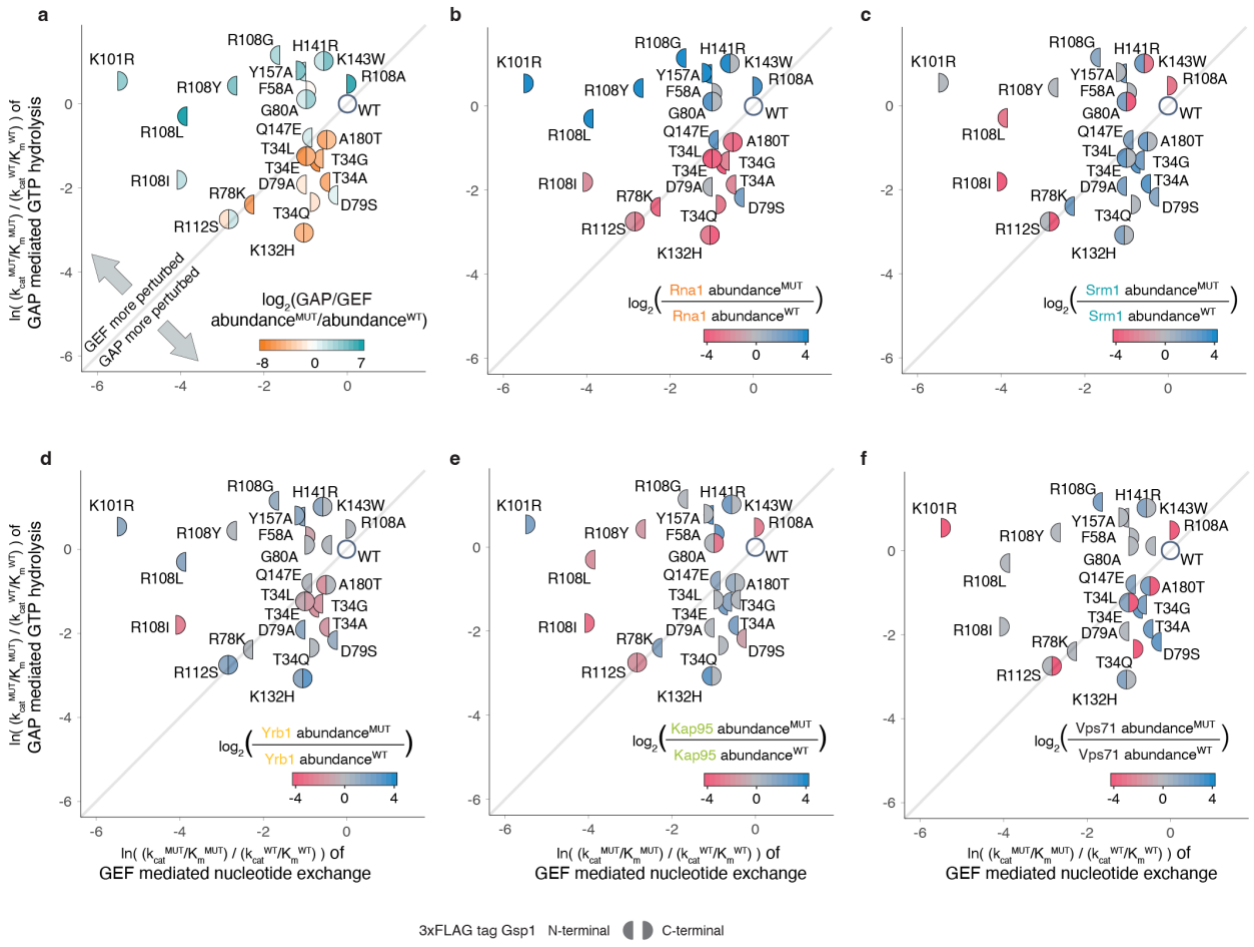


Figure 2.16 Relative prey protein abundance overlaid onto the effects of each mutation on relative in vitro efficiencies of GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange.
Relative GAP-mediated hydrolysis and GEF-mediated exchange efficiencies are plotted as $\ln(k_{cat}^{MUT}/K_m^{MUT} / k_{cat}^{WT}/K_m^{WT})$. Mutants that affect the efficiency (k_{cat}/K_m) of GEF-catalyzed nucleotide exchange more than the efficiency of GAP-catalyzed GTP hydrolysis are above the diagonal, and the mutants that affect the GAP-catalyzed hydrolysis are below the diagonal. Left semi-circle represents an amino-terminal 3xFLAG-tagged Gsp1 point mutant, and right semi-circle represents a carboxy-terminal 3xFLAG-tagged Gsp1 point mutant, relative to wild-type Gsp1 with the corresponding tag. **(a)** Color represents $\log_2$-transformed ratio of GAP and GEF abundance fold change for each Gsp1 point mutant compared to wild type defined as $\log_2((\text{abundance(Rna1)}^{MUT}/\text{abundance(Rna1)}^{WT}) / (\text{abundance(Srm1)}^{MUT}/\text{abundance(Srm1)}^{WT}))$. Orange colored mutants pull-down relatively less Rna1 (GAP) and teal mutants less Srm1 (GEF). **(b-f)** Color represents the log-transformed ratio of mutant and wild type pulled-down prey protein represented as $\log_2(\text{PREY abundance}^{MUT}/\text{PREY abundance}^{WT})$. Log-transformed relative abundance values are capped at +/− 4. Decreased prey abundance from AP-MS in pulled-down complexes with a mutant Gsp1 compared to complexes with the wild-type Gsp1 is represented in red and increased abundance in blue. Prey proteins: (b) Rna1 (GAP); (c) Srm1 (GEF); (d) Yrb1; (e) Kap95, and (f) Vps71. Yrb1 follows a pattern similar to that of Rna1 (GAP), whereas Kap95 and Vps71 are similar to Srm1 (GEF).

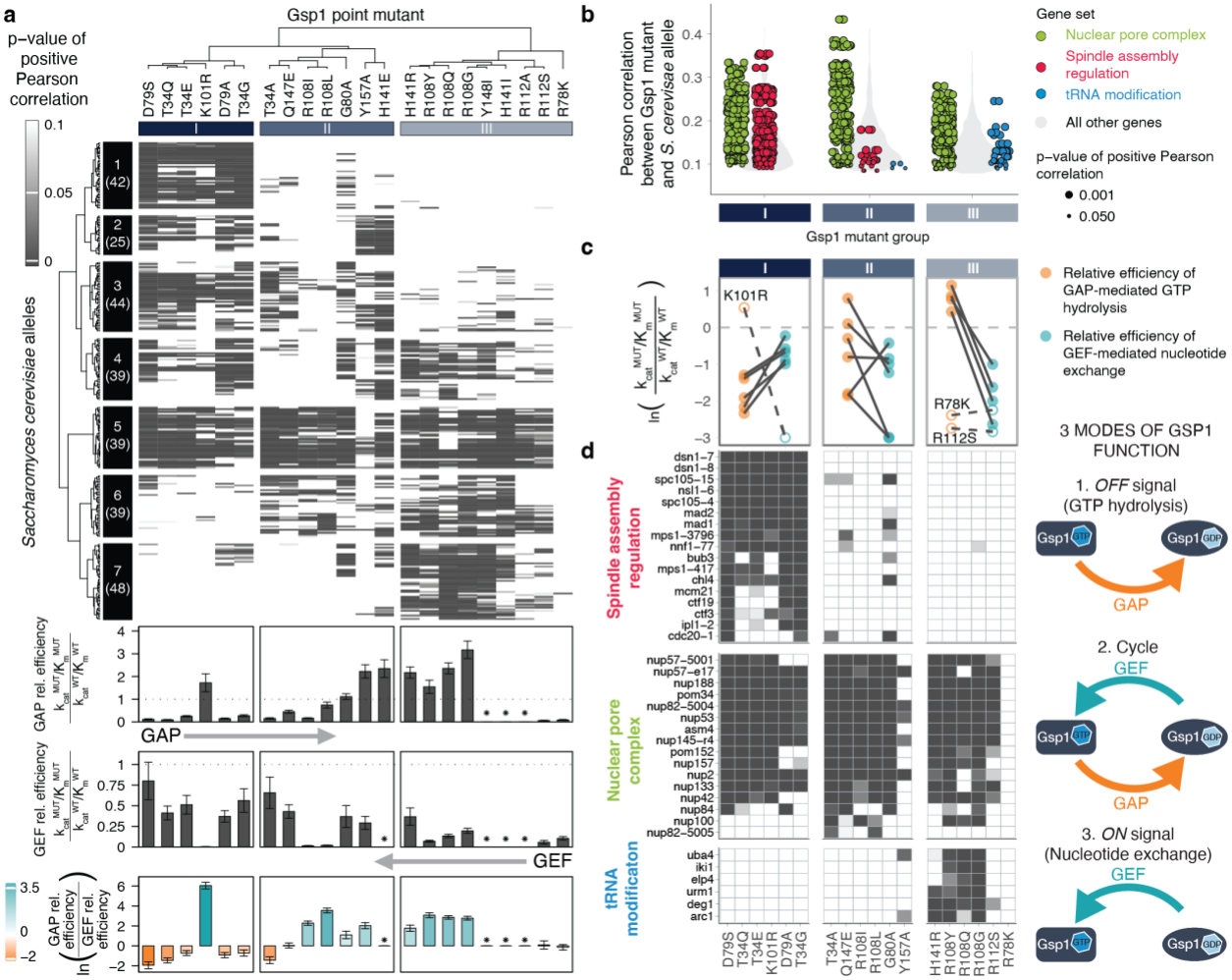


Figure 2.17 Cellular effects of interface mutations group by their effect on GTPase cycle kinetics.
(a) Clustering of 276 S. cerevisiae alleles and 22 strong *GSP1* point mutants by the p-value of Pearson correlations of their GI profiles compared to the relative efficiencies of GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange (asterisks, not measured). Grey scale: FDR-adjusted one-sided (positive) P value of the Pearson correlations. Numbers in parentheses indicate the number of genes in the cluster. **(b)** Distributions of Pearson correlations, separated by *GSP1* point mutant groups from column hierarchical clustering in (a). Green, red, or blue points are individual correlations with *S. cerevisiae* genes in three gene sets; grey violin plots show distributions of correlations with all other genes; point size corresponds to the FDR-adjusted one-sided (positive) p-value of the Pearson correlation. Only significant correlations (p-value < 0.05) are included. **(c)** Kinetic characteristics of Gsp1 mutant groups I to III. Outliers are shown as empty circles and dashed lines. The log ratio of relative catalytic efficiencies is capped at −3. **(d)** Heat maps of the FDR-adjusted one-sided (positive) p-value of the Pearson correlation for the three representative gene sets. *S. cerevisiae* genes for each gene set are clustered by p-value. The GTPase cycle schemes on the right represent three modes of Gsp1 function. In (c, d), only Gsp1 mutants with kinetics data are shown, grouped as in (a).

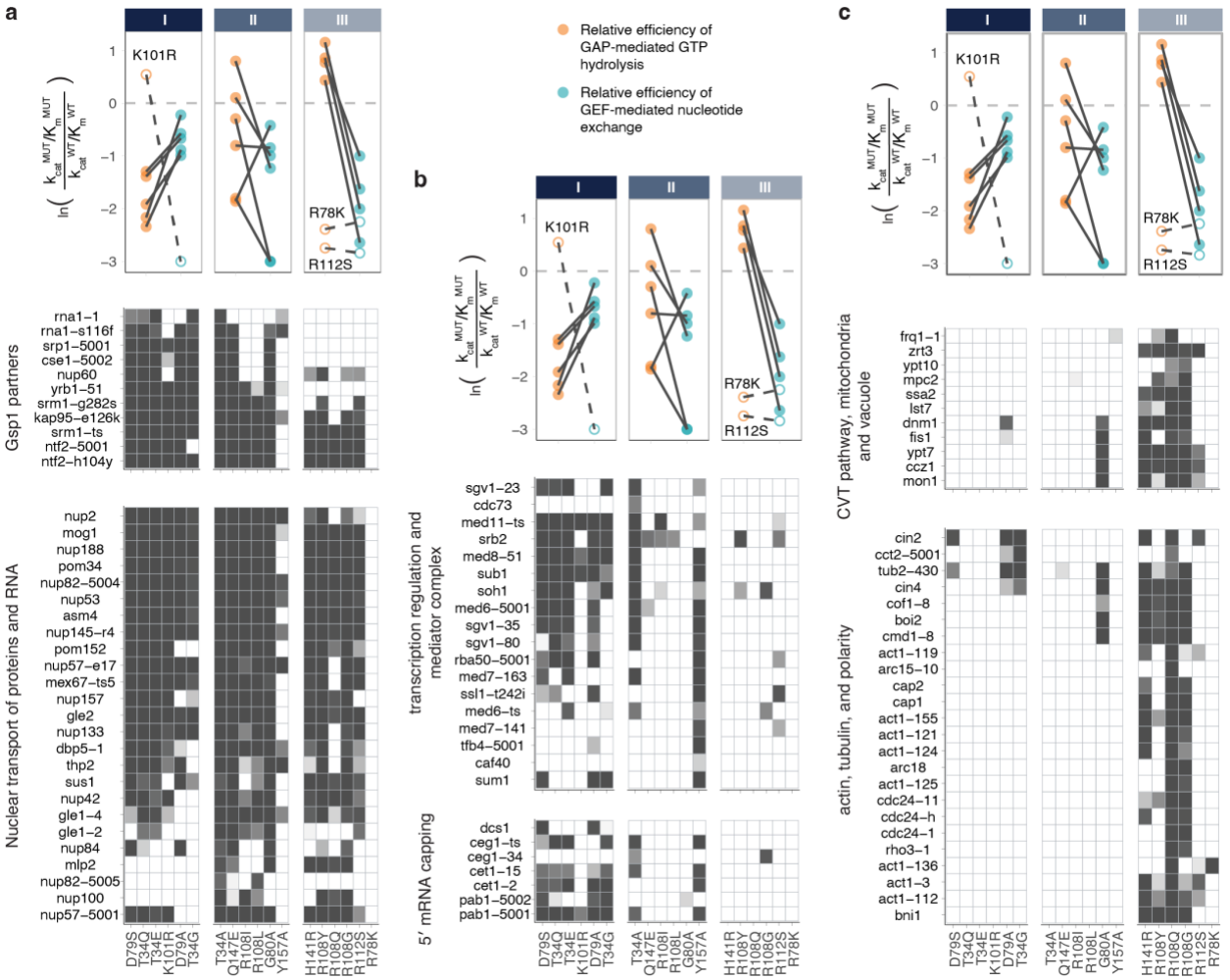


Figure 2.18 Sets of *S. cerevisiae* genes grouped by biological functions.
Heat maps of the false discovery rate adjusted one-sided (positive) p-values of the Pearson correlations between the GI profiles of 22 strong *GSP1* point mutants and GI profiles of knock-outs or knock-downs of *S. cerevisiae* genes from a previously published dataset.[65] The p-value is represented as a white to grey range as in **Figure 2.17**. Genes are organized in gene sets based on their biological function (Methods). The line plots above the heat maps are the same as in **Figure 2.17**. **(a)** *GSP1* point mutants and alleles of Gsp1 binding partners with available co-complex X-ray crystal structures, and *S. cerevisiae* genes involved in nuclear transport of RNA and proteins. **(b)** *GSP1* point mutants and *S. cerevisiae* genes involved in transcription regulation or 5′ mRNA capping. **(c)** *GSP1* point mutants and *S. cerevisiae* genes involved in the cytoplasm-to-vacuole targeting (CVT) pathway, and actin, tubulin, and cell polarity.

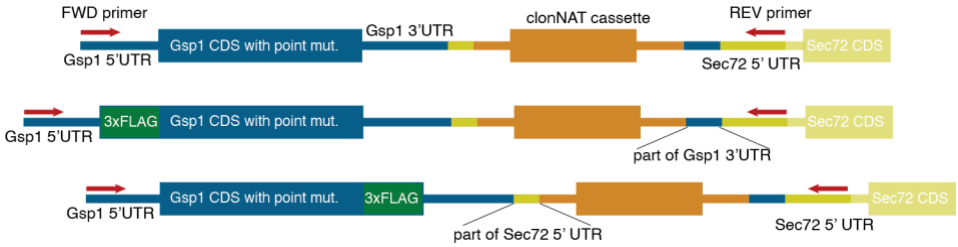


Figure 2.19 Schematic of genomically integrated GSP1 constructs.
For E-MAP experiments, wild type or mutant GSP1 cassettes including the clonNAT resistance cassette were integrated into the MAT:α strain. For AP-MS the constructs also included either an amino- (N) terminal or a carboxy- (C) terminal 3xFLAG tag (M**DYKDHDGDYKDHDIDYKDDDDK**GGGGA and GGGGA**DYKDHDGDYKDHDIDYKDDDDK**, respectively).

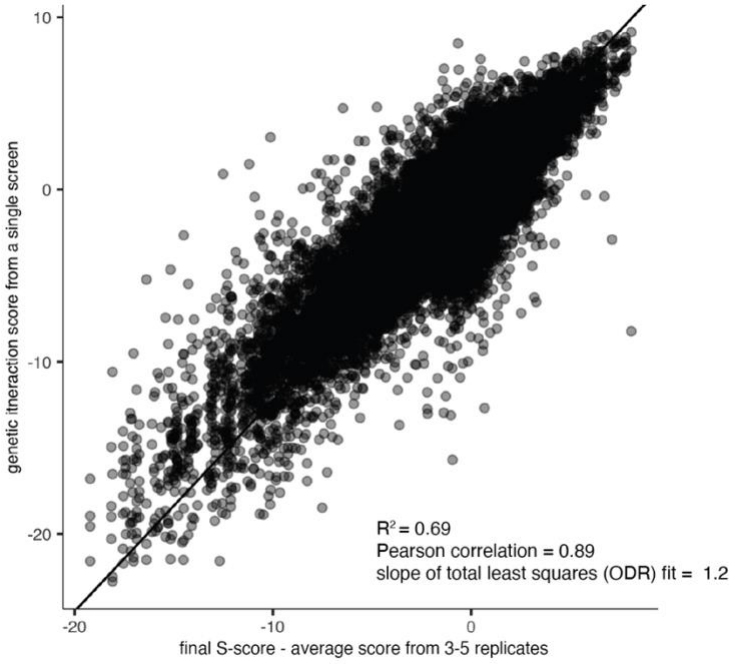


Figure 2.20 Reproducibility of GSP1 point mutant E-MAP screens.
A linear relationship between the genetic interaction S-score from a single E-MAP experiment and the final average S-score based on three or more replicates. The linear fit was calculated using the *odregress* function from the *pracma* R package.

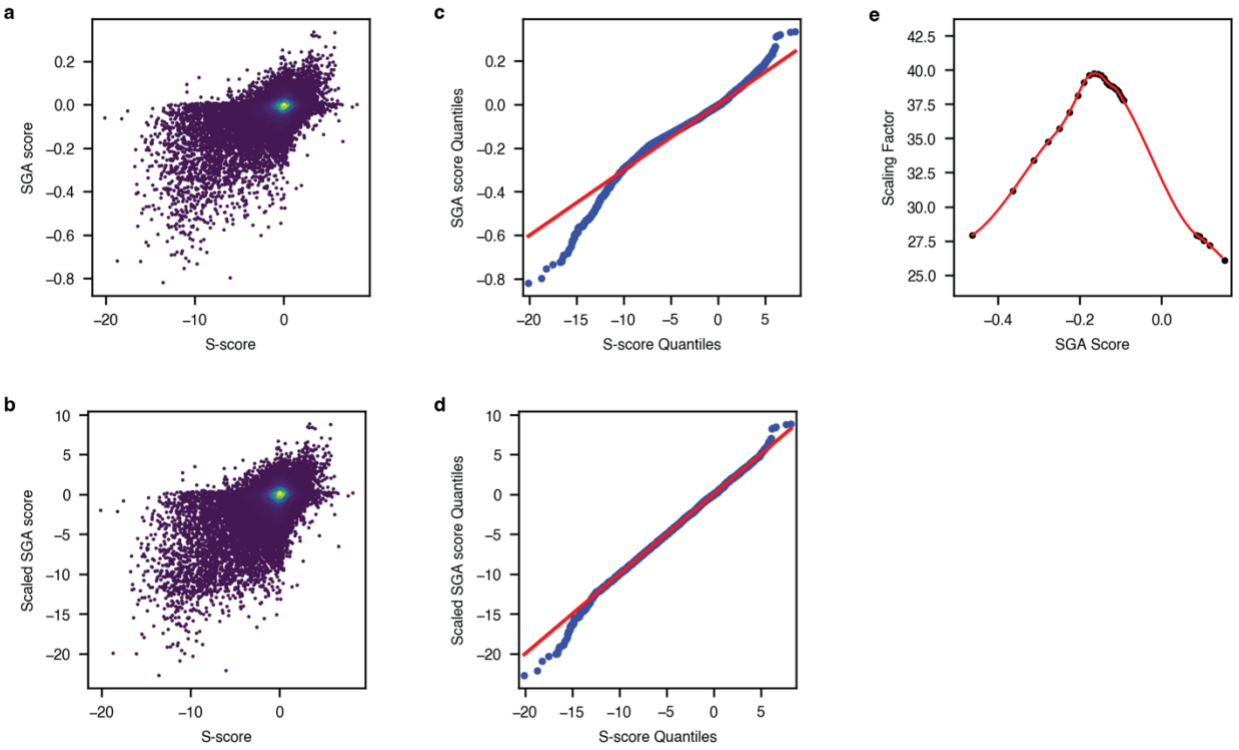


Figure 2.21 Non-linear scaling of SGA data from the Cell Map to E-MAP format.
(a) Distribution of S-scores from the chromatin biology E-MAP dataset[66] and the SGA score from the CellMap dataset.[65] **(b)** Distribution of S-scores from the chromatin biology E-MAP dataset and the *scaled* SGA score from the CellMap dataset. **(c)** Quantile-quantile plot showing the distribution of genetic interaction scores from the CellMap before scaling and the E-MAP chromatin biology datasets. **(d)** Quantile-quantile plot showing the distribution of genetic interaction scores from the CellMap after scaling and the E-MAP chromatin biology dataset. **(e)** The scaling function applied to the CellMap data. Red curve is the fitted spline of the scaling factors between the E-MAP S-scores and the SGA scores. Black dots represent the individual bins.

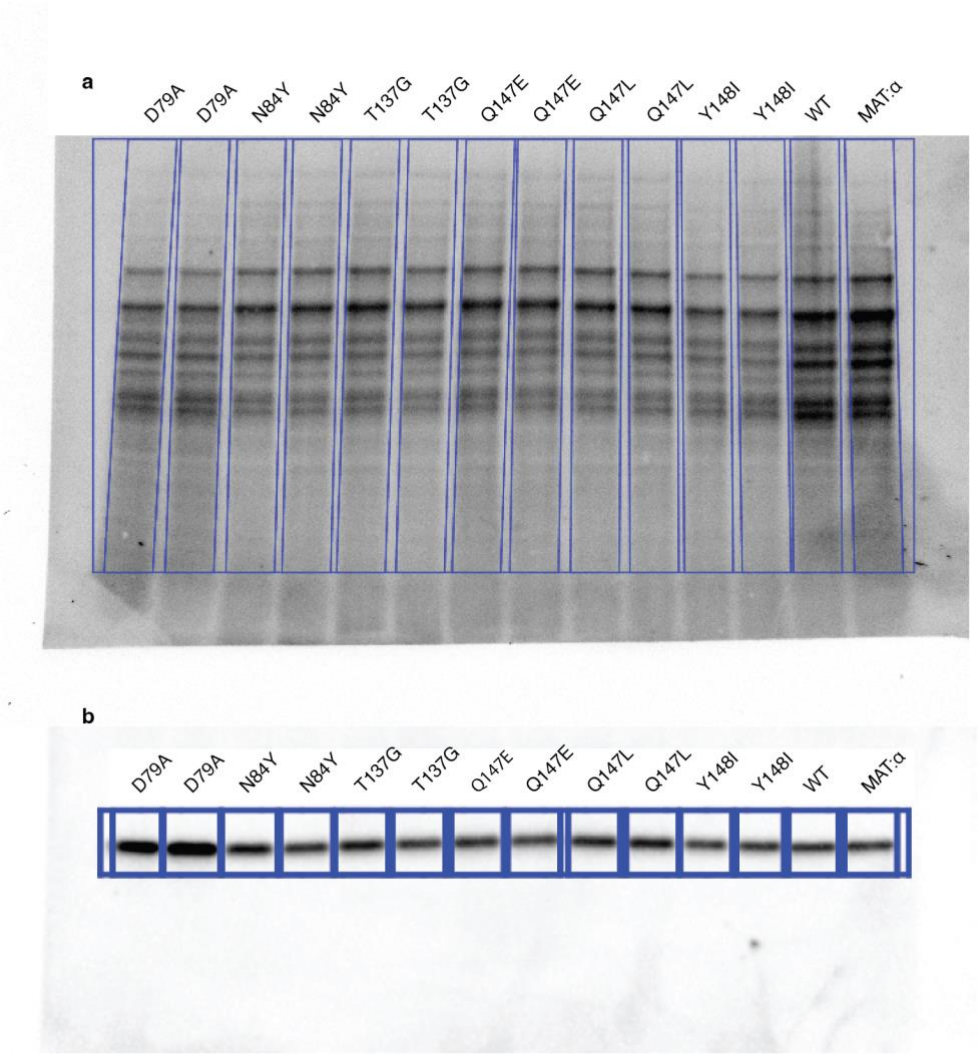


Figure 2.22 Example data for Gsp1 protein expression estimation by Western blot.
(a) Total protein staining. (**b)** Western blot of starting S. cerevisiae strain (MAT:α, see Methods for full strain description), wild type Gsp1 with clonNAT resistance marker (WT), and its mutants with anti-Ran antibody.

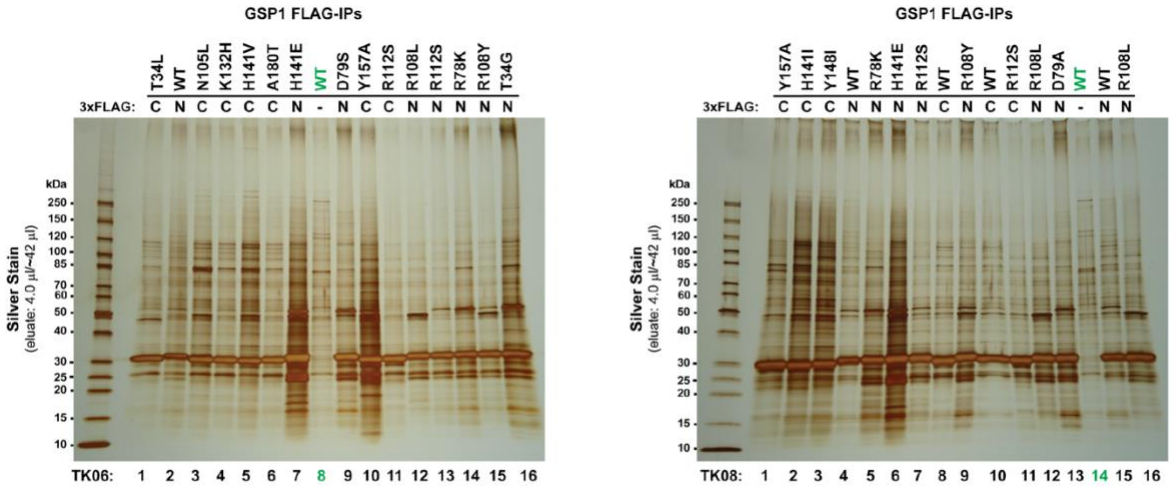


Figure 2.23 Silver stain gels after FLAG immunoprecipitation of amino- (N) or carboxy- (C) terminally 3xFLAG tagged genomically integrated Gsp1.
The strongest band at approximately 30 kDa corresponds to tagged Gsp1. Untagged wild type Gsp1 (lanes 8 and 14 in the left and right gel, respectively) were used as negative control for mass spectrometry analysis.

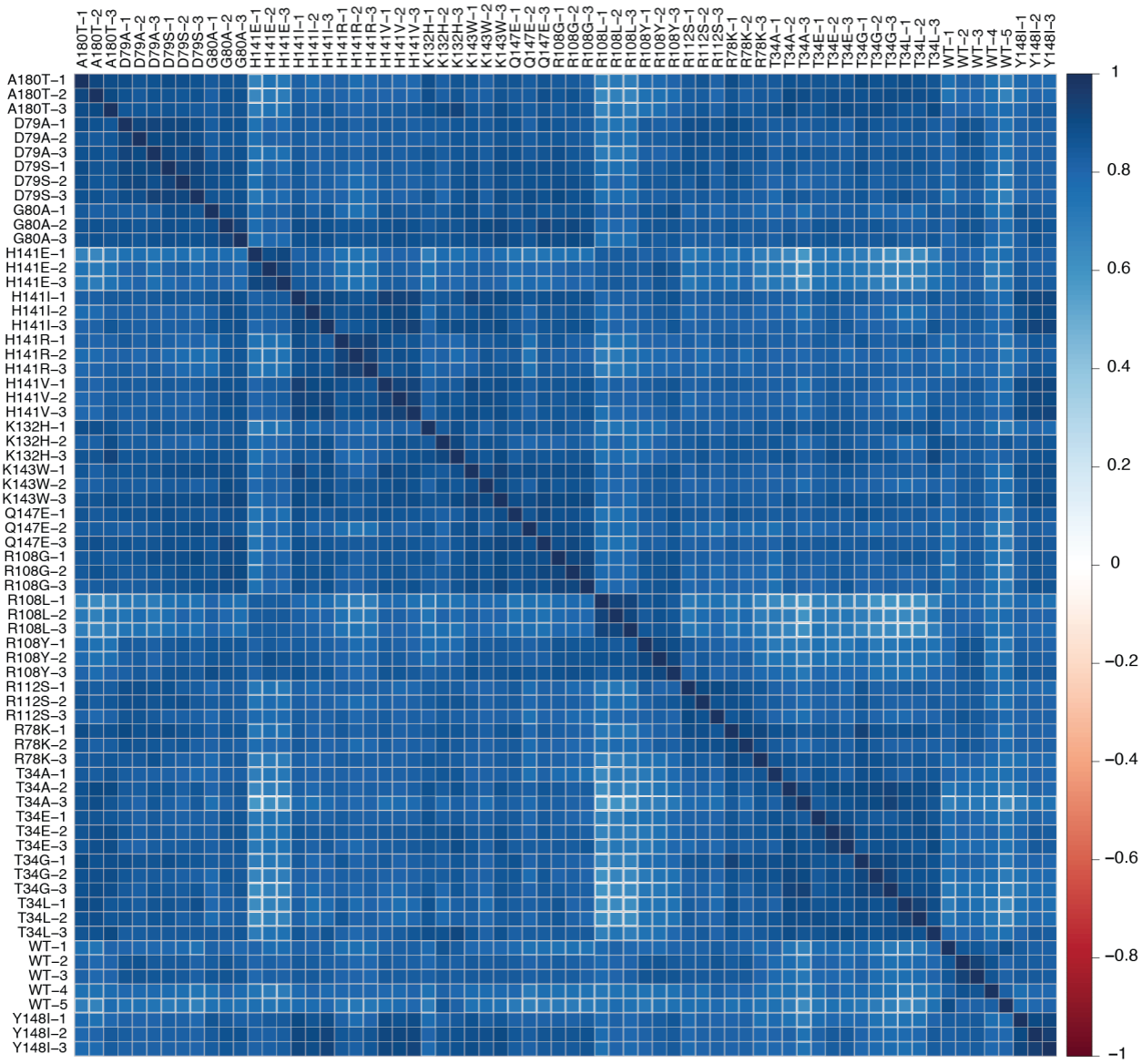


Figure 2.24 Clustering of individual AP-MS replicates based on correlations between protein abundance before the final scoring.
Data shown are for amino-terminally FLAG tagged wild type (WT) and Gsp1 mutants.

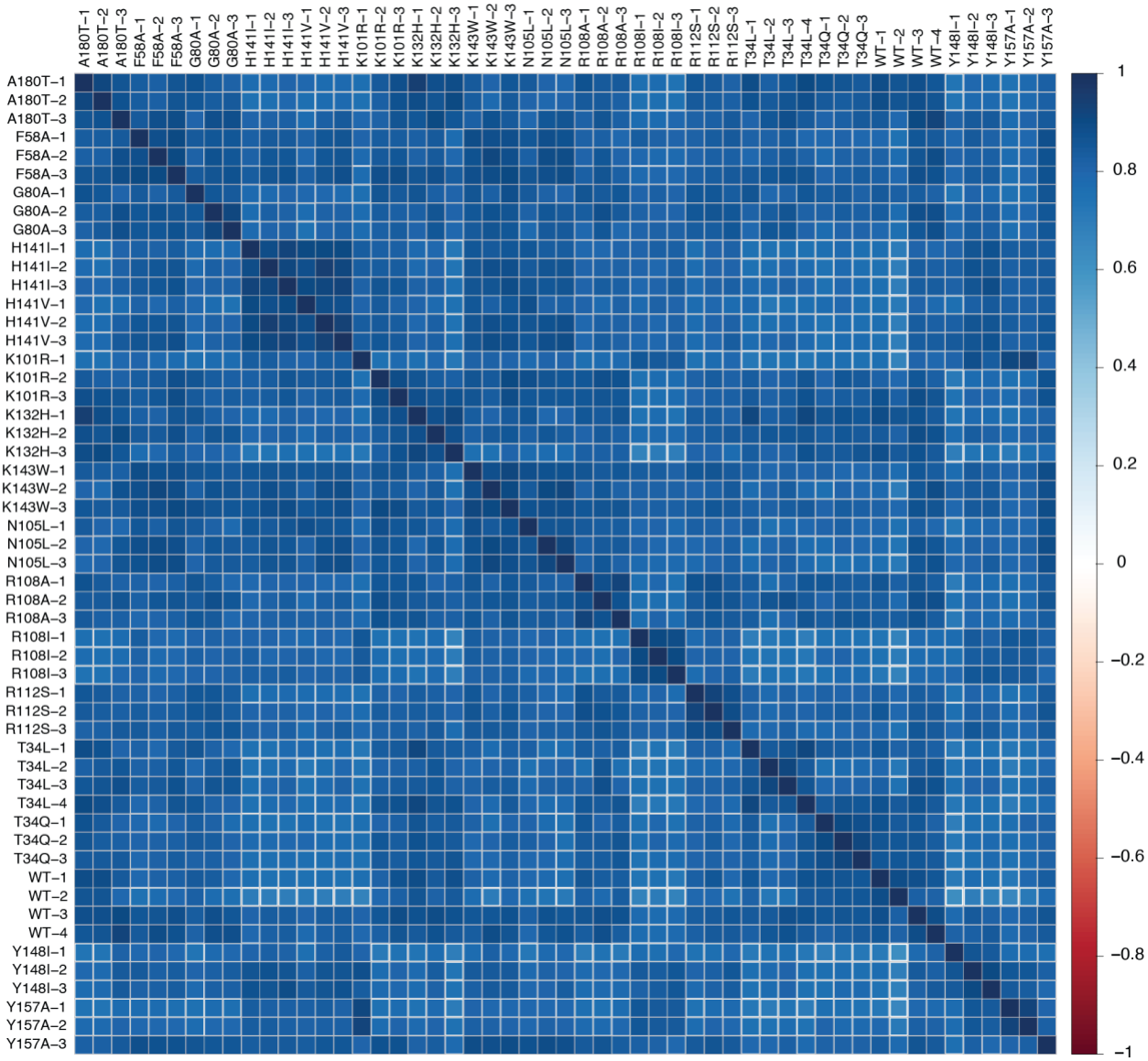


Figure 2.25 Clustering of individual AP-MS replicates based on correlations between protein abundance before the final scoring.
Data shown are for carboxy-terminally FLAG tagged wild type (WT) and Gsp1 mutants.

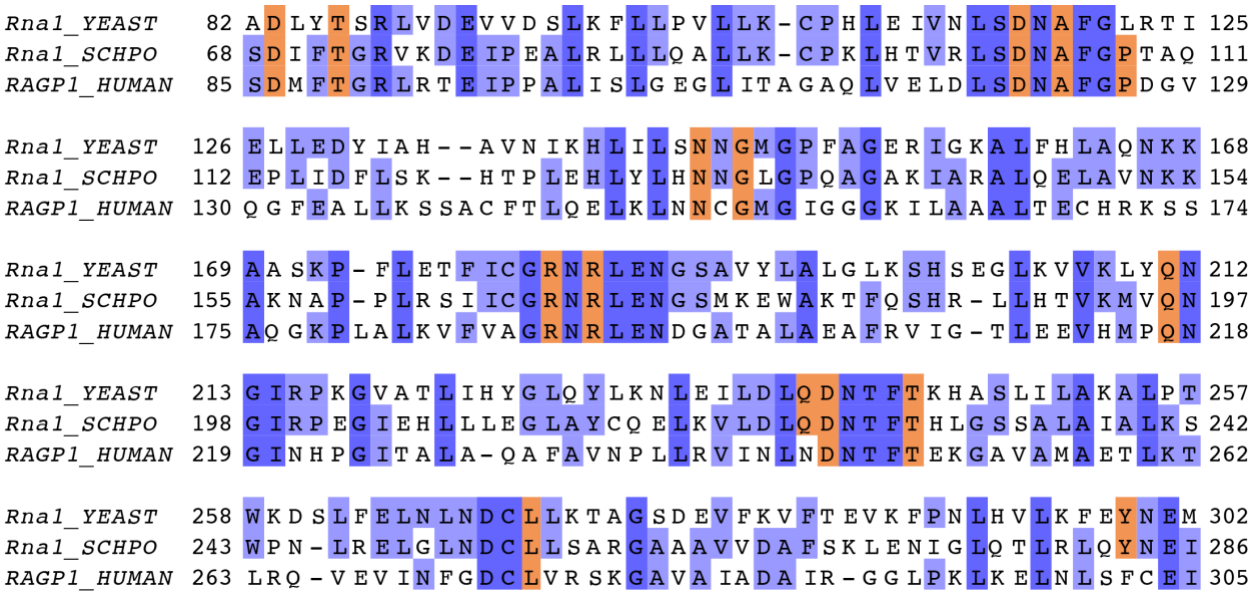


Figure 2.26 Multiple sequence alignment between Rna1 from *S. cerevisiae* (Rna1_YEAST) and S. pombe (Rna1_SCHPO), as well as human RanGAP (RAGP1_HUMAN, excluding the C-terminal SUMO conjugation domain which is absent in Fungi).
Overall sequence identity between *S. cerevisiae* and *S. pombe* Rna1 is 39%, with 53% sequence similarity. Interface core residues (based on the X-ray crystal structure between *S. pombe* Rna1 and mammalian Ran, PDB ID: 1k5d) are highlighted in orange. All interface core residues except Pro108 in *S. pombe* Rna1, which corresponds to Leu122 in *S. cerevisiae* Rna1, are conserved in sequence between *S. cerevisiae* and *S. pombe* Rna1.

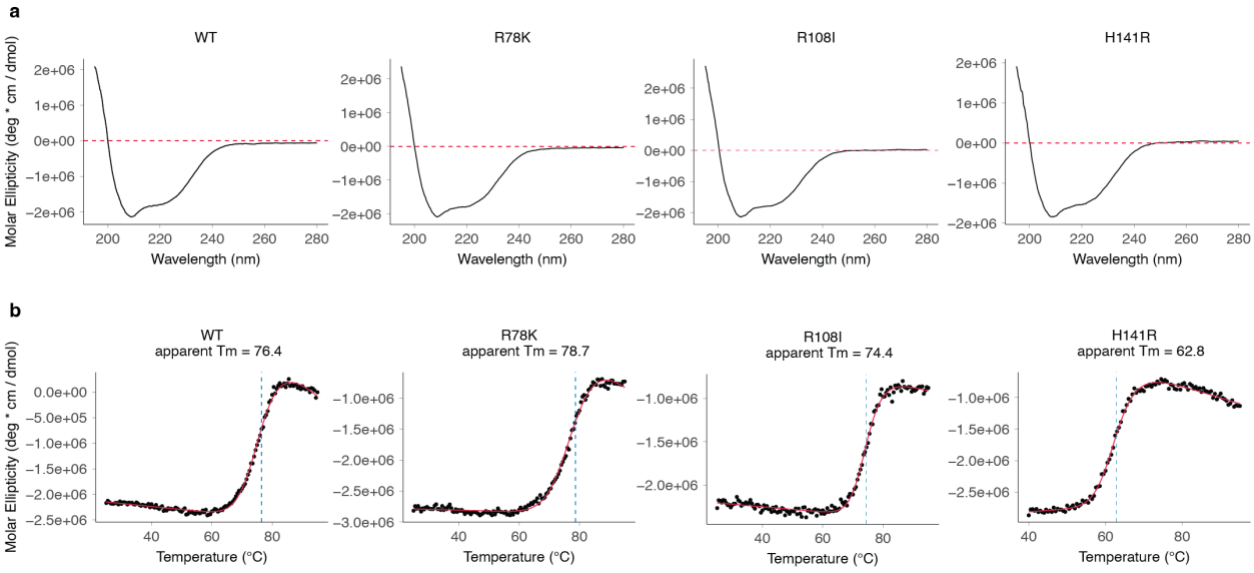


Figure 2.27 Circular dichroism (CD) data for wild type (WT) Gsp1 and select mutants.
(a) CD spectra. **(b)** Irreversible temperature melts.

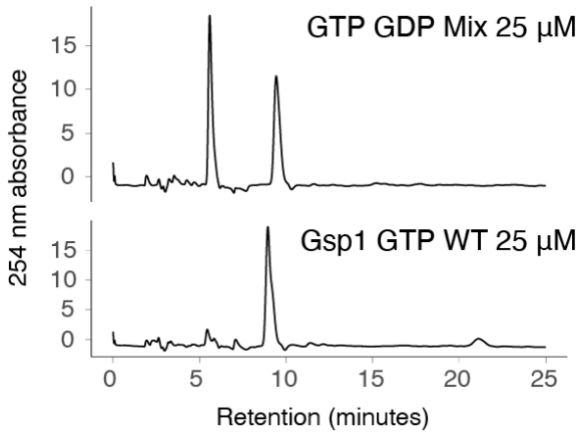


Figure 2.28 HPLC reverse phase chromatograms of a GTP/GDP mix (top) and that of a purified and GTP loaded wild type Gsp1 (bottom).

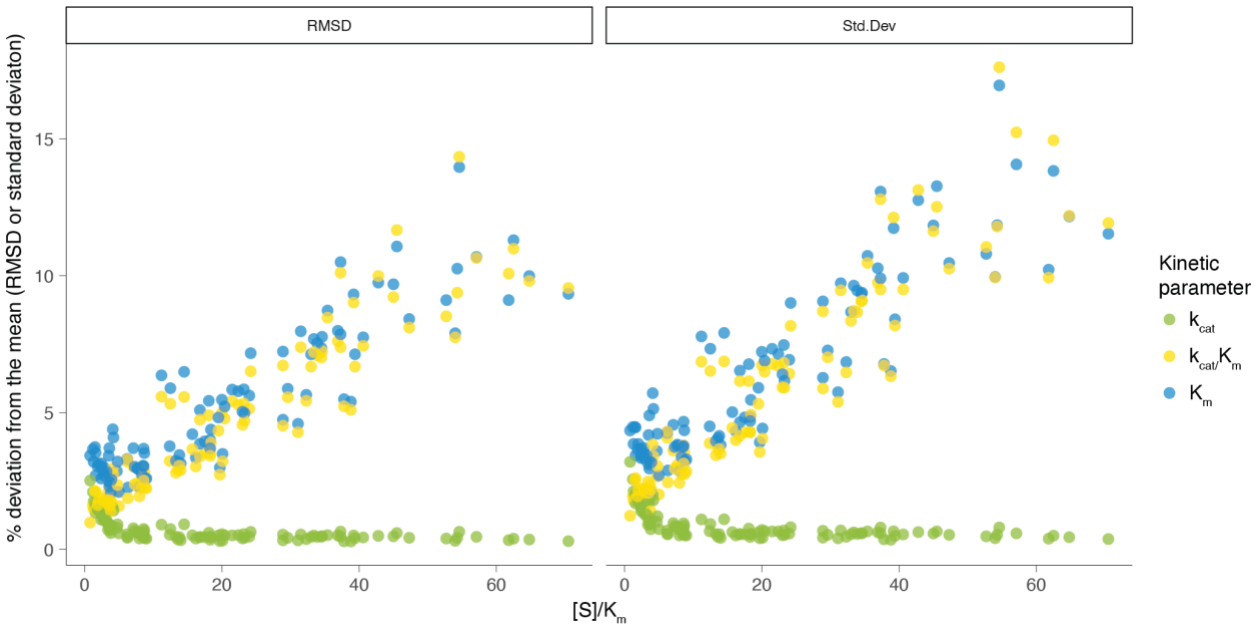


Figure 2.29 Accuracy estimation for determining the kinetic parameters of GAP-mediated GTP hydrolysis from individual time courses spanning [S] > K_m to [S] << K_m fit with an accurate solution of the integrated Michaelis Menten (IMM) equation.

Each time course was simulated using the experimentally determined parameters determined from the fitted IMM model, with added Gaussian noise similar to the experimental fluorescence signal noise. The deviation from the mean is plotted against a ratio of initial substrate (Gsp1:GTP) concentration [S] and the experimentally determined K_m. Deviation from the mean is reported either as standard deviation or $RMSD = \sqrt{\sum(simulation\_param - experimental\_param)^2 / N}$, where N = 100 simulations, and simulation_param and experimental_param are simulated and experimental k_{cat}, K_m, and k_{cat}/K_m, respectively. Here, simulated refers to the average of the fitted values for the simulated data sets.

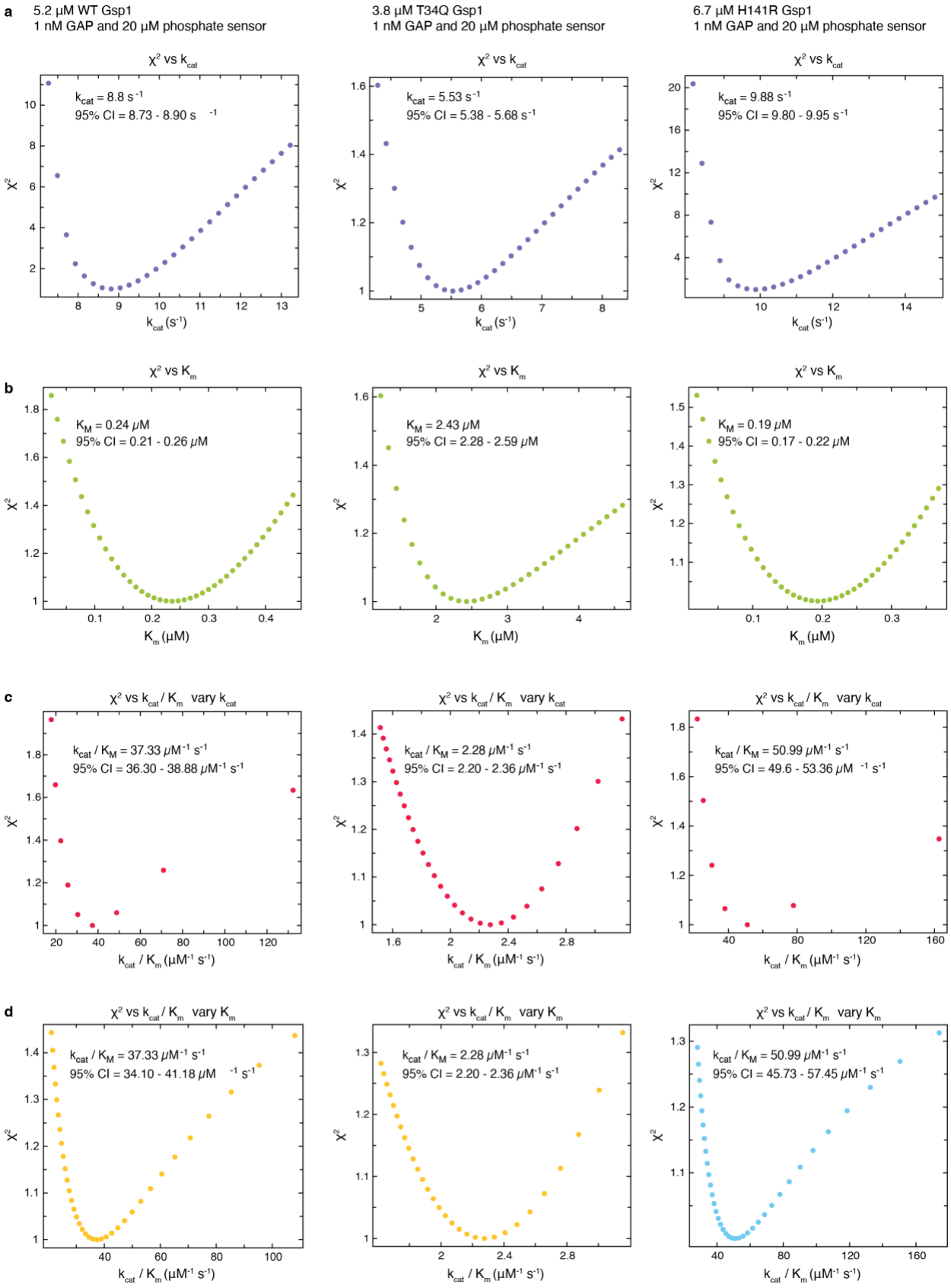

a
5.2 µM WT Gsp1
1 nM GAP and 20 µM phosphate sensor
3.8 µM T34Q Gsp1
1 nM GAP and 20 µM phosphate sensor
6.7 µM H141R Gsp1
1 nM GAP and 20 µM phosphate sensor
χ² vs k_cat
k_cat = 8.8 s⁻¹
95% CI = 8.73 - 8.90 s⁻¹
k_cat = 5.53 s⁻¹
95% CI = 5.38 - 5.68 s⁻¹
k_cat = 9.88 s⁻¹
95% CI = 9.80 - 9.95 s⁻¹
χ²
k_cat (s⁻¹)
b
χ² vs K_m
K_M = 0.24 µM
95% CI = 0.21 - 0.26 µM
K_M = 2.43 µM
95% CI = 2.28 - 2.59 µM
K_M = 0.19 µM
95% CI = 0.17 - 0.22 µM
K_m (µM)
c
χ² vs k_cat / K_m vary k_cat
k_cat / K_M = 37.33 µM⁻¹ s⁻¹
95% CI = 36.30 - 38.88 µM⁻¹ s⁻¹
k_cat / K_M = 2.28 µM⁻¹ s⁻¹
95% CI = 2.20 - 2.36 µM⁻¹ s⁻¹
k_cat / K_M = 50.99 µM⁻¹ s⁻¹
95% CI = 49.6 - 53.36 µM⁻¹ s⁻¹
k_cat / K_m (µM⁻¹ s⁻¹)
d
χ² vs k_cat / K_m vary K_m
k_cat / K_M = 37.33 µM⁻¹ s⁻¹
95% CI = 34.10 - 41.18 µM⁻¹ s⁻¹
k_cat / K_M = 2.28 µM⁻¹ s⁻¹
95% CI = 2.20 - 2.36 µM⁻¹ s⁻¹
k_cat / K_M = 50.99 µM⁻¹ s⁻¹
95% CI = 45.73 - 57.45 µM⁻¹ s⁻¹

Figure 2.30 Estimated error around the maximum likelihood estimated values of the Michaelis-Menten parameters.
Plotted is the change in χ^2 statistics as each of the parameters was fixed in gradual increments around the maximum likelihood value. The χ^2 values are relative to the maximum likelihood values. Error estimate analysis is shown for three of the Gsp1 variants: wild type Gsp1, the low efficiency Gsp1 T34Q mutant, and the high efficiency Gsp1 H141R mutant. 95% CI is the estimated 95% confidence interval for each value, based on the χ^2 surface. **(a)** Change of χ^2 statistics as the k_{cat} value is varied around the maximum likelihood value. **(b)** Change of χ^2 statistics as the K_m value is varied around the maximum likelihood value. **(c)** Change of χ^2 statistics as the k_{cat}/K_m value is varied around the maximum likelihood value and the K_m is kept fixed at the maximum likelihood value (k_{cat} is varied). **(d)** Change of χ^2 statistics as the k_{cat}/K_m value is varied around the maximum likelihood value and the k_{cat} is kept fixed at the maximum likelihood value (K_m is varied).

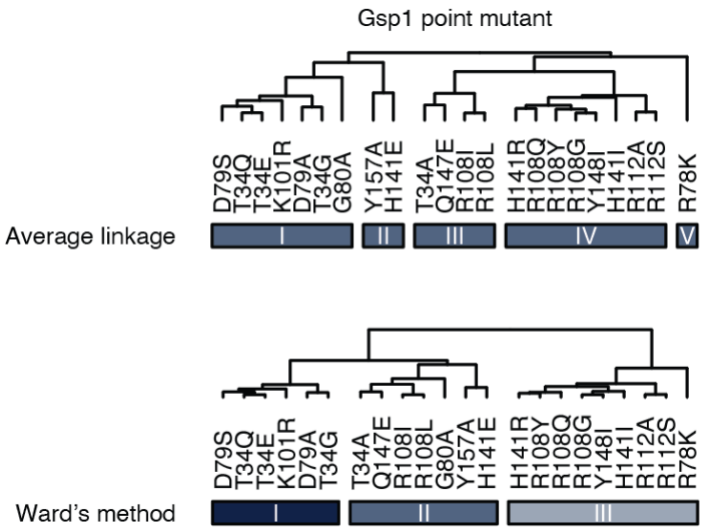


Figure 2.31 Hierarchical clustering of 22 strong Gsp1 point mutants by the p-value of Pearson correlations of their GI profiles and those of 276 *S. cerevisiae* alleles, using either the average linkage (top) or Ward's method (bottom) as linkage criterion.

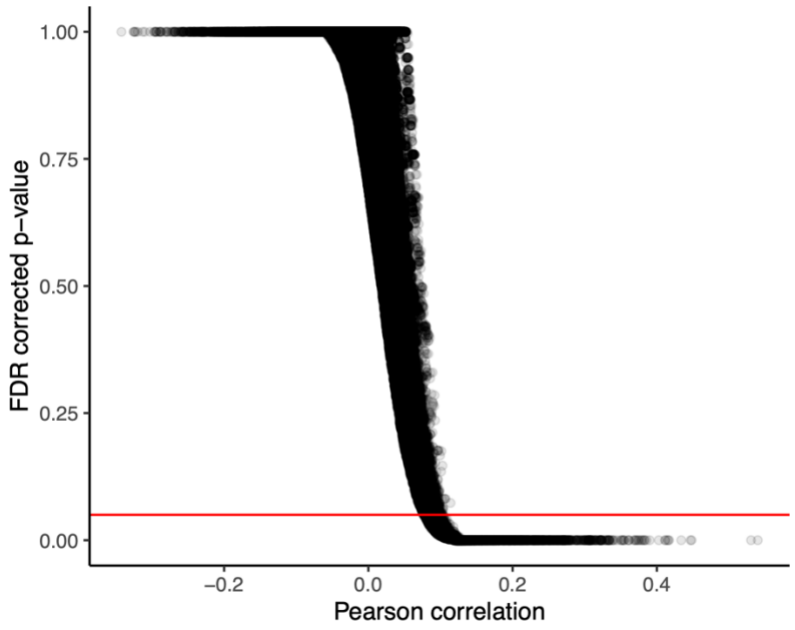


Figure 2.32 FDR corrected p-values for Pearson correlations of genetic interaction profiles between Gsp1 mutants and *S. cerevisiae* alleles.

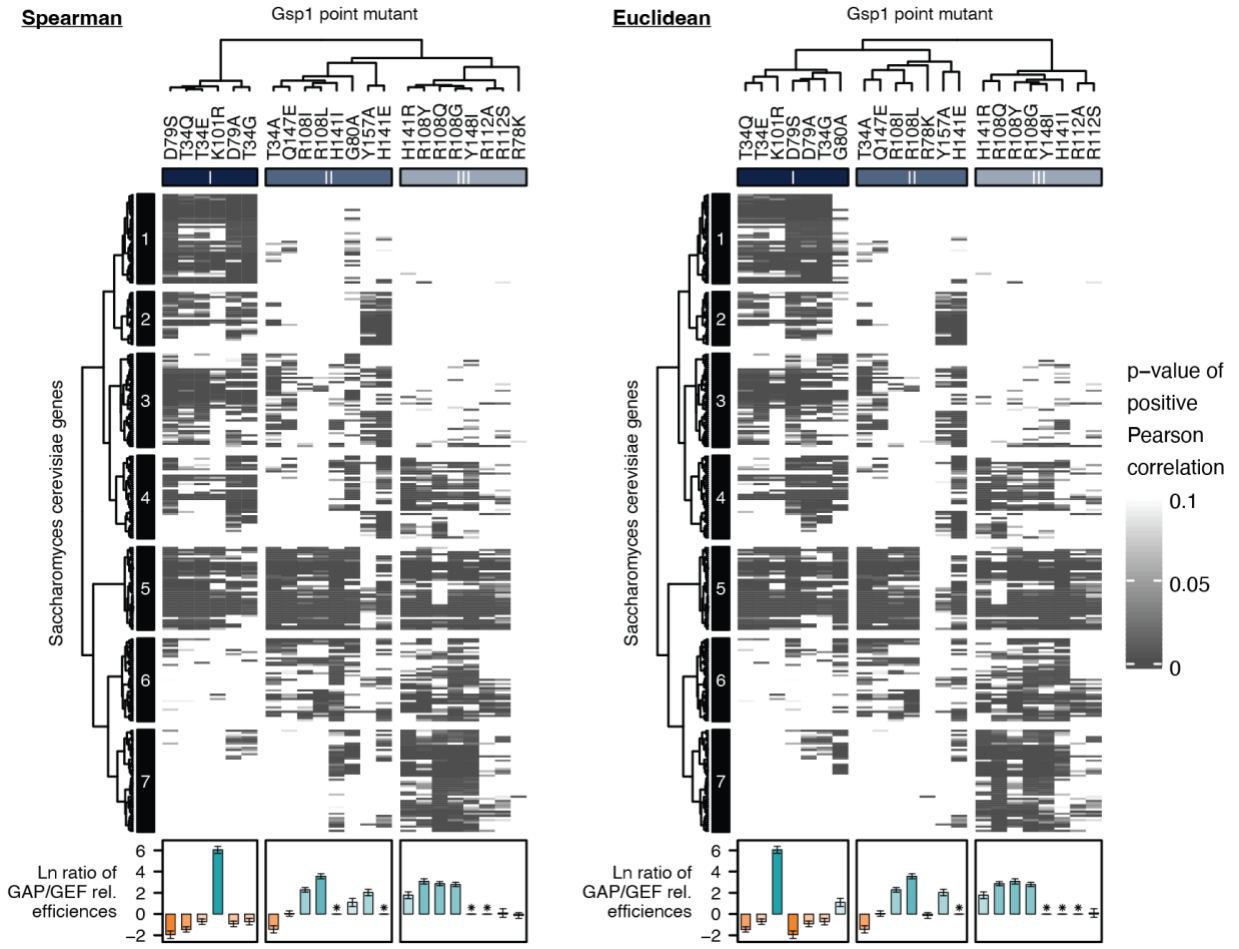


Figure 2.33 Clustering of *S. cerevisiae* alleles and strong Gsp1 point mutants by the p-value of Pearson correlations using alternative distance metrics.
Hierarchical clustering of 276 *S. cerevisiae* alleles and 22 strong Gsp1 point mutants by the p-value of Pearson correlations of their GI profiles compared to the relative efficiencies of GAP-mediated GTP hydrolysis and GEF-mediated nucleotide exchange as indicated (asterisks indicate not measured). The p-value is a false discovery rate adjusted one-sided (positive) p-value of the Pearson correlations (represented as gray scale). The underlying data are identical to those presented in **Figure 2.17A**, but the column clustering was performed using the Spearman correlation or the Euclidean distance rather than the Pearson correlation as a distance metric.

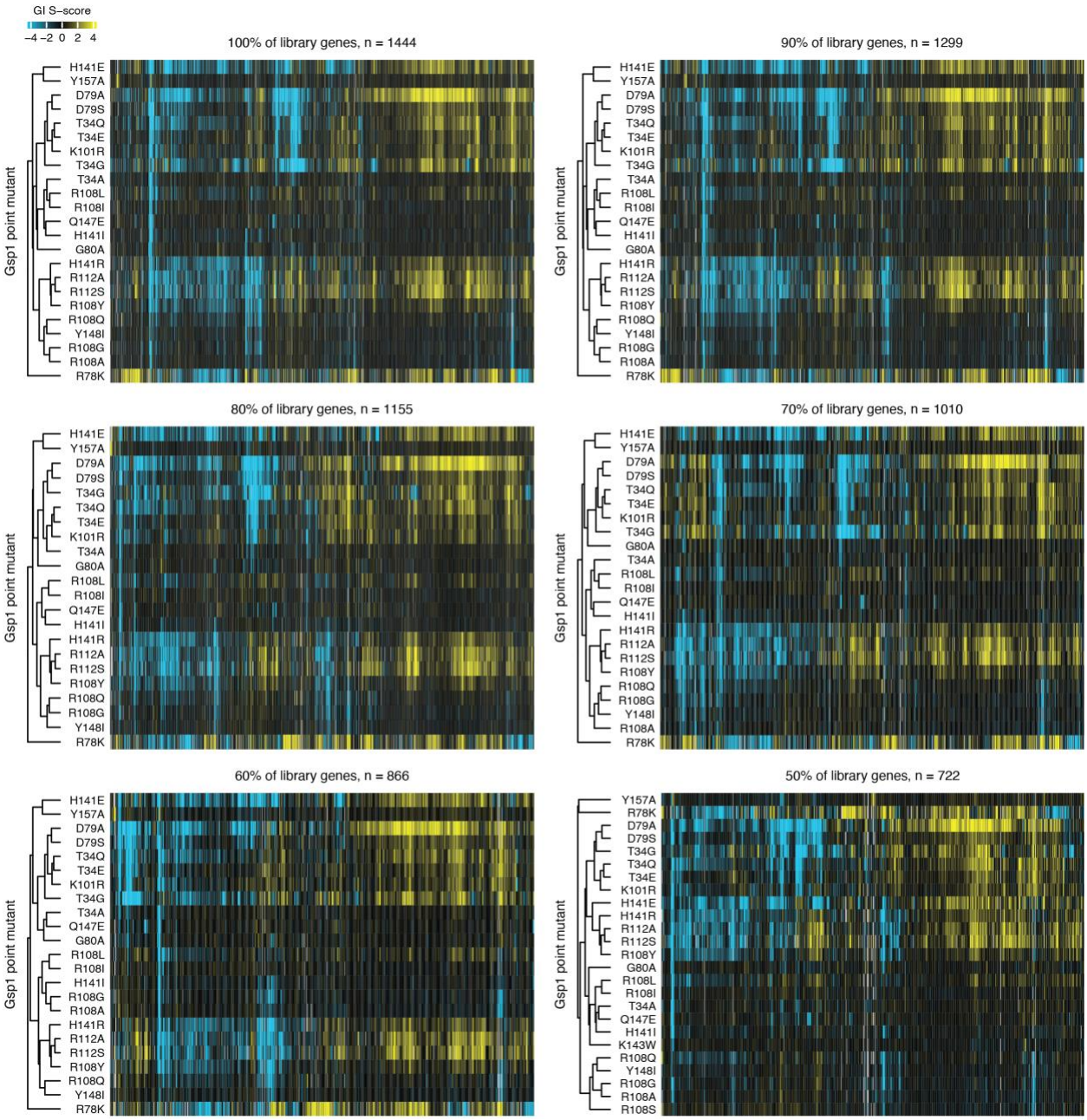


Figure 2.34 Subsampling of *S. cerevisiae* alleles maintains clustering of Gsp1 mutants based on their E-MAP profiles.

GI profiles of Gsp1 mutants. Negative S-score (blue) represents synthetic sick/lethal GIs, positive S-score (yellow) represents suppressive/epistatic GIs. Mutants and genes are hierarchically clustered by Pearson correlation. As in **Figure 2.1C**, all 55 point mutants are included in the clustering of columns, but only the dendrogram branch containing the strong mutants is shown. The clustering of mutants is robust to subsampling, with similar ordering of mutants observed down to removal of at least 60% of library genes.

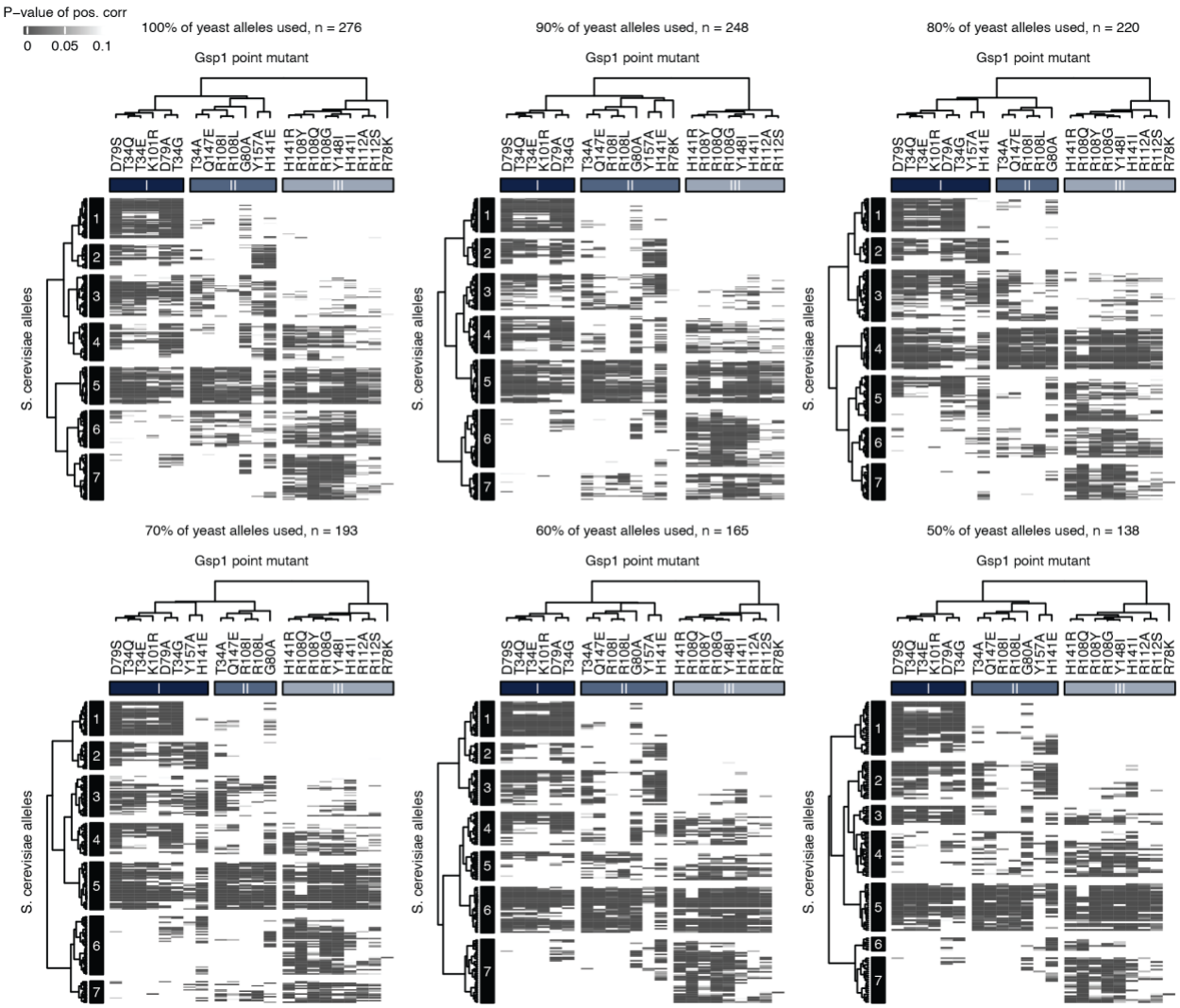


Figure 2.35 Random subsampling of *S. cerevisiae* alleles maintains clustering of Gsp1 mutants based on the p-value of Pearson correlations of their GI profiles.

The p-value is a false discovery rate adjusted one-sided (positive) p-value of the Pearson correlations (represented as gray scale). The grouping of mutants into the three observed groups is robust to subsampling, as the groups are maintained down to removal of at least 50% of alleles.

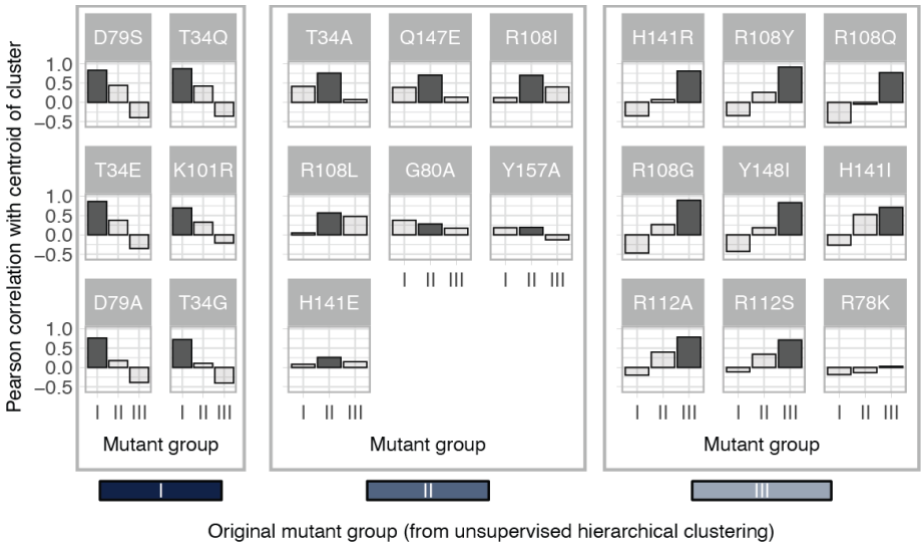


Figure 2.36 Leave-one-out analysis of Figure 2.17A.
Each bar graph shows the Pearson correlation values between the indicated withheld mutant and the centroid of each of the three groups identified by hierarchical clustering of the remaining 21 mutants. We grouped the bar graphs according to the original group to which each mutant was assigned in **Figure 2.17A**. For each of the withheld mutants, the dark bar represents the expected group (group I left bar, group II middle bar, group III right bar). With the exception of G80A, which is slightly more correlated with the group I centroid (Pearson correlation = 0.38) than the centroid of its original group, group II (Pearson correlation = 0.28), all other mutants have the highest correlation with their original groups.

Tables

Table 2.1 Co-complex X-ray crystal structures of Ran or Gsp1 with its partners.

Ran/Gsp1 binding partner				Ran/Gsp1	Overall sequence identity to *S. cerevisiae* homolog		Interface sequence identity to *S. cerevisiae* homolog	
Gene name	Partner protein name / function	PDB	source species	source species	Gsp1 [%]	partner [%]	Gsp1 [%]	partner [%]
Srm1	Guanine nucleotide exchange factor of Gsp1 (GEF)	1I2M	H. sapiens	H. sapiens	83	25	94	42
Rna1	Ran GTPase-activating protein 1 of Gsp1 (GAP)	1K5D	S. pombe	H. sapiens	83	39	84	71
Ntf2	Nuclear transport factor 2	1A2K	R. norvegicus	C. lupus	83	40	89	44
Nup1	FG-repeat nucleoporin	3CH5	R. norvegicus	H. sapiens	83	13	67	37
Nup60	FG-repeat nucleoporin	3CH5	R. norvegicus	H. sapiens	83	8	67	37
Yrb1	Ran-specific GTPase-activating protein 1	3M1I	S. cerevisiae	S. cerevisiae	100	100	100	100
Yrb2	Ran-specific GTPase-activating protein 2	3WYF	S. cerevisiae	S. cerevisiae	100	100	100	100
Srp1	Importin subunit alpha - receptor for simple and bipartite NLS	1WA5	S. cerevisiae	C. lupus	83	100	67	100
Kap95	Importin subunit beta-1 - receptor for cNLS	2BKU	S. cerevisiae	C. lupus	83	100	94	100
Crm1	Exportin-1 - Receptor for the leucine-rich nuclear export signal (NES)	3M1I	S. cerevisiae	S. cerevisiae	100	100	100	100
Los1	Exportin-T - tRNA nucleus export	3ICQ	S. pombe	S. cerevisiae	100	22	100	30
Pse1	Importin subunit beta-3 - receptor for cNLS and rg-NLS	3W3Z	S. cerevisiae	C. lupus	83	100	89	100
Kap104	Importin subunit beta-2 - receptor for rg-NLS and PY-NLS	1QBK	H. sapiens	H. sapiens	83	31	92	43
Msn5	Exportin and importin of proteins and tRNA	3A6P	H. sapiens	C. lupus	83	18	89	29
Cse1	Importin alpha re-exporter - export receptor for Srp1	1WA5	S. cerevisiae	C. lupus	83	100	89	100
Mtr10	mRNA transport regulator	4OL0	H. sapiens	H. sapiens	83	21	88	36

Table 2.2 Mutated residues in Gsp1 and their interface position and ΔrASA. CellMap alleles are annotated in parentheses.

Gsp1 residue number	Crm1	Cse1 (cse1-5002)	Kap104	Kap95 (kap95-e126k)	Los1 (los1)	Msn5 (msn5)
34				rim / 0.1		
58						
78	rim / 0.1	core / 0.34	core / 0.44	core / 0.2	rim / 0.33	rim / 0.18
79	core / 0.3	core / 0.29	support / 0.12	core / 0.37	support / 0.16	core / 0.28
80	core / 0.31	core / 0.27	core / 0.42	core / 0.29	core / 0.37	core / 0.32
84	rim / 0.3	rim / 0.21	rim / 0.3	rim / 0.41	rim / 0.09	rim / 0.09
101	rim / 0.17				rim / 0.13	rim / 0.02
102	support / 0.01				core / 0.08	
105	rim / 0.06			rim / 0.03	rim / 0.16	core / 0.25
108	core / 0.26	rim / 0.1	rim / 0.11	rim / 0.12	core / 0.43	core / 0.42
112	core / 0.55	core / 0.45	core / 0.44	core / 0.56	core / 0.4	core / 0.58
115	rim / 0.25	rim / 0.2	rim / 0.27	rim / 0.07	rim / 0.34	rim / 0.34
129		rim / 0.61	rim / 0.59	rim / 0.19		rim / 0.23
132	core / 0.12	rim / 0			rim / 0.03	rim / 0.12
137						
139	rim / 0.01			core / 0.04	rim / 0.02	
141	support / 0.14		support / 0.19	core / 0.15		
143	rim / 0.48	rim / 0.15	core / 0.35	rim / 0.27	rim / 0.09	rim / 0.01
147	core / 0.23		support / 0.23	core / 0.25	core / 0.07	core / 0.14
148	support / 0.11	support / 0.03	support / 0.13		support / 0.01	
154		core / 0.28		core / 0.38	rim / 0.13	
157	core / 0.38	rim / 0.13	core / 0.39	core / 0.29	rim / 0.05	
169	rim / 0.21			rim / 0.02		rim / 0.17
180			rim / 0.01			

Gsp1 residue number	Mtr10	Ntf2 (ntf2-h104y, ntf2-5001)	Nup1	Nup60	Pse1	Rna1 (rna1-1, rna1-s116f)
34						
58			rim / 0.28	rim / 0.28		
78	core / 0.25	core / 0.57	rim / 0	rim / 0	support / 0.03	support / 0.02
79	core / 0.36	rim / 0.1			core / 0.26	
80	core / 0.51	core / 0.27	support / 0.13	support / 0.13	core / 0.51	
84	rim / 0.36		rim / 0.25	rim / 0.25	rim / 0.2	
101						rim / 0.01
102	support / 0.1					support / 0.07
105	core / 0.21					
108	core / 0.18				rim / 0.12	
112	core / 0.57				core / 0.46	
115	rim / 0.12				rim / 0.37	
129						rim / 0
132						core / 0.44
137	core / 0.08					rim / 0.01
139	rim / 0.15				core / 0.18	
141	core / 0.14				support / 0	
143	core / 0.44				core / 0.52	
147	support / 0.09				core / 0.09	
148	support / 0.01					
154	rim / 0.03				rim / 0.05	
157	rim / 0.04					
169						
180						

Gsp1 residue number	Srm1 (srm1-g282s, srm1-ts)	Srp1 (srp1-5001)	Yrb1 (yrb1-51)	Yrb2
34			core / 0.4	rim / 0.24
58			core / 0.4	core / 0.39
78	rim / 0.46			
79	rim / 0.01			
80				
84				
101	core / 0.67	core / 0.47		
102	support / 0.15			
105	core / 0.44	rim / 0.03		
108	core / 0.47			
112	rim / 0.24			
115				
129		rim / 0.1		
132	rim / 0.16	core / 0.22		
137		core / 0.2		
139	core / 0.26	rim / 0.01		
141				
143	rim / 0.07			
147				
148				
154				
157				
169				
180			core / 0.64	rim / 0.5

Table 2.3 Gsp1 mutants and attempted yeast constructs.

Construct name	Gsp1 residue number	Gsp1 point mutation	Yeast strain successfully made
C-terminal 3xFLAG GSP1 T34L	34	T34L	yes
C-terminal 3xFLAG GSP1 T34Q	34	T34Q	yes
GSP1 T34A	34	T34A	yes
GSP1 T34D	34	T34D	yes
GSP1 T34E	34	T34E	yes
GSP1 T34G	34	T34G	yes
GSP1 T34L	34	T34L	yes
GSP1 T34Q	34	T34Q	yes
GSP1 T34S	34	T34S	yes
GSP1 T34Y	34	T34Y	yes
N-terminal 3xFLAG GSP1 T34A	34	T34A	yes
N-terminal 3xFLAG GSP1 T34E	34	T34E	yes
N-terminal 3xFLAG GSP1 T34G	34	T34G	yes
N-terminal 3xFLAG GSP1 T34L	34	T34L	yes
C-terminal 3xFLAG GSP1 F58A	58	F58A	yes
GSP1 F58A	58	F58A	yes
GSP1 F58L	58	F58L	yes
GSP1 R78K	78	R78K	yes
N-terminal 3xFLAG GSP1 R78K	78	R78K	yes
C-terminal 3xFLAG GSP1 D79A	79	D79A	yes
GSP1 D79A	79	D79A	yes
GSP1 D79S	79	D79S	yes
N-terminal 3xFLAG GSP1 D79A	79	D79A	yes
N-terminal 3xFLAG GSP1 D79S	79	D79S	yes
C-terminal 3xFLAG GSP1 G80A	80	G80A	yes
GSP1 G80A	80	G80A	yes
N-terminal 3xFLAG GSP1 G80A	80	G80A	yes
GSP1 N84Y	84	N84Y	yes
C-terminal 3xFLAG GSP1 K101R	101	K101R	yes
GSP1 K101R	101	K101R	yes
GSP1 N102I	102	N102I	yes
GSP1 N102K	102	N102K	yes
GSP1 N102M	102	N102M	yes
C-terminal 3xFLAG GSP1 N105L	105	N105L	yes
GSP1 N105L	105	N105L	yes
GSP1 N105V	105	N105V	yes
C-terminal 3xFLAG GSP1 R108A	108	R108A	yes
C-terminal 3xFLAG GSP1 R108I	108	R108I	yes
C-terminal 3xFLAG GSP1 R108Y	108	R108Y	yes
GSP1 R108A	108	R108A	yes
GSP1 R108D	108	R108D	yes
GSP1 R108G	108	R108G	yes

Construct name	Gsp1 residue number	Gsp1 point mutation	Yeast strain successfully made
GSP1 R108I	108	R108I	yes
GSP1 R108L	108	R108L	yes
GSP1 R108Q	108	R108Q	yes
GSP1 R108S	108	R108S	yes
GSP1 R108Y	108	R108Y	yes
N-terminal 3xFLAG GSP1 R108G	108	R108G	yes
N-terminal 3xFLAG GSP1 R108Y	108	R108Y	yes
C-terminal 3xFLAG GSP1 R112S	112	R112S	yes
GSP1 R112A	112	R112A	yes
GSP1 R112S	112	R112S	yes
N-terminal 3xFLAG GSP1 R112S	112	R112S	yes
GSP1 E115A	115	E115A	yes
GSP1 E115I	115	E115I	yes
GSP1 K129E	129	K129E	yes
GSP1 K129F	129	K129F	yes
GSP1 K129I	129	K129I	yes
GSP1 K129T	129	K129T	yes
C-terminal 3xFLAG GSP1 K132H	132	K132H	yes
GSP1 K132H	132	K132H	yes
N-terminal 3xFLAG GSP1 K132H	132	K132H	yes
GSP1 T137G	137	T137G	yes
GSP1 T139A	139	T139A	yes
GSP1 T139R	139	T139R	yes
C-terminal 3xFLAG GSP1 H141I	141	H141I	yes
C-terminal 3xFLAG GSP1 H141V	141	H141V	yes
GSP1 H141E	141	H141E	yes
GSP1 H141I	141	H141I	yes
GSP1 H141R	141	H141R	yes
GSP1 H141V	141	H141V	yes
N-terminal 3xFLAG GSP1 H141E	141	H141E	yes
N-terminal 3xFLAG GSP1 H141I	141	H141I	yes
N-terminal 3xFLAG GSP1 H141R	141	H141R	yes
N-terminal 3xFLAG GSP1 H141V	141	H141V	yes
C-terminal 3xFLAG GSP1 K143W	143	K143W	yes
GSP1 K143H	143	K143H	yes
GSP1 K143W	143	K143W	yes
GSP1 K143Y	143	K143Y	yes
N-terminal 3xFLAG GSP1 K143W	143	K143W	yes
GSP1 Q147E	147	Q147E	yes
GSP1 Q147L	147	Q147L	yes
N-terminal 3xFLAG GSP1 Q147E	147	Q147E	yes
C-terminal 3xFLAG GSP1 Y148I	148	Y148I	yes

Construct name	Gsp1 residue number	Gsp1 point mutation	Yeast strain successfully made
GSP1 Y148I	148	Y148I	yes
N-terminal 3xFLAG GSP1 Y148I	148	Y148I	yes
GSP1 K154M	154	K154M	yes
C-terminal 3xFLAG GSP1 Y157A	157	Y157A	yes
GSP1 Y157A	157	Y157A	yes
GSP1 K169I	169	K169I	yes
C-terminal 3xFLAG GSP1 A180T	180	A180T	yes
GSP1 A180T	180	A180T	yes
N-terminal 3xFLAG GSP1 A180T	180	A180T	yes
T34A Cter3xFL	34	T34A	no
T34E Cter3xFL	34	T34E	no
T34G Cter3xFL	34	T34G	no
T34Q Nter3xFL	34	T34Q	no
K39M	39	K39M	no
Y41A	41	Y41A	no
V49D	49	V49D	no
F58A Nter3xFL	58	F58A	no
G70N	70	G70N	no
Q71E	71	Q71E	no
K73Q	73	K73Q	no
G75N	75	G75N	no
R78K Cter3xFL	78	R78K	no
D79K	79	D79K	no
D79S Cter3xFL	79	D79S	no
G80N	80	G80N	no
G80S	80	G80S	no
I98F	98	I98F	no
K101R Nter3xFL	101	K101R	no
R108G Cter3xFL	108	R108G	no
R108I Nter3xFL	108	R108I	no
R108L Nter3xFL	108	R108L	no
R108Q Cter3xFL	108	R108Q	no
R108S Cter3xFL	108	R108S	no
K132M	132	K132M	no
K132Y	132	K132Y	no
T137E	137	T137E	no
H141E Cter3xFL	141	H141E	no
H141R Cter3xFL	141	H141R	no
Q147E Cter3xFL	147	Q147E	no
Y157K	157	Y157K	no

Table 2.4 Pearson correlations between Gsp1 mutants and the alleles of their direct interaction partners from the SGA CellMap. Ordered by correlation value

GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core		GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core
D79S	kap95-e126k	0.4146	TRUE		H141I	crm1_damp	0.2706	FALSE
Y148I	crm1_damp	0.4027	FALSE		T34Q	kap95-e126k	0.2681	FALSE
R108L	ntf2-h104y	0.3827	FALSE		T34A	yrb1-51	0.2676	TRUE
R108G	crm1_damp	0.3783	TRUE		T34Q	ntf2-5001	0.2588	FALSE
R108L	ntf2-5001	0.3612	FALSE		Y148I	ntf2-5001	0.2555	FALSE
R108Y	ntf2-h104y	0.3612	FALSE		K101R	ntf2-5001	0.255	FALSE
G80A	kap95-e126k	0.3545	TRUE		R108I	ntf2-5001	0.2544	FALSE
R112A	ntf2-h104y	0.3533	FALSE		D79A	srm1-ts	0.251	FALSE
R108Y	crm1_damp	0.3453	TRUE		D79S	yrb1-51	0.2502	FALSE
K101R	ntf2-h104y	0.3389	FALSE		H141I	ntf2-5001	0.2501	FALSE
R112S	ntf2-h104y	0.3353	FALSE		T34G	cse1-5002	0.2467	FALSE
R108Q	crm1_damp	0.3291	TRUE		D79A	ntf2-h104y	0.2459	FALSE
T34A	ntf2-5001	0.3231	FALSE		K101R	kap95-e126k	0.2449	FALSE
Q147E	kap95-e126k	0.323	TRUE		T34Q	yrb1-51	0.241	TRUE
H141R	crm1_damp	0.3199	FALSE		G80A	ntf2-h104y	0.2402	TRUE
K101R	srm1-ts	0.3197	TRUE		T34G	kap95-e126k	0.2365	FALSE
T34E	srm1-ts	0.3155	FALSE		K101R	srm1-g282s	0.2359	TRUE
T34Q	ntf2-h104y	0.3135	FALSE		G80A	cse1-5002	0.2357	TRUE
R112A	ntf2-5001	0.3117	FALSE		Y148I	ntf2-h104y	0.2355	FALSE
T34E	ntf2-h104y	0.3116	FALSE		T34E	ntf2-5001	0.2354	FALSE
R108Y	ntf2-5001	0.3091	FALSE		G80A	yrb1-51	0.2354	FALSE
D79S	ntf2-h104y	0.309	FALSE		D79S	srp1-5001	0.2343	FALSE
D79S	srm1-ts	0.3085	FALSE		T34G	srm1-ts	0.2321	FALSE
R112S	ntf2-5001	0.3043	FALSE		G80A	crm1_damp	0.2317	TRUE
D79S	cse1-5002	0.3022	TRUE		H141R	ntf2-5001	0.2296	FALSE
T34Q	srm1-ts	0.3015	FALSE		T34A	crm1_damp	0.2291	FALSE
T34A	ntf2-h104y	0.2946	FALSE		R108I	ntf2-h104y	0.2275	FALSE
H141R	ntf2-h104y	0.2929	FALSE		G80A	ntf2-5001	0.2275	TRUE
T34E	yrb1-51	0.2898	TRUE		T34E	srm1-g282s	0.2249	FALSE
T34A	srm1-ts	0.2881	FALSE		R108Q	ntf2-h104y	0.2245	FALSE
T34E	kap95-e126k	0.2813	FALSE		G80A	srm1-ts	0.2188	FALSE
T34A	kap95-e126k	0.2791	FALSE		H141E	rna1-s116f	0.2185	FALSE
R108L	srm1-ts	0.2773	TRUE		R108L	crm1_damp	0.2176	TRUE
D79A	kap95-e126k	0.2754	TRUE		D79S	ntf2-5001	0.2171	FALSE

GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core		GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core
R108G	ntf2-h104y	0.2171	FALSE		D79A	srp1-5001	0.171	FALSE
Q147E	srm1-ts	0.2142	FALSE		H141E	ntf2-h104y	0.1682	FALSE
R108G	ntf2-5001	0.2131	FALSE		D79A	yrb1-51	0.1672	FALSE
T34A	cse1-5002	0.2104	FALSE		R108G	yrb2_damp	0.1669	FALSE
Q147E	ntf2-h104y	0.2101	FALSE		Y148I	yrb2_damp	0.1654	FALSE
R108Y	srm1-ts	0.208	TRUE		D79S	srm1-g282s	0.1652	FALSE
R112A	crm1_damp	0.2075	TRUE		R108Y	yrb2_damp	0.165	FALSE
H141E	kap95-e126k	0.2073	FALSE		R108Y	kap95-e126k	0.1645	FALSE
R108I	srm1-ts	0.2059	TRUE		T34A	rna1-s116f	0.1637	FALSE
D79A	cse1-5002	0.2047	TRUE		D79A	ntf2-5001	0.1621	FALSE
Q147E	ntf2-5001	0.2044	FALSE		H141R	srm1-ts	0.1596	FALSE
T34Q	srm1-g282s	0.2038	FALSE		T34A	srp1-5001	0.1557	FALSE
H141I	ntf2-h104y	0.2025	FALSE		D79A	srm1-g282s	0.1534	FALSE
T34E	cse1-5002	0.1994	FALSE		T34Q	srp1-5001	0.1529	FALSE
R112S	srm1-ts	0.1945	FALSE		H141E	rna1-1	0.1529	FALSE
Q147E	yrb1-51	0.1942	FALSE		R108L	srm1-g282s	0.1527	TRUE
R108Q	ntf2-5001	0.1936	FALSE		Q147E	srm1-g282s	0.1524	FALSE
H141I	srm1-ts	0.1934	FALSE		Y157A	crm1_damp	0.1495	TRUE
R112A	srm1-ts	0.1929	FALSE		Q147E	cse1-5002	0.148	FALSE
K101R	yrb1-51	0.1912	FALSE		T34G	srm1-g282s	0.1476	FALSE
R108Q	yrb2_damp	0.1895	FALSE		R108G	srm1-ts	0.1462	TRUE
R112S	crm1_damp	0.1894	TRUE		T34G	ntf2-h104y	0.1453	FALSE
H141I	kap95-e126k	0.188	FALSE		H141E	srm1-ts	0.1446	FALSE
H141E	cse1-5002	0.1817	FALSE		Q147E	srp1-5001	0.1417	FALSE
T34E	srp1-5001	0.1805	FALSE		R108I	srm1-g282s	0.139	TRUE
T34G	yrb1-51	0.1781	TRUE		H141E	crm1_damp	0.138	FALSE
T34G	srp1-5001	0.1779	FALSE		Y148I	kap95-e126k	0.134	FALSE
G80A	srp1-5001	0.1775	FALSE		R112A	kap95-e126k	0.1302	TRUE
H141E	ntf2-5001	0.1762	FALSE		T34E	rna1-1	0.1295	FALSE
R108L	kap95-e126k	0.1753	FALSE		Q147E	crm1_damp	0.1269	FALSE
T34Q	cse1-5002	0.1738	FALSE		H141I	yrb2_damp	0.1254	FALSE
T34A	srm1-g282s	0.1729	FALSE		D79A	rna1-s116f	0.1242	FALSE
R108I	kap95-e126k	0.1719	FALSE		R108I	crm1_damp	0.1232	TRUE
H141R	yrb2_damp	0.1717	FALSE		T34Q	rna1-s116f	0.1232	FALSE

GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core		GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core
R108Q	srm1-ts	0.1214	TRUE		H141E	yrb2_damp	0.0828	FALSE
T34A	rna1-1	0.1214	FALSE		H141I	yrb1-51	0.082	FALSE
T34G	rna1-1	0.1199	FALSE		R108I	msn5	0.0815	TRUE
R112S	yrb2_damp	0.1175	FALSE		H141I	cse1-5002	0.0795	FALSE
R108I	yrb1-51	0.1171	FALSE		D79S	crm1_damp	0.0789	TRUE
Y157A	rna1-s116f	0.1168	FALSE		R108I	los1	0.075	TRUE
R108G	kap95-e126k	0.1162	FALSE		T34Q	crm1_damp	0.0745	FALSE
R112S	srm1-g282s	0.1154	FALSE		G80A	rna1-1	0.0735	FALSE
H141R	kap95-e126k	0.115	FALSE		T34E	crm1_damp	0.0732	FALSE
K101R	srp1-5001	0.1149	TRUE		R108I	cse1-5002	0.0721	FALSE
Q147E	rna1-s116f	0.1139	FALSE		K101R	los1	0.0704	FALSE
H141E	srp1-5001	0.1135	FALSE		H141I	rna1-1	0.0702	FALSE
R112S	kap95-e126k	0.1112	TRUE		R108L	los1	0.0699	TRUE
D79S	rna1-s116f	0.1081	FALSE		T34A	los1	0.0695	FALSE
G80A	srm1-g282s	0.1073	FALSE		Q147E	rna1-1	0.0666	FALSE
H141I	srm1-g282s	0.1062	FALSE		Y148I	yrb1-51	0.0651	FALSE
H141I	rna1-s116f	0.1045	FALSE		Y157A	cse1-5002	0.0637	FALSE
R108Y	srm1-g282s	0.104	TRUE		R108G	srm1-g282s	0.0626	TRUE
T34G	rna1-s116f	0.1031	FALSE		T34Q	los1	0.0616	FALSE
R112A	srm1-g282s	0.1025	FALSE		R108L	msn5	0.0612	TRUE
H141E	yrb1-51	0.1023	FALSE		R108G	rna1-1	0.0609	FALSE
T34E	rna1-s116f	0.1013	FALSE		Y157A	ntf2-5001	0.0609	FALSE
R112A	yrb2_damp	0.0975	FALSE		R108Q	kap95-e126k	0.0608	FALSE
T34Q	rna1-1	0.0959	FALSE		R108Q	msn5	0.0592	TRUE
G80A	rna1-s116f	0.0947	FALSE		H141I	srp1-5001	0.059	FALSE
Y148I	srm1-ts	0.0933	FALSE		R108G	msn5	0.0587	TRUE
T34E	los1	0.0932	FALSE		R108G	rna1-s116f	0.0578	FALSE
Y157A	kap95-e126k	0.092	TRUE		H141R	srm1-g282s	0.0573	FALSE
R108L	yrb1-51	0.0904	FALSE		D79A	rna1-1	0.0562	FALSE
D79S	rna1-1	0.089	FALSE		G80A	yrb2_damp	0.0522	FALSE
Y157A	rna1-1	0.0878	FALSE		R108I	srp1-5001	0.0516	FALSE
K101R	cse1-5002	0.0869	FALSE		D79S	los1	0.0504	FALSE
Y148I	rna1-s116f	0.086	FALSE		Y148I	rna1-1	0.0501	FALSE
Y157A	yrb1-51	0.0843	FALSE		Y148I	srm1-g282s	0.0482	FALSE

GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core
D79A	crm1_damp	0.0482	TRUE
R108L	srp1-5001	0.047	FALSE
R108I	rna1-1	0.045	FALSE
R112S	rna1-s116f	0.044	FALSE
Q147E	los1	0.0436	FALSE
R112S	rna1-1	0.0432	FALSE
R108Y	los1	-0.0089	TRUE
R78K	srp1-5001	-0.0124	FALSE
R108G	los1	-0.0126	TRUE
R78K	los1	-0.0127	FALSE
R112A	cse1-5002	-0.013	TRUE
Y157A	msn5	-0.0139	FALSE
R112S	los1	-0.0142	TRUE
D79A	msn5	-0.0197	TRUE
T34G	crm1_damp	-0.0216	FALSE
H141R	rna1-1	-0.0229	FALSE
R78K	msn5	-0.023	FALSE
H141I	los1	-0.0242	FALSE
H141I	msn5	-0.0244	FALSE
Y148I	msn5	-0.0253	FALSE
K101R	msn5	-0.0271	FALSE
T34Q	yrb2_damp	-0.0272	TRUE
R108G	srp1-5001	-0.0277	FALSE
R112S	cse1-5002	-0.0303	TRUE
R108Q	cse1-5002	-0.0317	FALSE
R108L	rna1-s116f	-0.032	FALSE
Q147E	msn5	-0.0322	FALSE
R78K	ntf2-h104y	-0.0322	TRUE
T34E	yrb2_damp	-0.033	TRUE
H141R	msn5	-0.0364	FALSE
R108Q	rna1-s116f	-0.0378	FALSE
T34Q	msn5	-0.0382	FALSE
Y157A	srm1-g282s	-0.039	FALSE
R78K	ntf2-5001	-0.0426	TRUE

GSP1 mutant	Partner strain name	Pearson correlation coefficient	Residue in interface core
H141E	msn5	-0.0441	FALSE
D79S	msn5	-0.0471	TRUE
T34G	msn5	-0.0479	FALSE
R78K	kap95-e126k	-0.0485	FALSE
H141R	los1	-0.0521	FALSE
R108Q	los1	-0.0593	TRUE
Y148I	los1	-0.0601	FALSE
K101R	yrb2_damp	-0.0651	FALSE
R108Q	srp1-5001	-0.0685	FALSE
R108Q	yrb1-51	-0.0696	FALSE
R78K	cse1-5002	-0.0696	TRUE
R78K	rna1-1	-0.0711	FALSE
T34E	msn5	-0.0741	FALSE
R78K	yrb1-51	-0.0814	FALSE
R78K	crm1_damp	-0.0839	TRUE
G80A	msn5	-0.0892	TRUE

Table 2.5 Interquartile range (IQR) of log2(fold change) values across all the Gsp1 mutants for each prey protein identified. Ordered by IQR.

Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)
Spa2	14.10		Apa1	2.53		Tub1	1.79		Mph1	1.55
Pup2	10.27		Eno1	2.49		Rpl39	1.78		Rpl26a	1.55
Cdc3	9.65		Hpm1	2.48		Swc5	1.78		Taf2	1.51
Rna1	6.86		Srm1	2.46		Ugp1	1.77		Net1	1.51
Mae1	6.34		Scj1	2.44		Tif4631	1.77		Msh2	1.51
Hrp1	6.25		Rtp1	2.44		Aim36	1.77		Egd1	1.49
Spb1	6.23		Tub3	2.34		Dbp3	1.76		Rpl29	1.47
Adr1	5.92		Idh2	2.30		Pln1	1.73		Hmo1	1.47
Rgr1	5.83		Idh1	2.20		Rpl37a	1.73		Tdh3	1.47
Ecm1	5.69		Tub2	2.19		Svf1	1.72		Azf1	1.46
Swi1	5.66		Aro9	2.17		Thi20	1.72		Nop2	1.45
Yar1	5.65		Krr1	2.17		Mcm6	1.70		Rps0b	1.44
Cmr1	5.30		Cia2	2.14		Npa3	1.70		Rpc82	1.43
Acf4	5.26		Ade5,7	2.13		Krs1	1.68		Ssa1	1.43
Vps71	5.25		Ura7	2.08		Siw14	1.68		Gbp2	1.42
Kri1	5.12		Afg2	2.02		Cbf5	1.67		Lcl2	1.42
Lcp5	5.09		Yap1	2.01		Zuo1	1.67		Rpp1a	1.41
Gcd14	4.99		Hef3	2.00		Rvb1	1.65		Mgm101	1.40
Srp54	4.94		Hsp60	1.99		Wtm1	1.65		Gfa1	1.39
Reh1	4.79		Rpa135	1.98		Vps13	1.64		Grs1	1.38
Gcd10	4.79		Vps1	1.98		Cdc14	1.64		Mcm4	1.38
Tdh1	4.78		Rvb2	1.96		Dpb4	1.64		Puf6	1.38
Srp68	4.75		Yrb30	1.96		Yku70	1.63		Rpl10	1.37
Srp1	4.61		Dep1	1.93		Fun12	1.62		Tra1	1.37
Rpc37	4.04		San1	1.92		Pwp1	1.61		Pro3	1.37
Kap120	3.24		Frs1	1.91		Rpc34	1.61		Nop4	1.35
Pol2	3.23		Rpc31	1.90		Aco1	1.60		Tfc3	1.35
Kap95	3.05		Oca1	1.90		Spt8	1.59		Spt20	1.35
Rix7	2.93		Mtc1	1.89		Orc1	1.58		Rpl3	1.34
Yef3	2.85		Tti1	1.87		Pse1	1.58		Rpl33b	1.33
Rpb8	2.62		Yrb1	1.87		Pdi1	1.57		Cdc9	1.33
Gpn3	2.59		Ptc3	1.85		Rpa190	1.57		Ubp15	1.32
Rea1	2.56		Sdd3	1.84		Sum1	1.56		Rpc11	1.31
Paa1	2.55		Dbp2	1.80		Yku80	1.56		Rpo21	1.31

Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)		Prey name	Interquartile range (IQR)
Rlp24	1.31		Rtg3	1.16		Rpc40	1.01		Rpl8b	0.87
Skn7	1.31		Rfm1	1.15		Ade3	1.01		Rsc6	0.87
Hsp42	1.31		Gdh1	1.14		Hho1	1.01		Mog1	0.86
Cys4	1.30		Sry1	1.13		Rpl5	1.00		Rsc9	0.84
Orc2	1.30		Chd1	1.12		Stm1	1.00		Rsc3	0.84
Hca4	1.29		Top2	1.11		Reb1	1.00		Gcd1	0.83
Ree1	1.29		Rpl31b	1.11		Ioc4	0.99		Rfc2	0.83
Ssz1	1.27		Cst6	1.11		Asg1	0.99		Swi3	0.83
Yta7	1.27		Rpl36a	1.10		Ioc3	0.99		Ies5	0.82
Pre6	1.26		Rpl4a	1.10		Msn1	0.99		Ioc2	0.82
Gtr2	1.26		Abf2	1.09		Adh6	0.99		Imh1	0.81
Hal5	1.25		Rpb5	1.09		Rpc19	0.99		Oye2	0.80
Rpb4	1.24		Spt7	1.08		Rpc53	0.98		Ies1	0.80
Nop6	1.24		Orc4	1.08		Adh3	0.98		Nhp10	0.79
Rpc25	1.23		Sin3	1.08		Rbg1	0.98		Arp9	0.79
Muk1	1.23		Rpo31	1.07		Raf1	0.98		Spt16	0.77
Caf40	1.23		Nur1	1.07		Orc3	0.98		Sfh1	0.76
Aat1	1.23		Rpb10	1.06		Rfc1	0.96		Htb2	0.76
Msh3	1.23		Sko1	1.06		Srl2	0.96		Enp2	0.76
Spt5	1.22		Rpp2b	1.06		Rpl24a	0.96		Taf10	0.76
Rok1	1.22		Rpo26	1.05		Top1	0.95		Bur6	0.76
Swr1	1.21		Vps72	1.05		Rpl6b	0.95		Isw2	0.75
Irc20	1.20		Rpl30	1.04		Isw1	0.95		Rsc2	0.75
Rpp2a	1.20		Hri1	1.04		Sth1	0.94		Taf5	0.74
Rim1	1.20		Nop10	1.03		Nhp2	0.94		Ies3	0.74
Hpc2	1.19		Pdr1	1.03		Egd2	0.93		Rpb11	0.74
Mcm5	1.19		Ald4	1.03		Npl6	0.90		Arp8	0.73
Rpl15a	1.19		Yra1	1.03		Rps29a	0.90		Rtt102	0.73
Rpa49	1.19		Nip7	1.02		Taf9	0.89		Cdc1	0.72
Ret1	1.18		Prp43	1.02		Gar1	0.89		Pob3	0.69
Tkl2	1.17		Rtt106	1.02		Rsc4	0.89		Htz1	0.68
Hst1	1.16		Hir2	1.02		Snf2	0.88		Spt15	0.68
Rpl9a	1.16		Arp5	1.02		Taf14	0.88		Rsc58	0.67
Elo1	1.16		Itc1	1.02		Grx1	0.87		Hir1	0.64

Prey name	Interquartile range (IQR)
Rfc3	0.62
Hos3	0.61
Mot1	0.61
Rsc8	0.61
Arp4	0.60
Pre2	0.60
Ies2	0.60
Arp7	0.55
Rpc10	0.54
Ant1	0.54
Abf1	0.54
Thi7	0.52
Lsm6	0.51
Rfc5	0.49
Hir3	0.49
Srp14	0.48
Rco1	0.45
Rfc4	0.43
Aim14	0.37
Sis1	0.33
Aah1	0.00
Aim41	0.00
Arl3	0.00
Cfd1	0.00
Gcn3	0.00
Lrs4	0.00
Opi1	0.00
Rad5	0.00
Rpl21b	0.00
Rrp8	0.00
Rrs1	0.00
Sen1	0.00
Slx9	0.00
Smc2	0.00
Snf12	0.00
Snf5	0.00
Spp41	0.00
Stb4	0.00
Sti1	0.00
Sub1	0.00
Tif4632	0.00

Table 2.6 Michaelis-Menten parameters of GAP-mediated GTP hydrolysis. The two Michaelis-Menten parameters and their ratio (enzymatic efficiency) are determined by an integrated Michaelis-Menten fit for each individual experiment. Standard error is based on three or more replicates.

Gsp1 mutant	k_{cat} [s^{-1}]	std.error k_{cat} [s^{-1}]	K_m [μM]	std.error K_m [μM]	k_{cat}/K_m [s^{-1} μM^{-1}]	std.error k_{cat}/K_m [s^{-1} μM^{-1}]
WT	9.2	0.66	0.4	0.04	26.0	2.57
T34A	9.8	3.65	2.3	0.63	4.0	0.56
T34E	8.9	0.23	1.4	0.09	6.5	0.36
T34G	5.0	0.81	0.8	0.12	7.1	0.99
T34L	15.2	1.27	2.0	0.10	7.5	0.88
T34Q	5.4	0.20	2.2	0.26	2.5	0.23
F58A	8.6	0.57	0.2	0.03	35.8	2.97
R78K	4.3	0.73	2.1	0.59	2.4	0.35
D79A	11.9	2.21	3.6	1.11	3.8	0.62
D79S	4.1	0.32	1.7	0.23	3.0	0.59
G80A	8.8	0.14	0.3	0.01	28.8	1.56
K101R	8.2	1.22	0.2	0.01	44.7	9.20
R108A	7.8	0.32	0.2	0.01	42.0	4.14
R108G	9.2	0.16	0.1	0.01	82.3	5.74
R108I	13.2	2.24	3.1	0.66	4.3	0.15
R108L	5.2	0.63	0.3	0.07	19.3	2.87
R108Q	9.2	0.03	0.2	0.00	61.2	1.18
R108Y	7.8	1.39	0.2	0.07	40.1	6.34
R112S	4.9	1.28	3.0	1.01	1.7	0.20
K132H	6.7	0.45	5.6	0.13	1.2	0.06
H141R	7.2	1.19	0.1	0.02	56.3	3.04
K143W	9.5	0.86	0.1	0.02	71.8	3.48
Q147E	7.6	0.65	0.7	0.04	11.6	1.58
Y157A	8.8	1.89	0.2	0.03	57.7	4.87
A180T	4.0	0.49	0.4	0.04	11.1	0.29

Table 2.7 Michaelis-Menten parameters of GEF-mediated nucleotide exchange. Standard error is based on the error of the Michaelis-Menten fit to the data.

Gsp1 mutant	k_{cat} [s^{-1}]	std.error k_{cat} [s^{-1}]	K_m [μM]	std.error K_m [μM]	k_{cat}/K_m [s^{-1} μM^{-1}]	std.error k_{cat}/K_m [s^{-1} μM^{-1}]
WT	3.0	0.08	0.9	0.12	3.3	0.44
T34A	1.8	0.10	0.9	0.22	2.1	0.55
T34E	1.7	0.07	1.0	0.17	1.7	0.29
T34G	2.5	0.14	1.4	0.28	1.8	0.39
T34L	2.0	0.11	1.6	0.35	1.2	0.27
T34Q	1.3	0.05	1.0	0.14	1.3	0.20
F58A	1.9	0.06	1.6	0.16	1.2	0.13
R78K	3.5	0.19	10.2	1.43	0.3	0.05
D79A	3.2	0.14	2.6	0.31	1.2	0.15
D79S	2.2	0.12	0.9	0.21	2.6	0.64
G80A	1.2	0.10	1.0	0.33	1.2	0.39
K101R	4.0	0.42	304.9	50.52	0.0	0.00
R108A	3.0	0.13	0.9	0.16	3.2	0.56
R108G	5.4	0.12	8.5	0.55	0.6	0.04
R108I	8.1	0.55	149.2	15.73	0.1	0.01
R108L	3.4	0.08	49.2	2.95	0.1	0.00
R108Q	3.8	0.10	8.7	0.64	0.4	0.03
R108Y	4.5	0.14	19.3	1.59	0.2	0.02
R112S	0.8	0.12	4.1	1.28	0.2	0.07
K132H	1.9	0.17	1.6	0.49	1.1	0.35
H141R	0.6	0.03	0.5	0.13	1.2	0.30
K143W	1.2	0.08	0.6	0.20	1.8	0.57
Q147E	1.9	0.07	1.4	0.18	1.4	0.19
Y157A	1.0	0.06	1.0	0.24	0.9	0.22
A180T	2.3	0.05	1.2	0.09	2.0	0.16

Table 2.8 Intrinsic GTP hydrolysis rate of wild type and mutant Gsp1. Standard deviation is based on data from 3 or more replicates.

Gsp1 mutant	Intrinsic GTP hydrolysis rate [s^{-1}]	std.error of intrinsic GTP hydrolysis rate [s^{-1}]
WT	2.5E-05	1.2E-06
T34A	7.4E-06	3.0E-06
T34E	8.7E-06	1.1E-06
T34G	2.0E-05	1.9E-06
T34L	1.8E-05	3.7E-07
T34Q	6.6E-06	3.0E-06
F58A	2.1E-05	2.7E-07
R78K	8.0E-06	3.9E-06
D79A	4.3E-05	1.2E-05
D79S	1.8E-05	2.9E-06
G80A	1.5E-05	7.3E-07
K101R	2.7E-05	2.1E-06
R108A	1.4E-05	4.9E-07
R108G	1.9E-05	1.2E-06
R108I	3.4E-05	8.8E-06
R108L	1.9E-05	9.4E-07
R108Q	1.9E-05	5.0E-07
R108Y	2.0E-05	2.4E-06
R112S	1.6E-05	5.9E-06
K132H	3.3E-05	4.9E-06
H141R	3.1E-05	8.8E-07
K143W	2.9E-05	7.6E-07
Q147E	1.6E-05	9.6E-08
Y157A	3.9E-05	5.5E-06
A180T	2.7E-05	1.4E-06

Table 2.9 Apparent T_m values estimated from the circular dichroism (CD) thermal melts. Mutants are ordered by apparent T_m.

Gsp1 mutant	Apparent Tm [ºC]
R78K	79
G80A	77
T34G	77
R108Y	77
N105L	77
R108G	77
WT	76
T34L	76
K101R	76
R108Q	75
R108I	74
A180T	74
K132H	74
Q147E	73
R108L	73
K143W	73
D79S	72
R112S	71
H141I	66
H141V	63
H141R	63
Y157A	63

Chapter 3. A complete allosteric map of a GTPase switch in its native network

Summary

Allosteric regulation is central to protein function in cellular networks.[77] However, despite technological advances[157,158] most studies of allosteric effects on function are conducted in heterologous environments,[157,159,160] limiting the discovery of allosteric mechanisms that rely on endogenous binding partners or posttranslational modifications to modulate activity. Here we report an approach that enables probing of new sites of allosteric regulation at residue-level resolution in essential eukaryotic proteins in their native biological context by comprehensive mutational scanning. We apply our approach to the central GTPase Gsp1/Ran. GTPases are highly regulated molecular switches that control signaling, with switching occurring via catalyzed GTP hydrolysis and nucleotide exchange. We find that 28% of 4,315 assayed mutations in Gsp1/Ran are highly deleterious, showing a toxic response identified by our assay as gain-of-function (GOF). Remarkably, a third of all positions enriched for GOF mutations (20/60) are outside the GTPase active site. Kinetic analysis shows that these distal sites are allosterically coupled to the active site, including a novel cluster of sites that alter the nucleotide preference of Gsp1 from GDP to GTP. We describe multiple distinct mechanisms by which allosteric mutations alter Gsp1/Ran cellular function by modulating GTPase switching. Our systematic discovery of new regulatory sites provides a functional map relevant to other GTPases such as Ras that could be exploited for targeting and reprogramming critical biological processes.

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.[77] Protein or ligand binding, posttranslational

modifications, and mutations can allosterically alter subsequent binding events or enzymatic activities to control metabolism[158] or signaling,[161,162] making allosteric regulation a driver of disease and attractive target for therapeutic drug design.[163] While it has been suggested that a considerable fraction of protein residues may be primed for allosteric regulation[157] and this priming may enable the evolution of new functional protein-protein interactions,[164] 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,[157] 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 methods for discovering new 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,[19] and whose misregulation is often associated with disease.[20] 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,[18] have been shown to affect the kinetics of biochemical switching function *in vitro*[161] and to lead to switch overactivation[159,160] and altered cellular function.[161] Additionally, one allosteric site of the GTPase Ras has been successfully targeted by small molecule inhibitors.[165] Despite these key findings, the vast majority of GTPase sites remain untested for allosteric regulation in their native biological networks[111]

when the functional context of opposing regulators, posttranslational modifications, interaction partners, and downstream signaling pathways is preserved (**Figure 3.1A**).

Here we introduce an approach to generate a complete allosteric 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.[59,61] 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.[161] Gsp1 is highly conserved, with 82% of its amino acid sequence identical to the human homolog Ran. With some notable exceptions,[160,166] prior mutational scanning experiments have revealed a tolerance to mutations even among highly conserved proteins,[167] suggesting missing biological context.[166,168] 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.

Results

Comprehensive mutational perturbation of Gsp1.

To systematically measure the effect of all Gsp1 mutations on cellular function (**Figure 3.1A**), we developed an approach derived from our EMPIRIC (extremely methodical and parallel investigation of randomized individual codons) method[60] but with a generalizable plasmid dropout selection to probe the function of essential genes (**Figure 3.1B**, 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 3.2**). 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 3.1C**, 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 3.1D**, 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%) were beneficial. 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 positions contacting the nucleotide (**Figure 3.1C** and **Figure 3.3**).

The distribution of STOP codon scores (**Figure 3.1D, 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). Residues 1-174 comprise a standard GTPase fold, whereas residues 175-219 comprise a C-terminal extension specific to the Ran subfamily (**Figure 3.1C** and **Figure 3.3A**). Thus, the first set of STOP codon mutants (residues 1-174) likely represent the growth defect of a null Gsp1 mutant, as

internal truncations in the GTPase fold likely result in nonfunctional proteins. 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 of 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 3.4A**).

Mapping structural locations of toxic/GOF mutations.

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 empirical distribution to a null distribution parameterized according to the total number of toxic/GOF mutations in the dataset (**Figure 3.5**, 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 residues were identical in amino acid identity between *S. cerevisiae* Gsp1 and human Ran.

Given most substantial fitness effects observed in mutational perturbation studies are typically from mutations at positions in active sites required for function, or at positions 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**

3.4B, blue) and an additional 10 positions are in the C-terminal extension. Thus, 20/60 toxic/GOF positions are at positions in the GTPase fold but distal to the active site (**Figure 3.4B,** 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 3.4C**, red), and 43 out of the 62 core positions are not toxic/GOF (**Figure 3.4C**, orange). Moreover, mutations in the active site would typically be expected to ablate function and therefore lead to a loss-of-function phenotype (similar to STOP). However, we observe 517 (61%) toxic/GOF mutations in the active site compared to only 93 (11%) STOP-like mutations (**Figure 3.4A**). Similarly, mutations 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 3.6, Figure 3.7**). 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.

Functional roles of toxic/GOF mutants.

The prevalence of toxic/GOF mutations in the C-terminal extension (**Figure 3.1C, 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,[169] which may explain the enrichment of cancer mutations in the C-terminus of Ran.[170] We therefore asked whether all toxic/GOF mutations 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 3.4B**), 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,[161] consistent with the proposed effect of mutations on regulated switching; and S155 is a known phosphorylation site[171] neighboring the conserved G5 SAK motif in the active site (**Figure 3.1C**). 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[172] (**Figure 3.8A**). We verified that toxic/GOF mutants 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 3.8B**), and that a C-terminal deletion variant was as toxic as the toxic/GOF mutations at these positions.

To examine whether toxic/GOF mutations perturbed switch function in this unexplained set of mutants, 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 positions that are clustered in the structure but distal from the active site. All purified mutants were well-folded and stable (**Figure 3.9, Figure 3.10**). 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 3.8C, Figure 3.11A**, Methods). All mutants except F159L had reduced or similar GEF-catalyzed nucleotide exchange rates compared to WT (**Figure 3.11B**). 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 (**Fig. 3.8D, Figure 3.11A-C**). 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 3.11D**). 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 3.12A, 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 3.12A**). These sites are at least 5 and up to 30Å away from the nucleotide (**Figure 3.12C**), showing that our method quantifying perturbations to cellular function in the native network identifies many non-local sites of allosteric regulation, even surpassing a recent study of allostery quantifying effects on biophysical function in peptide binding domains.[157]

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.[169] Second, distal positions in protein-protein interaction interfaces are in addition directly coupled to the switch by modulating the efficiency of GTP hydrolysis.[161] Third, we show here that a previously unknown allosteric cluster in the structure core (**Figure 3.12A, C**, red) is coupled to switch regulation by altering the nucleotide preference (**Figure 3.8**). Finally, the toxic/GOF positions also include 4 locations of

posttranslational modifications (PTMs).[171,173,174] 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.

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[159] (**Figure 3.12d**). 19/30 positions with activating mutations in H-Ras are also toxic/GOF positions in Gsp1 (**Figure 3.12E**). Those positions are enriched in the active site (**Figure 3.14**), 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 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 3.14**). 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[175] (**Figure 3.12F**). 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 3.14)**. 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 3.14**). 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 3.12A, C**). We propose a model where this sensitivity of the switch facilitates both its responsiveness to many biological inputs and its output signaling specificity.[161] We identify an altered switch balance as the common mechanism by which toxic/GOF mutations affect the cellular function of Gsp1. This finding suggests that the GTPase switch balance is finely tuned and that the sensitivity of this balance to mutations at many positions 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 have deleterious functional consequences. This finding is consistent with results from kinetic models of ultrasensitivity, where for switches controlled by opposing regulators (**Figure 3.1A**) small changes in the concentration or activity of regulators can result in sharp changes in the fraction of the switch "on" state.[110] Our study provides an important link between allosteric regulation of the switch balance[110] at the molecular level, and the ultrasensitivity of switches[110] and functional consequences for cellular regulation at the systems level.[161] Our residue-level functional map of a GTPase molecular switch and the discovery of new regulatory sites opens avenues to interrogate and target GTPases controlling many essential biological processes including intracellular transport, cell growth, differentiation, and cell survival.

Methods

Deep mutational scanning of Gsp1 in *S. cerevisiae*

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.[176] 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.[176] We used cassette ligation at NotI and AscI restriction sites downstream of Gsp1 gene to introduce an oligonucleotide cassette including an N_{18} random sequence into the pRS413Gsp1libmCherry variants. We used paired-end Illumina sequencing to associate the 18 base barcodes with the encoded Gsp1 variants.

To generate the plasmid swap strain, DBY681, we started with a heterozygous diploid Gsp1 knockout (BY4743 Gsp1::KanMX) ordered from GE. First, we introduced pRS416Gsp1GFP and selected for transformants on synthetic media lacking uracil. Next, we sporulated the diploid transformants 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[177] 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 x 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.

A sample of these "initial" cells were retained for sequencing and the remainder were collected by centrifugation at 5000 x 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.[176] 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 3.1E**) and correspond to C-terminal deletion mutants that are known to encode fully folded proteins with perturbed biochemical function.[169] 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. 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.

Expression levels of Gsp1 variants 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 x 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 # PA 1-5783, ThermoFisher Scientific) and Donkey anti-Rabbit HRP-linked secondary (CAT # NA934V, 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.[177] Transformed cells were recovered in SD-ura media for 6 hours and then 5 uL 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 uL of a 10x dilution series of cells were spotted on LB-amp plates.

Statistical modeling of the distribution of toxic/GOF mutations

A hypergeometric distribution was used to model the null distribution of toxic/GOF mutations partitioning among the 219 residue positions. This approach computes the probability that a certain number of toxic/GOF scores would be at the same position, given the number of toxic/GOF scores in the dataset and 21 possibilities at each position (20 amino acids and STOP). The calculation was performed using the *dhyper* function in the *stats* package of the programming language *R*.

Biochemical and biophysical assays

Protein purifications

Gsp1 variants were expressed from a pET-28 a (+) vector with an N-terminal 6xHis 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.[141] The autoinduction medium consisted of ZY medium (10 g/L tryptone, 5 g/L yeast extract) supplemented with the following stock mixtures: 20xNPS (1M Na_2HPO_4, 1M KH_2PO_4, and 0.5 M $(NH_4)_2SO_4$), 50x 5052 (25% glycerol, 2.5% glucose, and 10% α-lactose monohydrate), 1000x trace metal mixture (50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM each of $MnCl_2$ and $ZnSO_4$, 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[161] using reverse phase high performance liquid chromatography on a C18 column. Protein concentrations were confirmed by measuring at 10-50x 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 by Smith and Rittinger[142]). 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 Note** in Chapter 2. Srm1 was purified as Δ1-27Srm1 and GAP as a full-length protein. ScΔ1-27Srm1 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 ºC. 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 ºC.

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 $MgCl_2$, 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.[161] Gsp1 variants are purified as a Gsp1:GDP complex, as verified previously.[161] 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 30ºC in a Synergy H1 plate reader from BioTek, using Corning 3686 96-well half-area non-binding surface plates. Initial rates v_0 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 $p = v_0^{GTP}/v_0^{GDP}$, 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 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, CAT # 371701, Calbiochem) 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.[161] 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, Thermo Fisher) 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 30ºC 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[161] using the custom-made software DELA.[148] Specifically, each time course was fitted to an integrated Michaelis Menten equation:

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

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,

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

Computational methods

Structural bioinformatics

Protein structures were downloaded from the PDB-REDO databank web server.[178] 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[179] implemented in the PyRosetta interface (version *2020.28+release.8ecab77aa50*) to the Rosetta molecular modeling suite.[180]

Assignments of structural regions (structure core, interface core, and surface) of Gsp1 were previously reported[161] 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)[123] compared to the empirical maximum solvent accessible surface area for each of the 20 amino acids.[124] Annotations of the canonical Ras superfamily GTPase regions were taken from[181] as well as studies of Ran structures.[182–185] The key GEF binding region annotations were taken from Renault et al.[186]

Rosetta ΔΔG calculations

Stability calculations for all 19 possible point mutations were performed using the application *cartesian-ddg*[187,188] 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. ΔΔG 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 ΔΔG 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 ΔΔG values are available,[189,190] as recommended by the authors of the *cartesian-ddg* protocol.[187] 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 shown below. 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 ΔΔG values are available in full at the code repository at https://github.com/cjmathy/Gsp1_DMS_Manuscript.

relax flags:

```
<path/to/Rosetta>/main/source/bin/relax.default.linuxgccrelease \
-s <path/to/pdb_file> \
-out:path:all <path/to/output_dir> \
-database <path/to/Rosetta>/main/database \
-use_input_sc \
-in:file:movemap gsp1.movemap \ # for Gsp1 structures only
-constrain_relax_to_start_coords \
-ignore_unrecognized_res \
-nstruct 20 \
-relax:cartesian \
-relax:coord_constrain_sidechains  \
-relax:min_type lbfgs_armijo_nonmonotone \
-score:weights ref2015_cart \
-relax:script cart2.script
```

gsp1.movemap:

```
RESIDUE * BBCHI
RESIDUE 201 202 NO # Nucleotide and Mg in 3M1I. Use 208 209 for 3GJ0
JUMP * YES
```

cart2.script:

```
switch:cartesian
```

```
repeat 2
ramp_repack_min 0.02  0.01     1.0  50
ramp_repack_min 0.250 0.01     0.5  50
ramp_repack_min 0.550 0.01     0.0 100
ramp_repack_min 1     0.00001  0.0 200
accept_to_best
endrepeat
```

cartesian-ddg flags:

```
<path/to/Rosetta>/main/source/bin/cartesian_ddg.linuxgccrelease \
    -database <path/to/Rosetta>/main/database \
    -s <path/to/relaxed_pdb_file> \
    -out:path:all <path/to/output_dir> \
    -ddg:mut_file <path/to/mut_file> \
    -ddg:output_dir <path/to/output_dir> \
    -ddg:iterations 5 \
    -ddg::cartesian \
    -ddg::dump_pdbs true \
    -ddg::bbnbrs 1 \
    -fa_max_dis 9.0 \
    -score:weights ref2015_cart
```

Example *mut_file*, which specifies the mutation to make (here, F159L, which is residue 150 in the numbering scheme used by Rosetta, which always starts at 1 for the N-terminal residue in a chain). One such *mut_file* is used for each modeled mutation.

```
total 1 # specifies only one mutation is being made
1       # specifies only one mutation is being made
```

F 150 L

Comparison to H-Ras mutagenesis data

Alignment of sequence positions between Gsp1 and H-Ras was performed with the *bio3d* package[125] 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 3.13**. In total, 156 structurally aligned positions were included in the analysis. Fitness scores from the human H-Ras mutagenesis study[159] 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,[159] 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 (**Figure 3.12E**, **Figure 3.14**), 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 ($P = 7.711 \text{ e-}5$) 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[191] 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.[175]

Figures

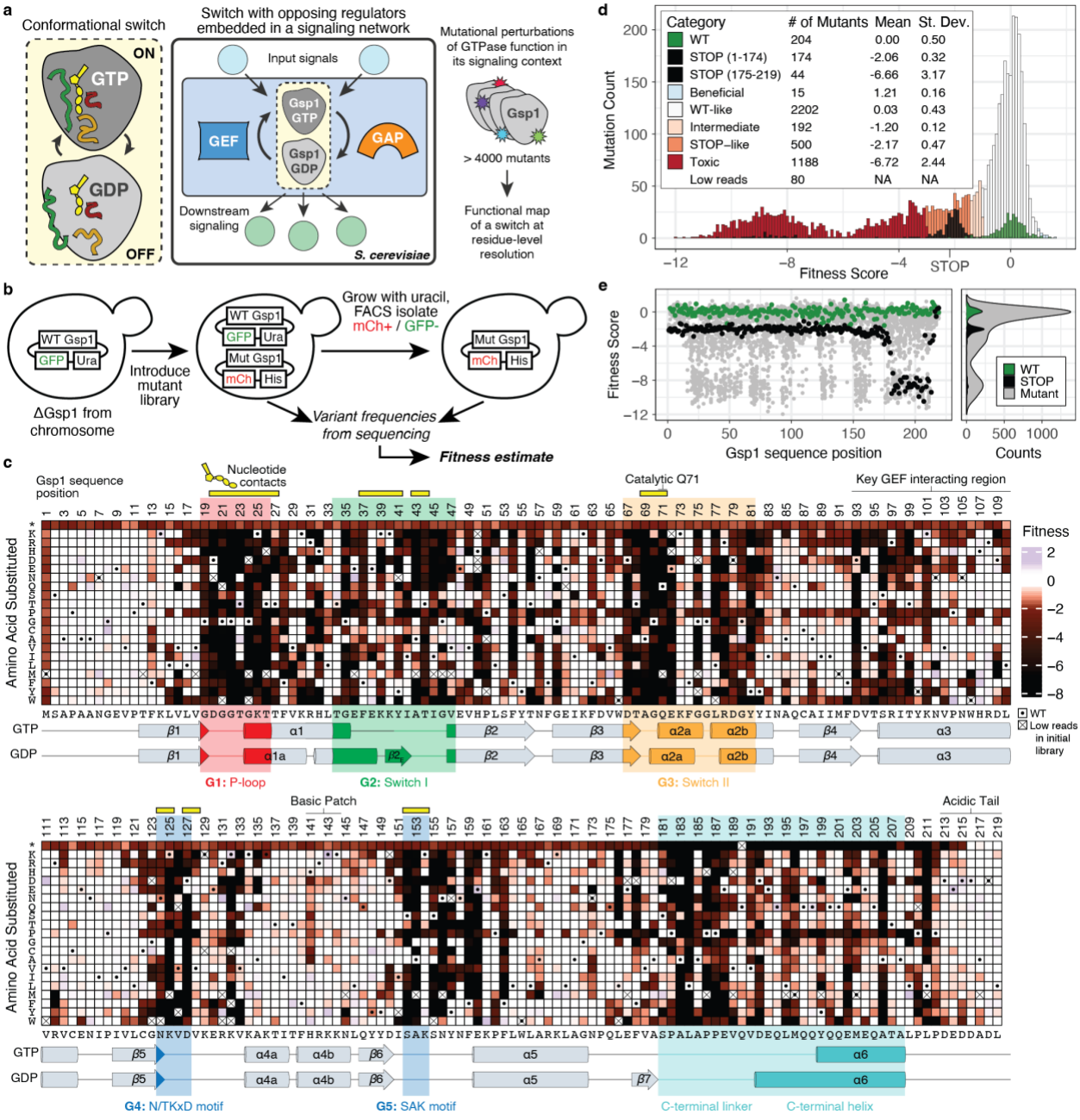


Figure 3.1 *In vivo* sensitivity of the GTPase Gsp1 to all possible single amino acid substitutions.
(a) Mutational perturbations exhaustively probe a switch in its native network. **(b)** Generalizable plasmid swap approach to probe essential genes by mutational mapping. **(c)** Heatmap showing quantitative fitness scores for all Gsp1 mutations after 6 generations of competitive growth. 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 in **Figure 3.3**. 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.[172] Positions 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).

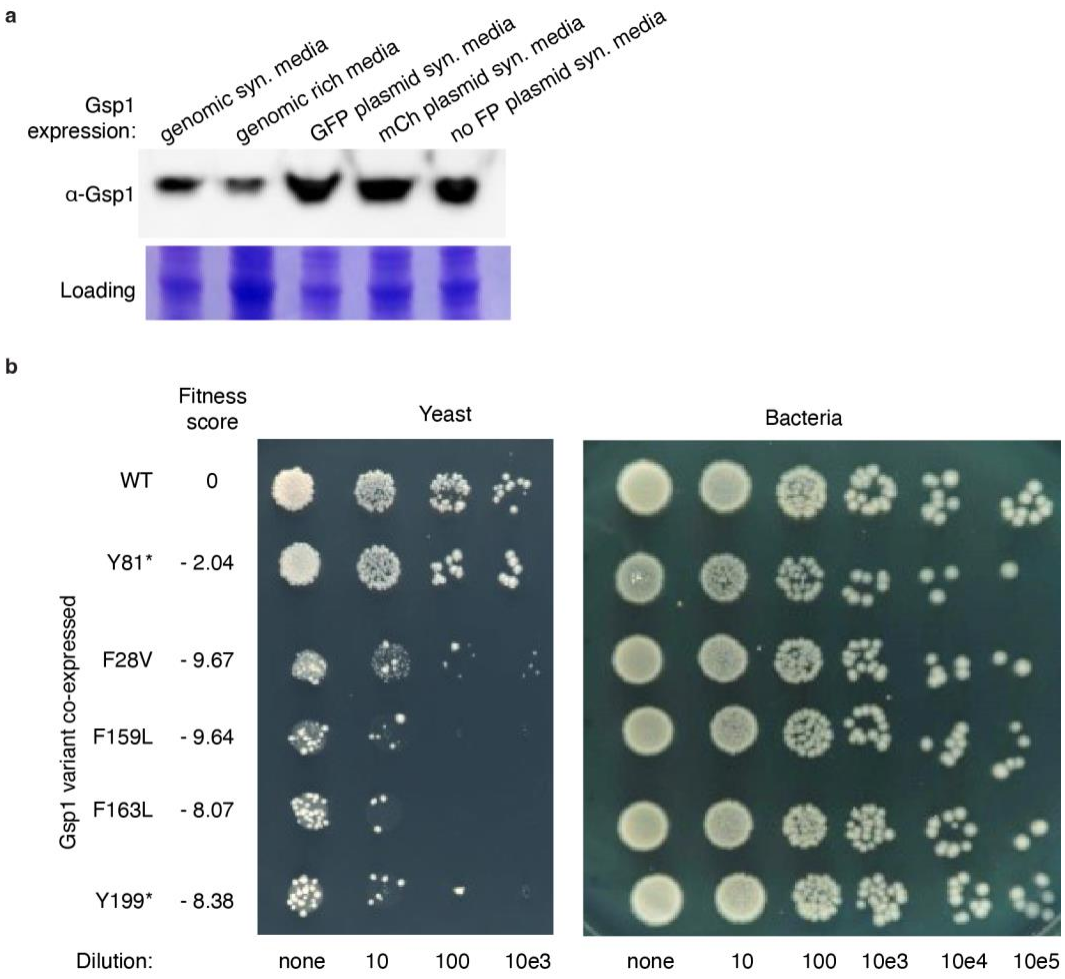


Figure 3.2 Verification of EMPIRIC plasmid expression and growth phenotypes of individual Gsp1 variants.

(a) Western blot comparing the expression of genomic Gsp1 (lanes 1 and 2) to that of Gsp1 expressed from an EMPIRIC plasmid driven by its native promoter (lanes 3-5) in the presence of a GFP marker (lane 3), mCherry marker (lane 4) or no marker (lane 5). **(b)** Dilution series for individual Gsp1 variants expressed from an EMPIRIC plasmid. *S. cerevisiae* (left) and *E. coli* (right) cells were transformed with equal amounts of plasmid and subsequently spotted onto selective plates in a dilution series. Series are shown for WT, one internal STOP codon mutant (Y81*), three toxic/GOF substitution variants (F28V, F159L, F163L), and one C-terminal extension STOP codon mutant (Y199*). Corresponding fitness scores from the EMPIRIC assay are provided. Bacterial dilutions show that the EMPIRIC plasmid itself is not generally toxic.

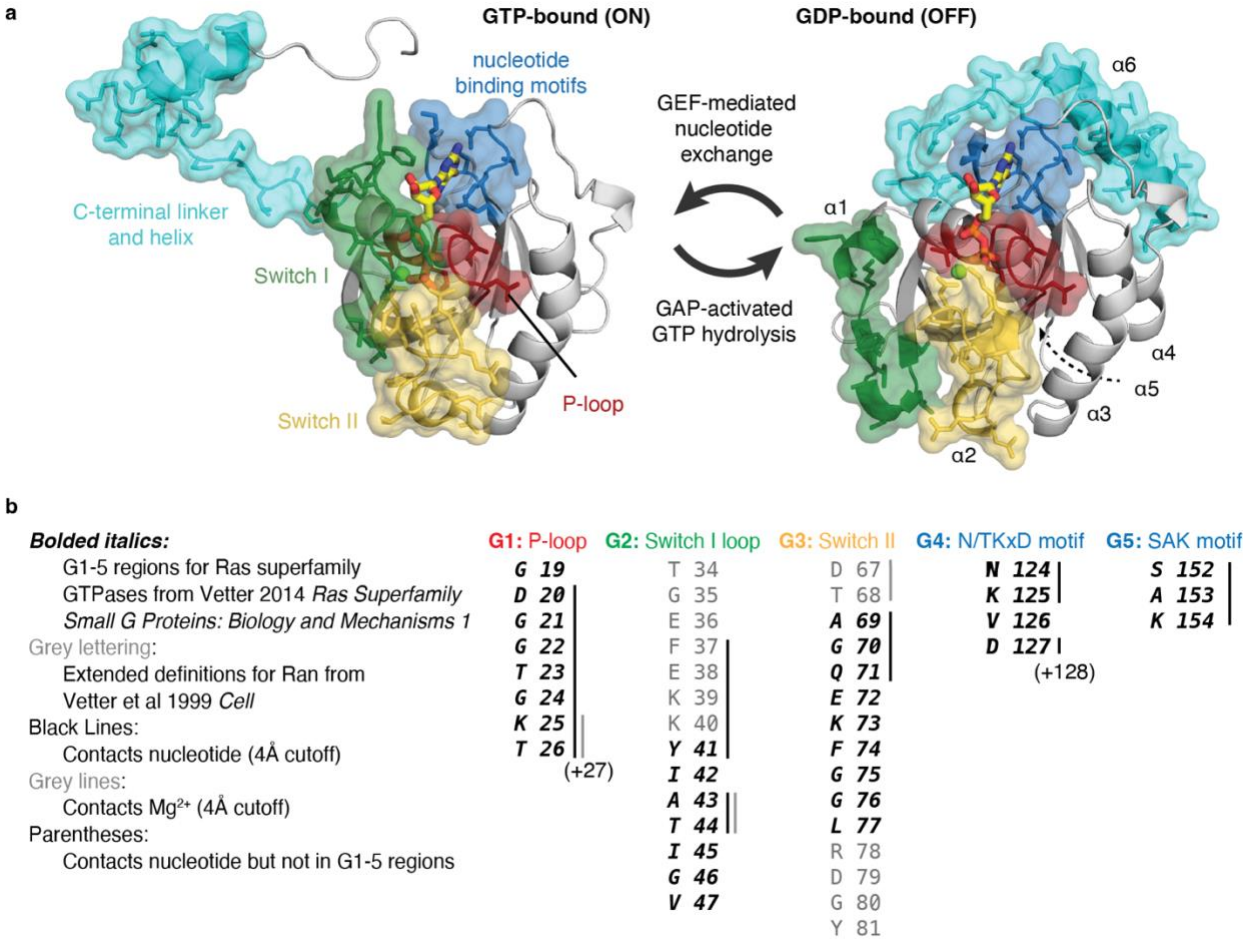


Figure 3.3 Structural annotation of Ran/Gsp1 GTPase regions.
(a) Structures of GTP-bound *S. cerevisiae* Gsp1 (PDB ID: 3M1I co-complex partners Yrb1 and Xpo1 not shown) and GDP-bound *H. sapiens* Ran (PDB ID: 3GJO). G1-5 regions corresponding to Fig. 1b are shown in surface representation: P-loop (red), Switch I loop (green), Switch II loop (yellow), nucleotide binding motifs (blue), C-terminal extension (cyan). Nucleotides are shown in yellow stick representation, and the Mg^{2+} cofactors are shown as green spheres. Large conformational changes associated with state-switching occur in the Switch I and II loops as well as the C-terminal extension. **(b)** Residues comprising the G regions, highlighting the distinction between canonical definitions derived from evolutionary conservation analysis of all Ras superfamily GTPases[181] and the Ran/Gsp1 specific definitions derived from structural characterization.[182] All sequence numbers shown correspond to *S. cerevisiae* Gsp1.

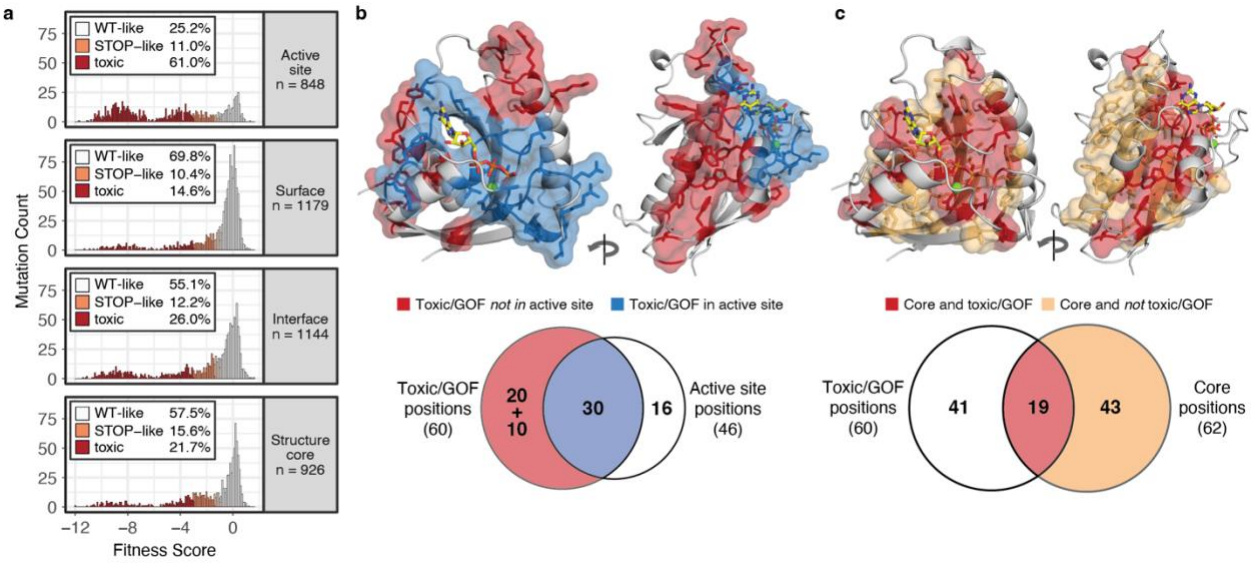


Figure 3.4 Locations of toxic/GOF positions outside of the active site.
(a) Histograms of fitness scores of mutants by structural regions; colors are as in **Figure 3.1D** (showing only point mutations, excluding changes that are WT synonymous or to/from STOP; intermediate and beneficial mutations make up the difference to 100%). Fractions are computed within each structural region; n indicates number of mutations. **(b, c)** Two views rotated by 180 degrees of the Gsp1-GTP structure (PDB ID: 3M1I) showing 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.

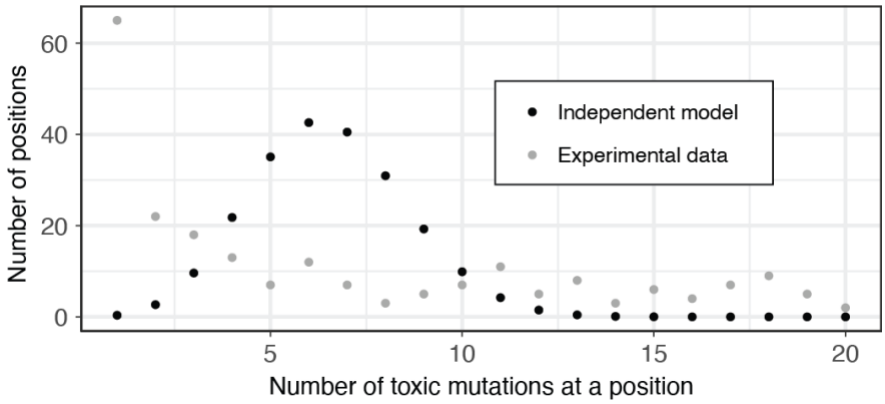


Figure 3.5 Null model of the distribution of toxic/GOF mutations used to define toxic/GOF positions. A null distribution for the number of toxic/GOF mutations observed at each of the possible 219 positions in Gsp1 was constructed from the hypergeometric distribution (Methods) and compared to the observed number of toxic/GOF mutations at each position. We chose a threshold of 10 or more toxic/GOF mutations at a position to define a position as toxic/GOF.

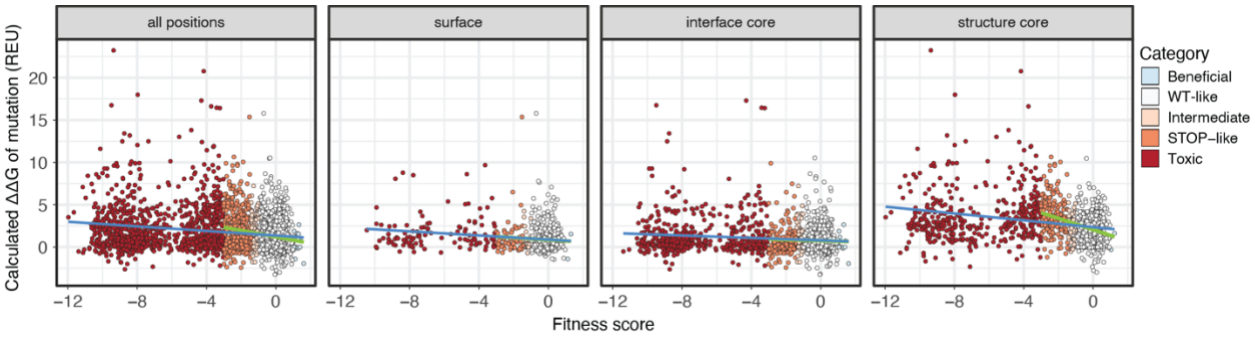


Figure 3.6 Prediction of effects of mutation on protein stability (ΔΔG) using Rosetta.
Scatterplot comparing the EMPIRIC fitness score with the calculated change in protein stability upon mutation (ΔΔG in Rosetta Energy Units, REU) predicted using the GTP-bound structure (PDB ID: 3M1I, residues 10-183). Points are colored by score category as in **Figure 3.1**. Scatterplots broken down by structural region are also shown. Lines indicate best fit when including (blue) or excluding (green) the toxic/GOF mutations.

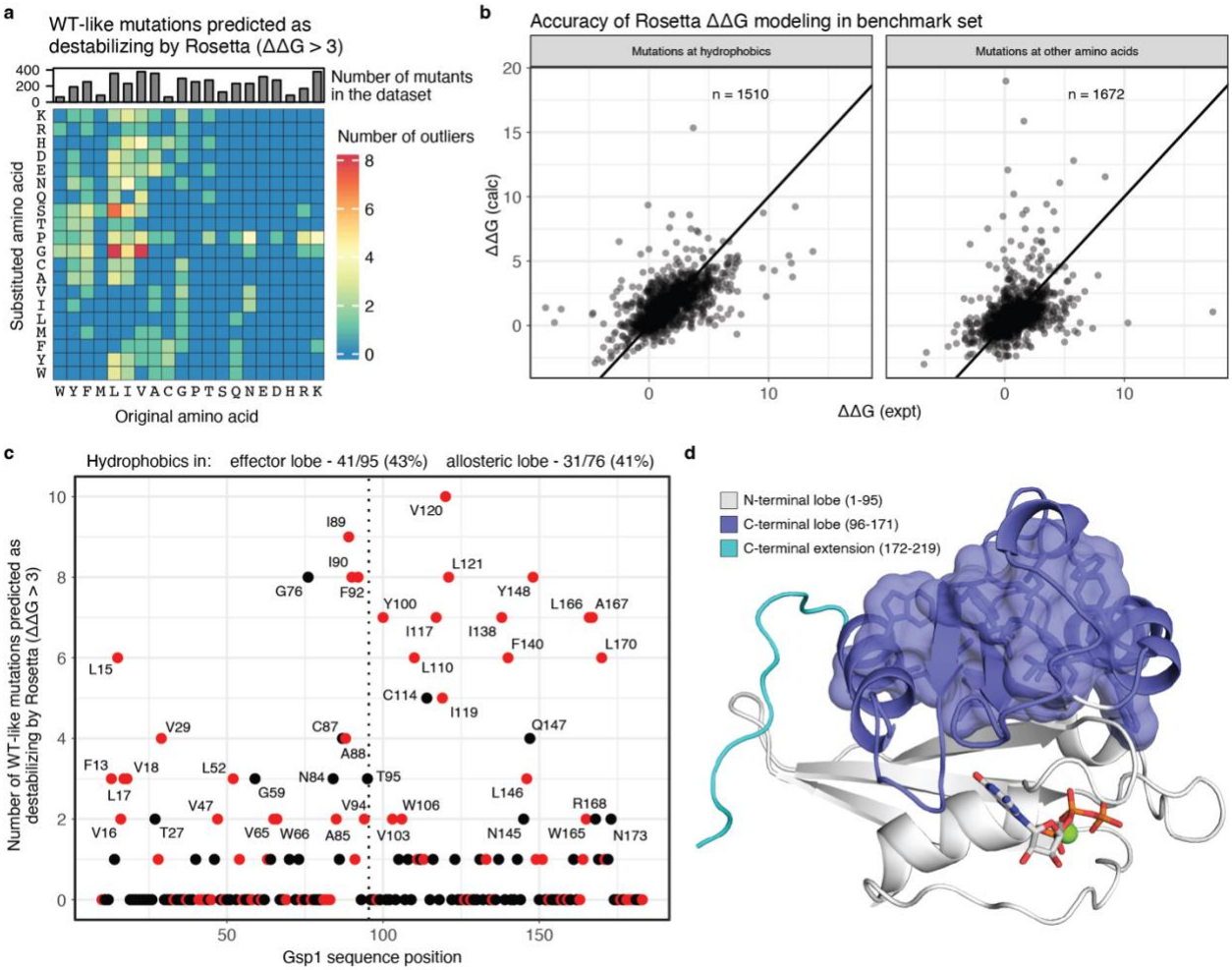


Figure 3.7 WT-like mutations predicted as destabilizing are predominantly at hydrophobic positions in the C-terminal lobe.

(a) Heatmap showing the number of mutations with WT-like fitness scores that are predicted as destabilizing (ΔΔG > 3 Rosetta Energy Units, REU) for Gsp1-GTP, broken down by original and substituted amino acid. **(b)** Experimentally measured and computed ΔΔG values in a large benchmark dataset separated by mutations at positions with a hydrophobic WT amino acid residue (W, Y, F, M, L, I, V, A), vs. those at all other residues. **(c)** Number of mis-predicted WT-like mutations as in panel (a), ordered by Gsp1 sequence position. Positions with two or more mutations mis-predicted are labeled. Mutations at hydrophobic positions shown in red. A large set of mutations at core hydrophobic residues in the C-terminal lobe of Gsp1 are predicted as destabilizing but have WT-like fitness scores, a discrepancy that cannot be explained by poor method accuracy of Rosetta given high performance on similar mutations in a large benchmark dataset as seen in panel (c). **(d)** Structural representation of Gsp1-GTP (PDB ID: 3M1I) showing the N-terminal lobe (white), C-terminal lobe (purple), and a stretch of C-terminal extension residues (cyan). Hydrophobic sidechains comprising the C- terminal lobe core (red dots after sequence position 87 in panel (d)) are shown in sticks and surface representation. Nucleotide shown in white sticks and Mg^{2+} cofactor shown as a green sphere.

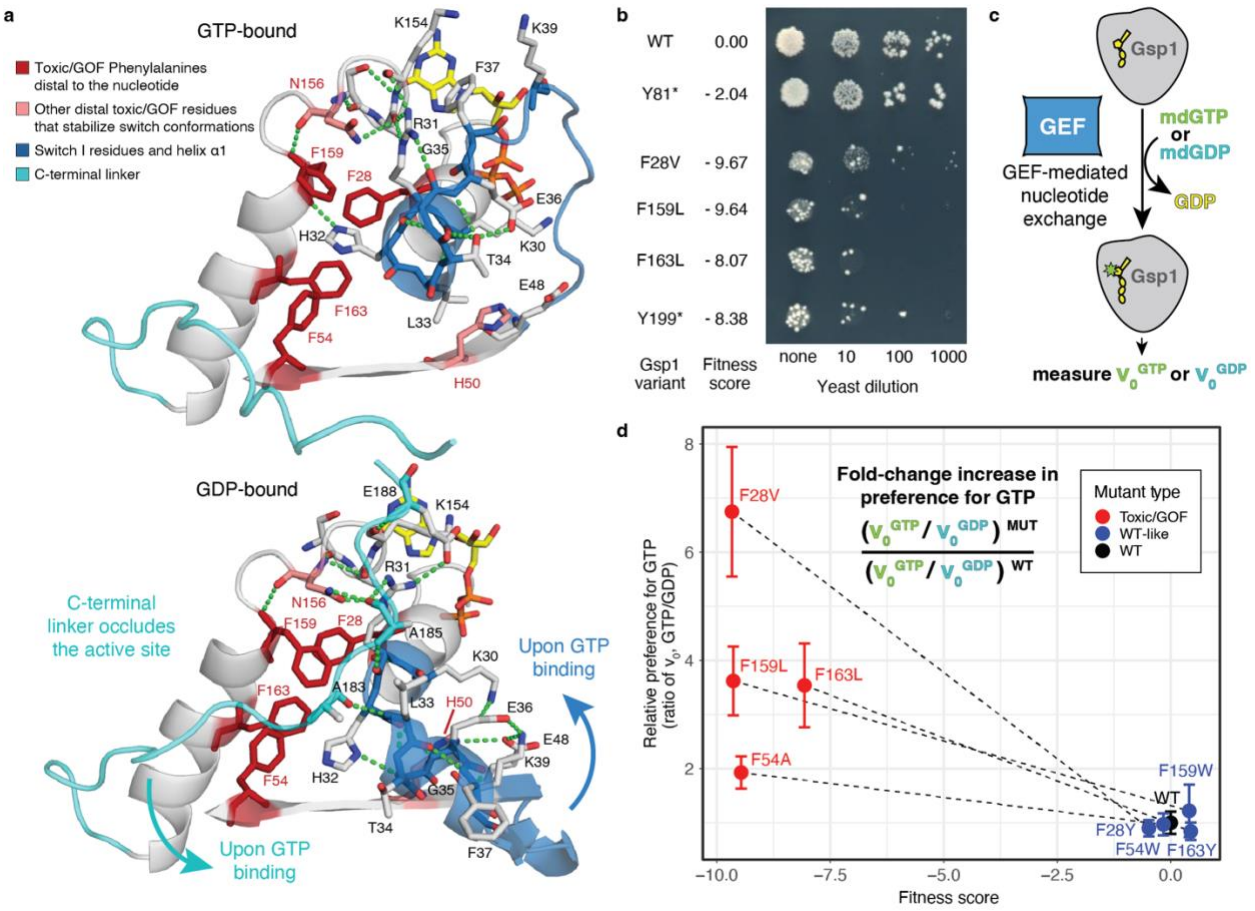


Figure 3.8 Distal toxic/GOF mutations allosterically alter the balance of the switch states.
(a) Structural depiction of extended networks of interactions in the GTP-bound (top, PDB ID: 3M1I) and GDP-bound states (bottom, PDB ID: 3GJ0). Toxic/GOF mutants characterized in panels (b) and (d) shown in red. 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 positions 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. Error bars represent the standard deviations of v_0 measurements propagated across the division operator (Methods).

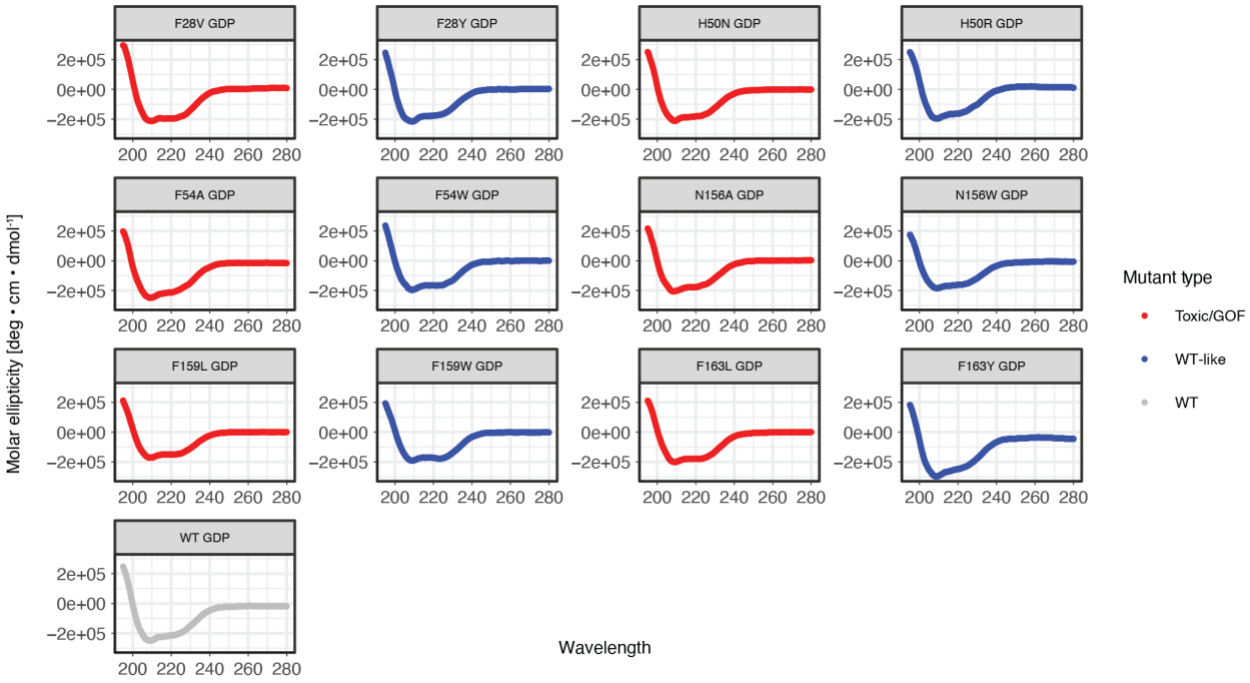


Figure 3.9 Circular dichroism (CD) spectra of purified Gsp1 variants.
CD spectra for Gsp1 variants at 25ºC. Variants with a toxic/GOF mutation are shown in red, variants with a WT-like mutation in blue, and WT in gray.

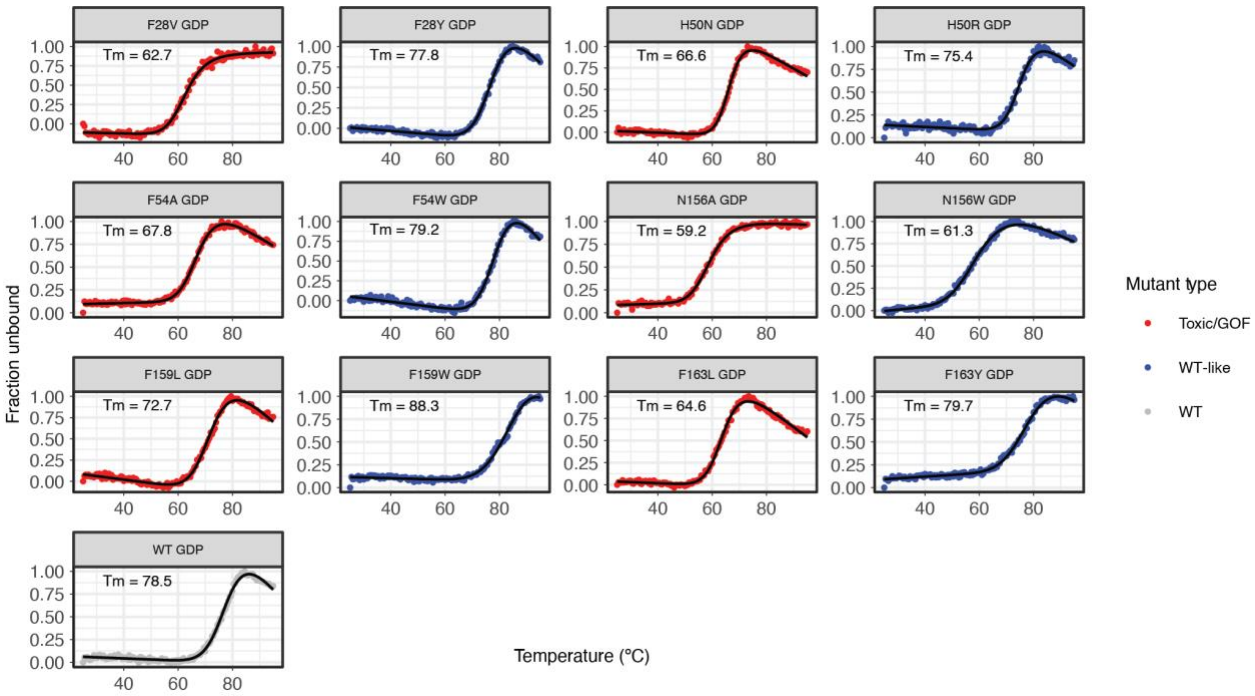


Figure 3.10 Circular dichroism (CD) thermal melts of purified Gsp1 variants.
CD melts of Gsp1 variants from 25 - 95°C. Variants with a toxic/GOF mutation are shown in red, variants with a WT-like mutation in blue, and WT in gray. Apparent melting temperature (T_m^{app}) values were computed by fitting melts to a two-state unfolding equation (see Methods). CD melts are not reversible. All variants are stable up to at least 50°C, although toxic/GOF mutations resulted in slightly decreased apparent melting temperatures (T_m^{app} 62.7–72.7°C) compared to WT (T_m^{app} 78.5°C) or WT-like mutations (T_m^{app} 77.8–88.3°C) at the same positions.

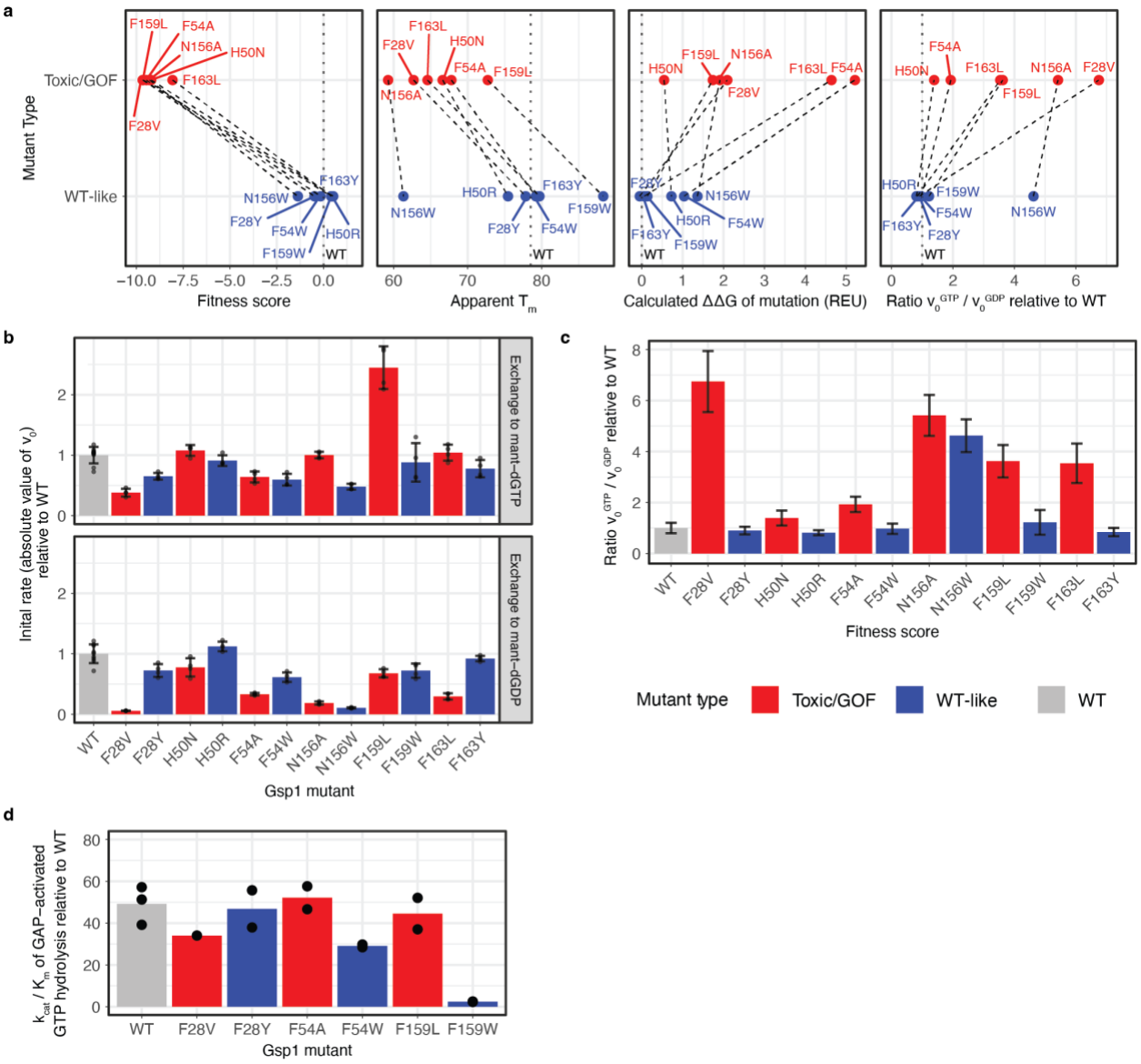


Figure 3.11 Biophysical properties of toxic/GOF and WT-like mutations.
(a) Comparison of toxic/GOF vs WT-like mutations for various parameters (from left to right): EMPIRIC fitness score, apparent T_m as measured by irreversible CD melts, calculated ΔΔG for the GDP-bound structure (*H. sapiens* Ran-GDP, PDB ID: 3GJ0) using Rosetta, and the change in preference for GTP over GDP relative to WT as measured using the GEF-mediated nucleotide exchange assay. Variants colored by category: Toxic/GOF (red) and WT-like (blue). WT values shown as black dotted line. Colored dotted lines connect mutants at the same sequence position. **(b)** Initial rates of GEF-mediated nucleotide exchange, normalized to the WT values. Measurements performed with n >= 4 replicates. Error bars represent the standard deviations of v_0 measurements propagated across the division operator (Methods). **(c)** Relative change in nucleotide preference, 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. Error bars represent the standard deviations of v_0 measurements propagated across the division operator. **(d)** Catalytic efficiency (k_{cat}/K_m) of GAP-activated GTP hydrolysis, normalized to the WT value. Individual replicates are shown as points on each bar.

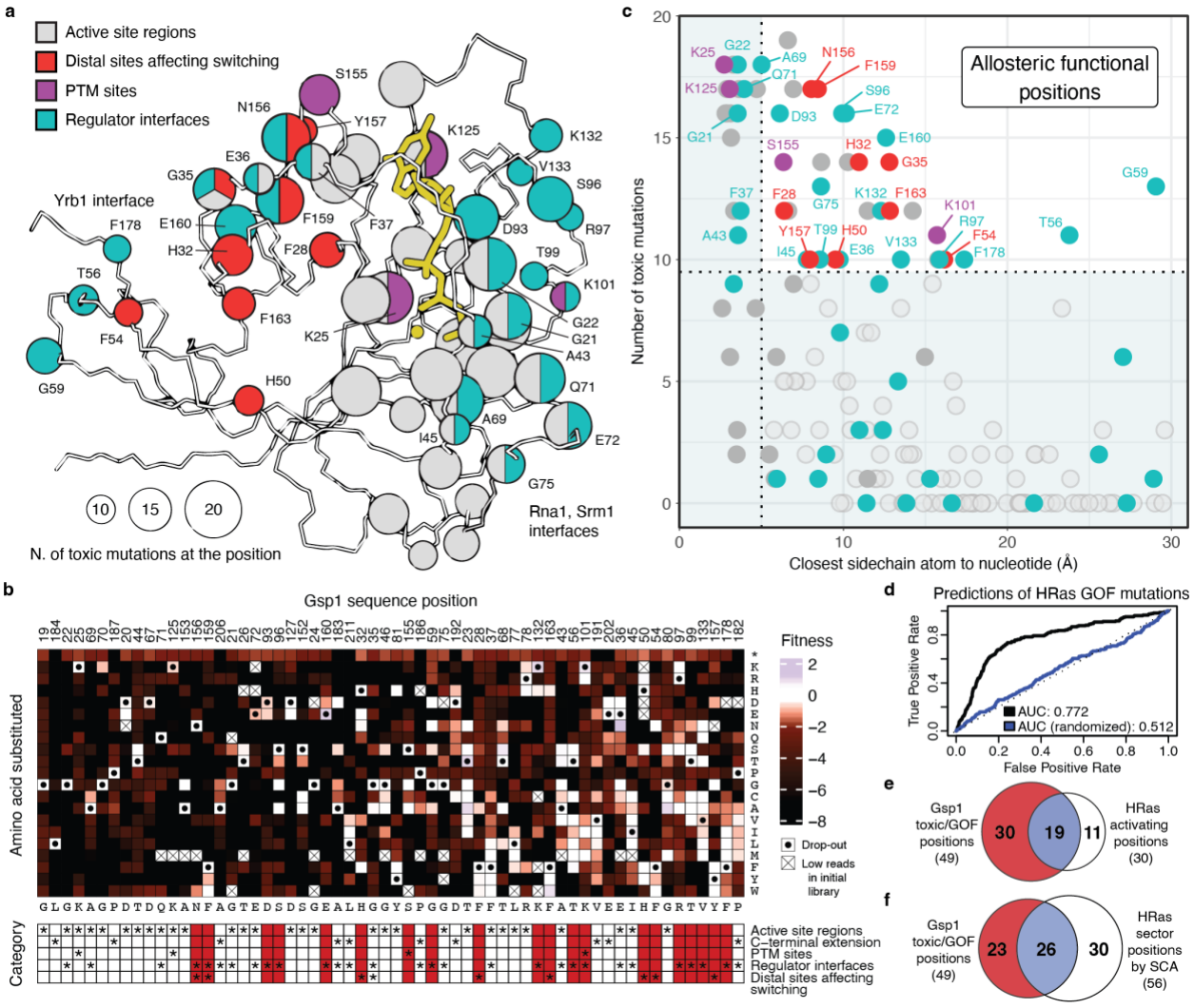


Figure 3.12 Allosteric map of the Gsp1 GTPase switch.
(a) Wire representation of Gsp1-GTP (PDB ID: 3M1I, residues 1-180). Toxic/GOF positions are shown in sphere representation. Sphere radius represents number of toxic/GOF mutations at each position. Spheres are colored by functional categories, see (b). The nucleotide and Mg^{2+} cofactor are shown in yellow. **(b)** Heatmap showing fitness scores 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 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* HRas mutant as activating, as defined by Hidalgo et al.[159] Datasets were trimmed to the 156 sequence positions alignable for Gsp1 and HRas (**Figure 3.13**). **(e, f)**, Overlap of functional sites defined as Gsp1 toxic/GOF and either **(e)** HRas activating or **(f)** comprising an HRas sector defined by statistical coupling analysis (SCA)[175] (**Table 3.1**).

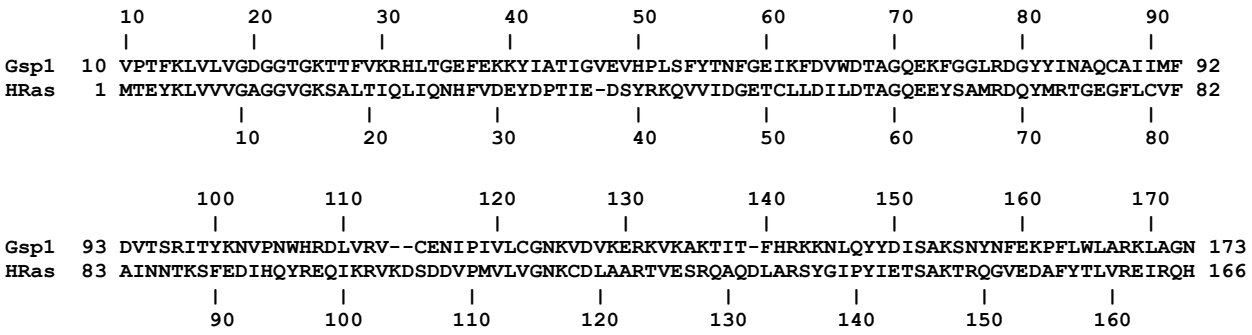


Figure 3.13 Sequence alignment of Gsp1 – H-Ras based on structural alignment.

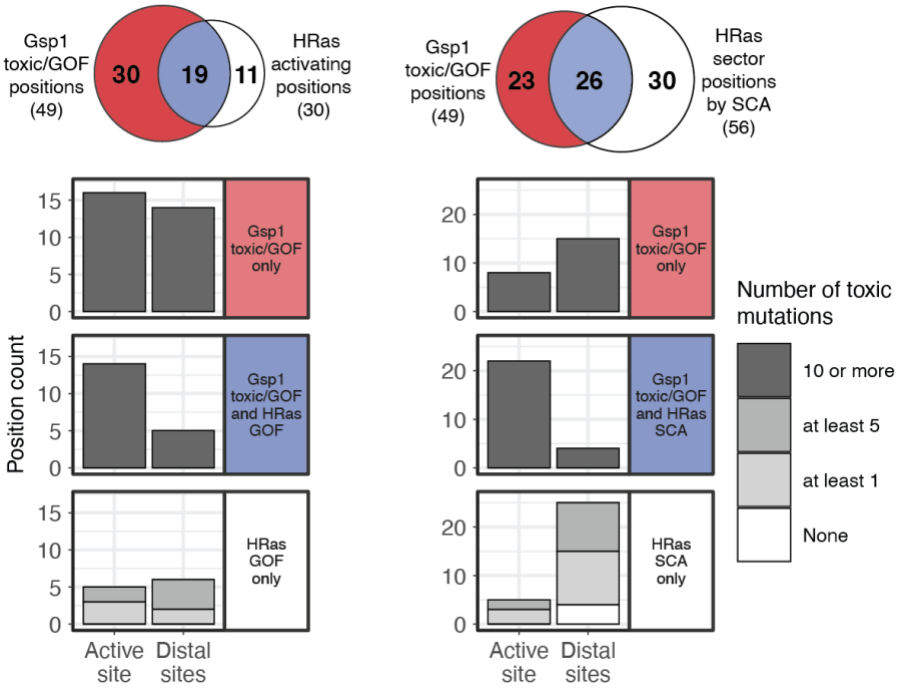


Figure 3.14 Locations of functional positions identified by the Gsp1 generalized EMPIRIC assay, HRas activation assay, or HRas statistical coupling analysis.
Active site defined as in **Figure 3.1**, including the canonical G1-5 regions conserved across Ras-superfamily GTPases, as well as residues included in the expanded definitions for Ran/Gsp1 based on structural analysis (**Figure 3.3**, Methods). Venn diagrams are as in **Figure 3.12E, F** and repeated here for comparison. Bar graphs underneath indicate number of positions in each of the categories (red, blue, and white) from the Venn diagram. Bars are shaded by number of toxic/GOF mutations in the Gsp1 generalized EMPIRIC assay.

Table

Table 3.1 Overlap of positions annotated as toxic/GOF in Gsp1, activating in H-Ras, or part of an H-Ras sector by SCA. Definitions for categories are from Gsp1 analysis, shown in Fig. 4. Sets of H-Ras sites for each category were constructed based on a structural alignment, and the subset of sites labeled as activating[159] or as part of the SCA sector[175] are listed.

Gsp1 residue category (Figure 3.12)	Toxic/GOF positions in Gsp1 (10 or more toxic mutations)	Activating positions in H-Ras (>=2 activating mutations)	Sector positions in H-Ras (from SCA analysis)
Active site regions	G19, D20, G21, G22, T23, G24, K25, T26, G35, E36, F37, A43, T44, I45, G46, D67, T68, A69, G70, Q71, E72, G75, L77, R78, G80, Y81, K125, D127, S152, A153	G12, G13, V14, K16, A18, P34, T58, A59, G60, Q61, E63, R68, N116, K117, D119, L120, S145, A146, K147	G10, A11, V14, G15, K16, S17, F28, Y32, P34, T35, I36, D57, T58, A59, G60, Q61, E62, E63, Y64, R68, Y71, N116, K117, D119, S145, A146, K147
Distal sites affecting switching	F28, H32, G35, H50, F54, N156, Y157, F159, F163	L19, Q22, L23, V152, F156	Q22, L23, A134, F156
PTM sites in Gsp1	K25, K101, K125, S155	K16, K117	K16, K117
Regulator Interface	G21, G22, G35, E36, F37, A43, I45, T56, G59, A69, Q71, E72, G75, D93, S96, R97, T99, K101, K132, V133, N156, F159, E160	G12, G13, P34, A59, Q61, E63, A83, V152	F28, Y32, P34, I36, A59, Q61, E62, E63, Y64, A83, Q99, R123, V125

Chapter 4. Conclusion

Switching constrains the genotype-phenotype map of Gsp1

The concept of a genotype-phenotype map is vital to biology. Derived from the mathematical concept of a "mapping" between two sets, it firstly captures the linkage between hereditary information (genotype) and an organization of molecules that maintains homeostasis in a dynamic natural environment (phenotype, or the behavior of a living system). But a genotype-phenotype map also specifies that this linkage is well-defined, i.e. we can discover the rules that uniquely determine how a phenotype arises from a given genotype, and under similar environmental constraints we can reproduce a phenotype by deploying the corresponding genotype.

This *reproducibility* of genotype-phenotype maps is vital to bioengineering. While natural living systems are optimized through Darwinian survival of the fittest, where selection occurs on phenotypes sampled through random mutation of genotypes, bioengineers can much more efficiently design optimal living systems by first "charting" and then "navigating" these genotype-phenotype maps (to use a corny but suitable *double-entendre* of the word "map"). Charting the underlying mechanics of living systems through experimentation and model building can guide the design of effective treatments and of interesting and useful biotechnologies for non-medical applications, such as in agriculture, bioremediation, or even the arts.

Unfortunately, genotype-phenotype maps are notoriously hard to chart. This is for many valid reasons: many mechanisms occur simultaneously in a single cell, and our inventory of molecules is incomplete; very precise tools are often required to make repeated, non-destructive measurements of living systems; the space of possible genotypes is massive; and physical systems often behave in unique ways at certain scales, meaning we have to be careful to not base our models

too heavily on intuition from other fields – for instance noting that a torpedo pool toy will move very differently than a similarly shaped bacterium in fluids at low Reynold's number.[192]

The particular characteristic complicating genotype-phenotype maps that motivated this dissertation is the possibility for complex propagation of effects across multiple scales (**Figure 4.1**). There is a distinct separation in scale between genotypes encoded at the level of molecules and phenotypes which undergoes selection at the level of cells or even entire tissues. Our understanding of how sequence changes alter the structure of proteins has indeed taken a huge leap recently due to the emergence of methods for protein structure prediction based on deep learning and evolutionary conservation.[193] However, it remains challenging to predict how these structural changes translate to altered protein dynamics or biochemical functions like binding and catalysis, how changes in biochemical function propagate through a molecular interaction network, and how the state of a molecular network determines cellular fitness in any number of possible environments.

The studies presented in this dissertation sought to identify if, in the Gsp1 network, this complex propagation was dominated by a single mechanism. By making measurements of the effects of mutations at each scale – measuring structure and dynamics with ^{31}P NMR, biochemical function with *in vitro* enzymology, network rewiring with AP-MS, and cellular phenotype with the E-MAP and EMPIRIC assays – we identified the kinetic rates of GTPase switching as the key parameters constraining the genotype-phenotype map of Gsp1.

Whether this genotype-phenotype map is of use for bioengineers will ultimately be determined by future efforts to understand, modulate, and design the function of GTPases and other protein switches. However, several existing studies suggest that the importance of regulated tuning of switch kinetics is not only broadly important for other biological switches, but is also engineerable.

First, misregulation of the balance between switch states is a major mechanism of oncogenesis driven by Ras mutations,[194] and the ratio of expression levels of Ras GAPs and Ras GEFs is perturbed in cancerous tissues despite being consistent across healthy tissues.[120] This suggests that targeted modulation of the rates of GTPase activation and deactivation may be a valid therapeutic strategy, and indeed at least one inhibitor targeting the Ras GEF SOS has advanced to clinical trials.[195] Second, the ability for molecular interactions of certain kinetic rates to selectively occur over other interactions, known as kinetic proofreading,[196] is important for reducing error in many cellular processes, including tRNA charging[197], activation of SOS,[198] and antigen discrimination by T cell receptors.[199] Proofreading has been successfully engineered by tuning dwell times of ligand-receptor complexes via affinity engineering.[200] At the systems level, different kinetics of activation or dwelling can also result in selective activation of downstream pathways in a mechanism known as kinetic filtering.[107] The sensitivity of downstream processes to altered kinetics of Gsp1 switching adds to this large set of kinetic mechanisms that underly complex biology and thus might be of interest for porting to new engineered systems.

An open question for future efforts in engineering pathway responses to differential switch kinetics will be whether the phenomenon requires some or all of the characteristic features of the Ran/Gsp1 system, such as the spatial gradient of activity due to separation of the GEF and GAP into different compartments, the large set of effector proteins with varying affinities,[38] the lack of functional homologs whose presence might introduce pathway crosstalk, or the thermodynamic coupling between effector interfaces and the molecular switch. If these features are dispensable or readily engineerable, it may be possible to introduce similar multifunctionality to that seen in Gsp1 into other two-state molecular switches with opposing regulators, such proteins whose regulatory function is dependent on one or more PTMs.

Protein engineering of systems level function – could we design a protein like Gsp1?

Protein engineering aims to identify protein sequences with a desired function, through the purposeful introduction of mutations into existing proteins by randomized[201] or guided processes (such as through the use of structural[202], functional[203], or evolutionary data[204,205]). Key to this idea is the view of a genotype-phenotype map in terms of a *protein fitness landscape*: for a protein of sequence length N, we define a fitness function $f(N)$ which represents how evolutionarily favored that sequence is in a certain set of environmental conditions.[206] All protein engineering methods go through iterations of sampling different sequences, scoring them based on a measurable proxy of fitness, and selecting one or more sequences that have desirable traits (which in addition to intended function may also include biophysical properties such as stability or reduced immunogenicity). An especially fit sequence is said to lie at a fitness peak, and the ruggedness of the fitness landscape refers to how likely neighboring sequences are to be similarly fit.

Fitness landscapes first served as a useful theoretical framework for understanding natural evolution and developing protein engineering strategies,[206] but in recent years machine learning-based methods that explicitly model fitness landscapes have enabled new levels of efficiency in identifying fit variants.[207,208] For instance, an approach known as UniRep[209] first trained an unsupervised recurrent neural network on raw protein sequences, learning how to distinguish protein-like from non-protein-like sequences, and then re-weighted the model using green fluorescent protein (GFP) sequences to learn the fitness landscape of that specific family. UniRep was highly accurate in predicting mutations that reduced fluorescence,[209] and was able to identify new better-than-WT variants for both GFP and the enzyme TEM-1 β-lactamase after re-weighting the model with sequences from only 24 initially characterized variants for each task.[210] This low ratio of sequences initially characterized to those identified as more fit represents an enormous

increase in efficiency over directed evolution approaches, given that they often rely on screening libraries of 10^6-10^9 (or greater) variants,[211] and that the selection of optimal variants by UniRep is fast once the models have been re-trained on the experimental data. While computational protein design methods similarly offer improved efficiency by bypassing any experimental step for selecting fit variants, validation experiments have found that the percent of stable, functional designs from large scale studies is still relatively low.[212]

Given that Ran/Gsp1 is an essential and conserved piece of eukaryotic machinery, a natural question for protein engineers is: could we design a protein like Gsp1? While many of the key mechanisms of the Ran/Gsp1 system can currently be engineered, such as conformational switching,[21] allosteric regulation,[23,24] protein-protein interactions of high affinity and specificity,[213,214] and precise protein interaction network connectivity,[215] the systems level property of multifunctionality – the selective, simultaneous, and independent tuning of downstream processes – poses a design task that is difficult to achieve using current methods. Whether via directed evolution, computational design, or machine learning-guided engineering, most protein engineering methods rely on single parameter objective functions to serve as a proxy for fitness. For instance, surface display screens typically select for variants with optimal binding affinity to only one partner, perhaps in the presence of a competitor. On the other hand, growth assays of protein variant libraries in their native context do account for the impact of new sequences on all functions, but which of the multiple functions was improved or diminished for a given variant is indistinguishable from the growth rate alone. This was evidenced in the deep mutational scanning study presented in Chapter 3, as toxic/gain-of-function mutations of Gsp1 at many different sites and with different molecular mechanisms of perturbed switching had similar fitness scores.

For a multifunctional protein like Gsp1, an alternative approach might be to monitor the effects of each variant on several functions simultaneously using systems level phenotypic data, and then applying computational tools to integrate the data for many variants and model a fitness landscape. The potential of this approach is apparent in the results of the E-MAP study presented in Chapter 2: the functional genetics data was sufficient to classify mutations based on their effect on all individual cellular functions of Gsp1. However, that study was limited to only 55 mutations, due to the intensive process of yeast strain engineering and the medium-throughput array-based format of genetic interaction profiling. Fortunately, advanced high-throughput methods have been recently developed for engineering mutations[216,217] and performing functional genomics with transcriptomic readouts[218–221]. Furthermore, a pioneering enzyme engineering campaign has shown how systems level measurements might guide protein engineering: monitoring the *S. cerevisiae* metabolome using microfluidic sorting and matrix-assisted laser desorption ionization (MALDI)-MS enabled the discrimination of enzyme variants that preferentially and to varying degrees metabolize a substrate to one product instead of an alternative product.[222] Developing general computational methods for constructing fitness landscapes from high-dimensional systems level data will also enable other forms of phenotype data to be used in protein engineering, such as high content imaging data,[223–225] or genome-wide measurements of global protein complex formation.[226–228]

Summary

Together, the systems biology studies of the *S. cerevisiae* GTPase Gsp1 presented in this dissertation culminated in a new model for how switch cycle kinetics enable the independent regulation of distinct downstream processes, as well as the identification of new allosteric

regulatory sites in GTPases. Future work will ask to what extent these mechanisms generalize to human Ran, to other eukaryotic small GTPases, and to protein switches in general, and whether the integrated experimental and computational approaches for uncovering new biochemical phenomena presented here can drive new biological inquiry and the productive engineering of living systems.

Figures

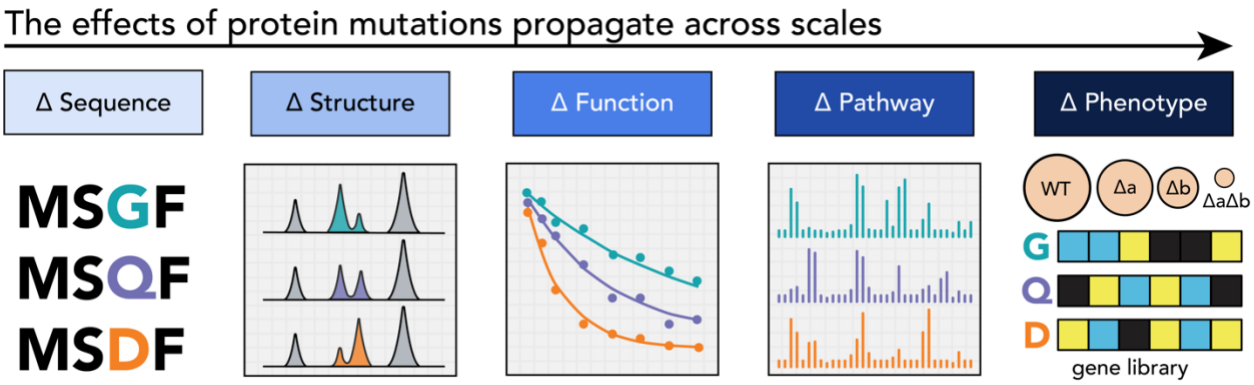


Figure 4.1 Complex propagation of mutational effects to cellular phenotypes.
The effects of mutations in proteins (Δ Sequence) can exert their effect at multiple scales, and the propagation from each scale to the next adds complexity to the genotype-phenotype map of a protein. Cartoon representations are shown for: how mutations can induce differential population of structural states (Δ Structure); how populating different structural states can directly alter biochemical function, such as the kinetic rates of enzymes (Δ Function); how changes in structural states and biochemical functions can rewire molecular interaction networks, as seen in different protein complex abundances identified by affinity-purification mass spectrometry experiments (Δ Pathway); and how the many pathways acting together in a cell ultimate result in a measurable phenotype, shown as yeast colony sizes and genetic interaction profiles (Δ Phenotype).

References

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
2. Tunnacliffe, E. & Chubb, J. R. What is a transcriptional burst? *Trends Genet.* **36**, 288–297 (2020).
3. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of transcriptional bursting in bacteria. *Cell* **158**, 314–326 (2014).
4. Antolović, V., Miermont, A., Corrigan, A. M. & Chubb, J. R. Generation of Single-Cell Transcript Variability by Repression. *Curr. Biol.* **27**, 1811-1817.e3 (2017).
5. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
6. Kitano, H. Computational systems biology. *Nature* **420**, 206–210 (2002).
7. Strange, K. The end of "naïve reductionism": rise of systems biology or renaissance of physiology? *American Journal of Physiology-Cell Physiology* vol. 288 C968–C974 (2005).
8. Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).
9. Kholodenko, B. N. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.* **267**, 1583–1588 (2000).
10. Pomerening, J. R., Sontag, E. D. & Ferrell, J. E., Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat. Cell Biol.* **5**, 346–351 (2003).
11. Guidi, G. M. & Goldbeter, A. Bistability without Hysteresis in Chemical Reaction Systems: The Case of Nonconnected Branches of Coexisting Steady States. *The Journal of Physical Chemistry A* vol. 102 7813–7820 (1998).

12. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).

13. Lim, W. A., Lee, C. M. & Tang, C. Design principles of regulatory networks: searching for the molecular algorithms of the cell. *Mol. Cell* **49**, 202–212 (2013).

14. Gordley, R. M., Bugaj, L. J. & Lim, W. A. Modular engineering of cellular signaling proteins and networks. *Curr. Opin. Struct. Biol.* **39**, 106–114 (2016).

15. Shah, N. A. & Sarkar, C. A. Robust network topologies for generating switch-like cellular responses. *PLoS Comput. Biol.* **7**, e1002085 (2011).

16. Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).

17. Wittinghofer, A. *Ras Superfamily Small G Proteins: Biology and Mechanisms 1: General Features, Signaling*. (Springer Vienna, 2014).

18. Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299–1304 (2001).

19. Kholodenko, B. N. Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165–176 (2006).

20. Cherfils, J. & Zeghouf, M. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* **93**, 269–309 (2013).

21. Alberstein, R. G., Guo, A. B. & Kortemme, T. Design principles of protein switches. *Curr. Opin. Struct. Biol.* **72**, 71–78 (2022).

22. Glasgow, A. A. *et al.* Computational design of a modular protein sense-response system. *Science* **366**, 1024–1028 (2019).

23. Quijano-Rubio, A. *et al.* De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).

24. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301**, 1904–1908 (2003).

25. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* **25**, 660–662 (2007).

26. Stewart, M. Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell Biol.* **8**, 195–208 (2007).

27. Macara, I. G. Transport into and out of the Nucleus. *Microbiol. Mol. Biol. Rev.* **65**, 570–594 (2001).

28. Stewart, M. Ran in Nucleocytoplasmic Transport. in *Ras Superfamily Small G Proteins: Biology and Mechanisms 2: Transport* (ed. Wittinghofer, A.) 109–124 (Springer International Publishing, 2014).

29. Ren, M., Coutavas, E., D'Eustachio, P. & Rush, M. G. Effects of mutant Ran/TC4 proteins on cell cycle progression. *Mol. Cell. Biol.* **14**, 4216–4224 (1994).

30. Mirallas, O. *et al.* Intertwined control of the cell cycle and nucleocytoplasmic transport by the cyclin-dependent kinase Pho85 and RanGTPase Gsp1 in Saccharomyces cerevisiae. *Microbiol. Res.* **206**, 168–176 (2018).

31. Kalab, P., Pu, R. T. & Dasso, M. The ran GTPase regulates mitotic spindle assembly. *Curr. Biol.* **9**, 481–484 (1999).

32. Scott, R. J., Cairo, L. V., Van de Vosse, D. W. & Wozniak, R. W. The nuclear export factor Xpo1p targets Mad1p to kinetochores in yeast. *J. Cell Biol.* **184**, 21–29 (2009).

33. Ryan, K. J., McCaffery, J. M. & Wente, S. R. The Ran GTPase cycle is required for yeast nuclear pore complex assembly. *J. Cell Biol.* **160**, 1041–1053 (2003).

34. Hayashi, N. *et al.* Mutations in Ran system affected telomere silencing in Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun.* **363**, 788–794 (2007).

35. Clément, M., Deshaies, F., de Repentigny, L. & Belhumeur, P. The nuclear GTPase Gsp1p can affect proper telomeric function through the Sir4 protein in Saccharomyces cerevisiae. *Mol. Microbiol.* **62**, 453–468 (2006).

36. Fan, S. *et al.* Induction of Ran GTP drives ciliogenesis. *Mol. Biol. Cell* **22**, 4539–4548 (2011).

37. Wing, C. E., Fung, H. Y. J. & Chook, Y. M. Karyopherin-mediated nucleocytoplasmic transport. *Nat. Rev. Mol. Cell Biol.* **23**, 307–328 (2022).

38. Hahn, S. & Schlenstedt, G. Importin β-type nuclear transport receptors have distinct binding affinities for Ran-GTP. *Biochem. Biophys. Res. Commun.* **406**, 383–388 (2011).

39. Beddow, A. L., Richards, S. A., Orem, N. R. & Macara, I. G. The Ran/TC4 GTPase-binding domain: identification by expression cloning and characterization of a conserved sequence motif. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3328–3332 (1995).

40. Mueller, L., Cordes, V. C., Bischoff, F. R. & Ponstingl, H. Human RanBP3, a group of nuclear RanGTP binding proteins. *FEBS Lett.* **427**, 330–336 (1998).

41. Partridge, J. R. & Schwartz, T. U. Crystallographic and biochemical analysis of the Ran-binding zinc finger domain. *J. Mol. Biol.* **391**, 375–389 (2009).

42. Shiomi, T. *et al.* Human dis3p, which binds to either GTP- or GDP-Ran, complements Saccharomyces cerevisiae dis3. *J. Biochem.* **123**, 883–890 (1998).

43. Suzuki, N. *et al.* The Saccharomyces cerevisiae small GTPase, Gsp1p/Ran, is involved in 3' processing of 7S-to-5.8S rRNA and in degradation of the excised 5'-A0 fragment of 35S pre-rRNA, both of which are carried out by the exosome. *Genetics* **158**, 613–625 (2001).

44. Nishitani, H. *et al.* Full-sized RanBPM cDNA encodes a protein possessing a long stretch of proline and glutamine within the N-terminal region, comprising a large protein complex. *Gene* **272**, 25–33 (2001).

45. Wang, D., Li, Z., Schoen, S. R., Messing, E. M. & Wu, G. A novel MET-interacting protein shares high sequence similarity with RanBPM, but fails to stimulate MET-induced Ras/Erk signaling. *Biochem. Biophys. Res. Commun.* **313**, 320–326 (2004).

46. Lampert, F. *et al.* The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *Elife* **7**, (2018).

47. Baker, R. P., Harreman, M. T., Eccleston, J. F., Corbett, A. H. & Stewart, M. Interaction between Ran and Mog1 is required for efficient nuclear protein import. *J. Biol. Chem.* **276**, 41255–41262 (2001).

48. Bao, X. *et al.* Mitosis-specific acetylation tunes Ran effector binding for chromosome segregation. *J. Mol. Cell Biol.* **10**, 18–32 (2018).

49. Oki, M. & Nishimoto, T. Yrb1p interaction with the gsp1p C terminus blocks Mog1p stimulation of GTP release from Gsp1p. *J. Biol. Chem.* **275**, 32894–32900 (2000).

50. Nicolás, F. J., Moore, W. J., Zhang, C. & Clarke, P. R. XMog1, a nuclear ran-binding protein in Xenopus, is a functional homologue of Schizosaccharomyces pombe mog1p that co-operates with RanBP1 to control generation of Ran-GTP. *J. Cell Sci.* **114**, 3013–3023 (2001).

51. Belhumeur, P. *et al.* GSP1 and GSP2, genetic suppressors of the prp20-1 mutant in Saccharomyces cerevisiae: GTP-binding proteins involved in the maintenance of nuclear organization. *Mol. Cell. Biol.* **13**, 2152–2161 (1993).

52. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005).

53. Drivas, G. T., Shih, A., Coutavas, E., Rush, M. G. & D'Eustachio, P. Characterization of four novel ras-like genes expressed in a human teratocarcinoma cell line. *Mol. Cell. Biol.* **10**, 1793–1798 (1990).

54. Fersht, A. R. Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry* **26**, 8031–8037 (1987).

55. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).

56. Cunningham, B. C. & Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081–1085 (1989).

57. Massova, I. & Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society* vol. 121 8133–8143 (1999).

58. Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004**, l2 (2004).

59. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).

60. Hietpas, R., Roscoe, B., Jiang, L. & Bolon, D. N. A. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat. Protoc.* **7**, 1382–1396 (2012).

61. Dunham, A. S. & Beltrao, P. Exploring amino acid functions in a deep mutational landscape. *Mol. Syst. Biol.* **17**, e10305 (2021).

62. Tong, A. H. Y. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).

63. Collins, S. R., Schuldiner, M., Krogan, N. J. & Weissman, J. S. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7**, R63 (2006).

64. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).

65. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).

66. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).

67. Braberg, H. *et al.* Genetic interaction mapping informs integrative structure determination of protein complexes. *Science* **370**, (2020).

68. Braberg, H. *et al.* From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**, 775–788 (2013).

69. Sundell, G. N. *et al.* Proteome-wide analysis of phospho-regulated PDZ domain interactions. *Molecular Systems Biology* vol. 14 (2018).

70. Bompard, G. *et al.* Subgroup II PAK-mediated phosphorylation regulates Ran activity during mitosis. *J. Cell Biol.* **190**, 807–822 (2010).

71. Wang Xiaodong & Hayes Jeffrey J. Acetylation Mimics within Individual Core Histone Tail Domains Indicate Distinct Roles in Regulating the Stability of Higher-Order Chromatin Structure. *Mol. Cell. Biol.* **28**, 227–236 (2008).

72. de Boor, S. *et al.* Small GTP-binding protein Ran is regulated by posttranslational lysine acetylation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3679-88 (2015).

73. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). *Cell* **38**, 835–840 (1984).

74. Weinkam, P., Chen, Y. C., Pons, J. & Sali, A. Impact of mutations on the allosteric conformational equilibrium. *J. Mol. Biol.* **425**, 647–661 (2013).

75. Krojer, T., Fraser, J. S. & von Delft, F. Discovery of allosteric binding sites by crystallographic fragment screening. *Curr. Opin. Struct. Biol.* **65**, 209–216 (2020).

76. Geyer, M. *et al.* Conformational states of the nuclear GTP-binding protein Ran and its complexes with the exchange factor RCC1 and the effector protein RanBP1. *Biochemistry* **38**, 11250–11260 (1999).

77. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).

78. Whitty, A. Cooperativity and biological complexity. *Nat. Chem. Biol.* **4**, 435–439 (2008).

79. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).

80. Humphris, E. L. & Kortemme, T. Design of multi-specificity in protein interfaces. *PLoS Comput. Biol.* **3**, e164 (2007).

81. Erijman, A., Aizner, Y. & Shifman, J. M. Multispecific recognition: mechanism, evolution, and design. *Biochemistry* **50**, 602–611 (2011).

82. Flock, T. *et al.* Selectivity determinants of GPCR–G-protein binding. *Nature* **545**, 317–322 (2017).

83. Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M. & Sexton, P. M. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **19**, 638–653 (2018).

84. Brown, J. H. Breaking symmetry in protein dimers: designs and functions. *Protein Sci.* **15**, 1–13 (2006).

85. Flynn, J. M., Mishra, P. & Bolon, D. N. A. Mechanistic Asymmetry in Hsp90 Dimers. *J. Mol. Biol.* **427**, 2904–2911 (2015).

86. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575 (2015).

87. Korkmazhan, E., Tompa, P. & Dunn, A. R. The role of ordered cooperative assembly in biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **22**, 647–648 (2021).

88. Brandman, O. & Meyer, T. Feedback loops shape cellular signals in space and time. *Science* **322**, 390–395 (2008).

89. El-Samad, H. Biological feedback control—Respect the loops. *Cell Systems* **12**, 477–487 (2021).

90. Zanzoni, A., Ribeiro, D. M. & Brun, C. Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cell. Mol. Life Sci.* **76**, 4407–4412 (2019).

91. Espinosa-Cantú, A., Cruz-Bonilla, E., Noda-Garcia, L. & DeLuna, A. Multiple Forms of Multifunctional Proteins in Health and Disease. *Front Cell Dev Biol* **8**, 451 (2020).

92. Espinosa-Cantú, A. *et al.* Protein Moonlighting Revealed by Noncatalytic Phenotypes of Yeast Enzymes. *Genetics* **208**, 419–431 (2018).

93. Jeffery, C. J. Proteins with neomorphic moonlighting functions in disease. *IUBMB Life* **63**, 489–494 (2011).

94. Arkin, A. & Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **67**, 560–578 (1994).

95. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical "wires." *Nature* **469**, 212–215 (2011).

96. Chen, Z. *et al.* De novo design of protein logic gates. *Science* **368**, 78–84 (2020).

97. Goldbeter, A. & Koshland, D. E., Jr. Ultrasensitivity in biochemical systems controlled by covalent modification. Interplay between zero-order and multistep effects. *J. Biol. Chem.* **259**, 14441–14447 (1984).

98. Ferrell, J. E., Jr & Ha, S. H. Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends Biochem. Sci.* **39**, 612–618 (2014).

99. Cao, Y., Lopatkin, A. & You, L. Elements of biological oscillations in time and space. *Nat. Struct. Mol. Biol.* **23**, 1030–1034 (2016).

100. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

101. Ferrell, J. E., Jr. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14**, 140–148 (2002).

102. Goldbeter, A. Dissipative structures in biological systems: bistability, oscillations, spatial patterns and waves. *Philos. Trans. A Math. Phys. Eng. Sci.* **376**, (2018).

103. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).

104. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Mol. Cell* **36**, 872–884 (2009).

105. Adler, M., Szekely, P., Mayo, A. & Alon, U. Optimal Regulatory Circuit Topologies for Fold-Change Detection. *Cell Syst* **4**, 171-181.e8 (2017).

106. Vert, G. & Chory, J. Crosstalk in cellular signaling: background noise or the real thing? *Dev. Cell* **21**, 985–991 (2011).

107. Gerardin, J., Reddy, N. R. & Lim, W. A. The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation. *Cell Syst* **9**, 297-308.e2 (2019).

108. Holehouse, A. S. & Pappu, R. V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* **57**, 2415–2423 (2018).

109. Ferrell, J. E., Jr & Ha, S. H. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends Biochem. Sci.* **39**, 496–503 (2014).

110. Goldbeter, A. & Koshland, D. E., Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6840–6844 (1981).

111. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405**, 823–826 (2000).

112. Rush, M. G., Drivas, G. & D'Eustachio, P. The small nuclear GTPase Ran: how much does it run? *Bioessays* **18**, 103–112 (1996).

113. Dasso, M. The Ran GTPase: theme and variations. *Curr. Biol.* **12**, R502-8 (2002).

114. Bischoff, F. R. & Ponstingl, H. Ran regulation by ran GEF and ran GAP. in *The Small GTPase Ran* 163–176 (Springer US, 2001).

115. Moore, M. S. & Blobel, G. The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. *Nature* **365**, 661–663 (1993).

116. Köhler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nat. Rev. Mol. Cell Biol.* **8**, 761–773 (2007).

117. Arnaoutov, A. & Dasso, M. The Ran GTPase regulates kinetochore function. *Dev. Cell* **5**, 99–111 (2003).

118. Ren, M. *et al.* Separate domains of the Ran GTPase interact with different factors to regulate nuclear protein import and RNA processing. *Mol. Cell. Biol.* **15**, 2117–2124 (1995).

119. Henriksen, P. *et al.* Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **11**, 1510–1522 (2012).

120. Besray Unal, E. *et al.* Systems level expression correlation of Ras GTPase regulators. *Cell Commun. Signal.* **16**, 46 (2018).

121. Görlich, D., Seewald, M. J. & Ribbeck, K. Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. *EMBO J.* **22**, 1088–1100 (2003).

122. Barr, F. A. Review series: Rab GTPases and membrane identity: causal or inconsequential? *J. Cell Biol.* **202**, 191–199 (2013).

123. Levy, E. D. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J. Mol. Biol.* **403**, 660–670 (2010).

124. Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J. & Wilke, C. O. Maximum allowed solvent accessibilites of residues in proteins. *PLoS One* **8**, e80635 (2013).

125. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).

126. Rojas, A. M., Fuentes, G., Rausell, A. & Valencia, A. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **196**, 189–201 (2012).

127. Collins, S. R., Roguev, A. & Krogan, N. J. Quantitative genetic interaction mapping using the E-MAP approach. *Methods Enzymol.* **470**, 205–231 (2010).

128. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002).

129. Cherry, J. M. *et al.* SGD: Saccharomyces Genome Database. *Nucleic Acids Res.* **26**, 73–79 (1998).

130. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

131. Usaj, M. *et al.* Thecellmap.org: A web-accessible database for visualizing and mining the global yeast genetic interaction network. *G3 (Bethesda)* **7**, 1539–1549 (2017).

132. Ryan, C. J. *et al.* Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell* **46**, 691–704 (2012).

133. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

134. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).

135. Jäger, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2011).

136. Jäger, S. *et al.* Purification and characterization of HIV-human protein complexes. *Methods* **53**, 13–19 (2011).

137. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

138. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

139. Teo, G. *et al.* SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J. Proteomics* **100**, 37–43 (2014).

140. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).

141. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

142. Smith, S. J. M. & Rittinger, K. Preparation of GTPases for structural and biophysical analysis. *Methods Mol. Biol.* **189**, 13–24 (2002).

143. Klebe, C., Prinz, H., Wittinghofer, A. & Goody, R. S. The kinetic mechanism of Ran--nucleotide exchange catalyzed by RCC1. *Biochemistry* **34**, 12543–12552 (1995).

144. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization

of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J. Biomol. NMR* **12**, 1–23 (1998).

145. Mishra, A. K. & Lambright, D. G. High-throughput assay for profiling the substrate specificity of Rab GTPase-activating proteins. *Methods Mol. Biol.* **1298**, 47–60 (2015).

146. Seewald, M. J. *et al.* Biochemical characterization of the Ran-RanBP1-RanGAP system: are RanBP proteins and the acidic tail of RanGAP required for the Ran-RanGAP GTPase reaction? *Mol. Cell. Biol.* **23**, 8124–8136 (2003).

147. Goudar, C. T., Sonnad, J. R. & Duggleby, R. G. Parameter estimation using a direct solution of the integrated Michaelis-Menten equation. *Biochim. Biophys. Acta* **1429**, 377–383 (1999).

148. Malaby, A. W. *et al.* Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. *J. Appl. Crystallogr.* **48**, 1102–1113 (2015).

149. Klebe, C., Bischoff, F. R., Ponstingl, H. & Wittinghofer, A. Interaction of the nuclear GTP-binding protein Ran with its regulatory proteins RCC1 and RanGAP1. *Biochemistry* **34**, 639–647 (1995).

150. Braberg, H. *et al.* Quantitative analysis of triple-mutant genetic interactions. *Nat. Protoc.* **9**, 1867–1881 (2014).

151. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *J. Classif.* **31**, 274–295 (2014).

152. Becker, J. *et al.* RNA1 encodes a GTPase-activating protein specific for Gsp1p, the Ran/TC4 homologue of Saccharomyces cerevisiae. *J. Biol. Chem.* **270**, 11860–11865 (1995).

153. Gideon, P. *et al.* Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity. *Mol. Cell. Biol.* **12**, 2050–2056 (1992).

154. Brinkmann, T. *et al.* Rap-specific GTPase activating protein follows an alternative mechanism. *J. Biol. Chem.* **277**, 12525–12531 (2002).

155. Schnell, S. Validity of the Michaelis-Menten equation--steady-state or reactant stationary assumption: that is the question. *FEBS J.* **281**, 464–472 (2014).

156. Hanson, S. M. & Schnell, S. Reactant stationary approximation in enzyme kinetics. *J. Phys. Chem. A* **112**, 8654–8658 (2008).

157. Faure, A. J. *et al.* Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022).

158. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* vol. 172 358-372.e23 (2018).

159. Hidalgo, F. *et al.* A saturation-mutagenesis analysis of the interplay between stability and activation in Ras. *Elife* **11**, (2022).

160. Bandaru, P. *et al.* Deconstruction of the Ras switching cycle through saturation mutagenesis. *Elife* **6**, 727 (2017).

161. Perica, T. *et al.* Systems-level effects of allosteric perturbations to a model molecular switch. *Nature* 1–6 (2021).

162. Jameson, N., Gavagan, M. & Zalatan, J. G. A kinetic mechanism for systems-level behavior in GTPase signaling. *Trends Biochem. Sci.* (2022) doi:10.1016/j.tibs.2022.01.006.

163. Dokholyan, N. V. Controlling Allosteric Networks in Proteins. *Chem. Rev.* **116**, 6463–6487 (2016).

164. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990 (2007).

165. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).

166. Stiffler, M. A., Hekstra, D. R. & Ranganathan, R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. *Cell* **160**, 882–892 (2015).

167. Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D. & Bolon, D. N. A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425**, 1363–1377 (2013).

168. Hietpas, R. T., Bank, C., Jensen, J. D. & Bolon, D. N. A. Shifting fitness landscapes in response to altered environments. *Evolution* **67**, 3512–3522 (2013).

169. Richards, S. A., Lounsbury, K. M. & Macara, I. G. 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 (1995).

170. Zhou, J. *et al.* GEF-independent Ran activation shifts a fraction of the protein to the cytoplasm and promotes cell proliferation. *Molecular Biomedicine* **1**, 18 (2020).

171. Holt, L. J. *et al.* Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682–1686 (2009).

172. Neuwald, A. F., Kannan, N., Poleksic, A., Hata, N. & Liu, J. S. 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 (2003).

173. Weinert, B. T. *et al.* Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. *Cell Rep.* **4**, 842–851 (2013).

174. Swaney, D. L. *et al.* Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* **10**, 676–682 (2013).

175. Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-Based Functional Decomposition of Proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).

176. Flynn, J. M. *et al.* Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife* **9**, (2020).

177. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).

178. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).

179. Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Research* vol. 43 D364–D368 (2015).

180. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

181. Vetter, I. R. The Structure of the G Domain of the Ras Superfamily. in *Ras Superfamily Small G Proteins: Biology and Mechanisms 1: General Features, Signaling* (ed. Wittinghofer, A.) 25–50 (Springer Vienna, 2014).

182. Vetter, I. R., Arndt, A., Kutay, U., Görlich, D. & Wittinghofer, A. Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell* **97**, 635–646 (1999).

183. Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39–46 (1999).

184. Chook, Y. M. & Blobel, G. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* **399**, 230–237 (1999).

185. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature* **374**, 378–381 (1995).

186. Renault, L., Kuhlmann, J., Henkel, A. & Wittinghofer, A. Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). *Cell* **105**, 245–255 (2001).

187. Park, H. *et al.* Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

188. Frenz, B. *et al.* Prediction of Protein Mutational Free Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. *Front Bioeng Biotechnol* **8**, 558247 (2020).

189. Ó Conchúir, S. *et al.* A Web Resource for Standardized Benchmark Datasets, Metrics, and Rosetta Protocols for Macromolecular Modeling and Design. *PLoS One* **10**, e0130433 (2015).

190. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838 (2011).

191. Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).

192. Purcell, E. M. Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).

193. AlQuraishi, M. Machine learning in protein structure prediction. *Curr. Opin. Chem. Biol.* **65**, 1–8 (2021).

194. Simanshu, D. K., Nissley, D. V. & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17–33 (2017).

195. Stalnecker, C. A. & Der, C. J. RAS, wanted dead or alive: Advances in targeting RAS mutant cancers. *Sci. Signal.* **13**, eaay6013 (2020).

196. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 4135–4139 (1974).

197. Hopfield, J. J., Yamane, T., Yue, V. & Coutts, S. M. Direct experimental evidence for kinetic proofreading in amino acylation of tRNAIle. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1164–1168 (1976).

198. Huang, W. Y. C. *et al.* A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science* **363**, 1098–1103 (2019).

199. McKeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5042–5046 (1995).

200. Martinez-Fabregas, J. *et al.* Kinetics of cytokine receptor trafficking determine signaling and functional selectivity. *Elife* **8**, (2019).

201. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5618–5622 (1993).

202. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019).

203. Wittmann, B. J., Johnston, K. E., Wu, Z. & Arnold, F. H. Advances in machine learning for directed evolution. *Curr. Opin. Struct. Biol.* **69**, 11–18 (2021).

204. Gumulya, Y. & Gillam, E. M. J. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. *Biochem. J.* **474**, 1–19 (2017).

205. Ovchinnikov, S. & Huang, P.-S. Structure-based protein design with deep learning. *Curr. Opin. Chem. Biol.* **65**, 136–144 (2021).

206. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

207. Romero, P. A., Krause, A. & Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E193-201 (2013).

208. Freschlin, C. R., Fahlberg, S. A. & Romero, P. A. Machine learning to navigate fitness landscapes for protein engineering. *Curr. Opin. Biotechnol.* **75**, 102713 (2022).

209. Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M. & Church, G. M. Unified rational protein engineering with sequence-based deep representation learning. *Nat. Methods* **16**, 1315–1322 (2019).

210. Biswas, S., Khimulya, G., Alley, E. C., Esvelt, K. M. & Church, G. M. Low-N protein engineering with data-efficient deep learning. *Nat. Methods* **18**, 389–396 (2021).

211. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455 (1994).

212. Pan, X. & Kortemme, T. Recent advances in de novo protein design: principles, methods, and applications. *J. Biol. Chem.* 100558 (2021).

213. Kapp, G. T. *et al.* Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5277–5282 (2012).

214. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).

215. Chen, Z. & Elowitz, M. B. Programmable protein circuit design. *Cell* **184**, 2284–2301 (2021).

216. Roy, K. R. *et al.* Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.* **36**, 512–520 (2018).

217. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

218. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

219. Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* **17**, 629–635 (2020).

220. Ursu, O. *et al.* Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-021-01160-7.

221. Bock, C. *et al.* High-content CRISPR screening. *Nat Rev Methods Primers* **2**, (2022).

222. Xu, L. *et al.* Mapping enzyme catalysis with metabolic biosensing. *Nat. Commun.* **12**, 6803 (2021).

223. de Groot, R., Lüthi, J., Lindsay, H., Holtackers, R. & Pelkmans, L. Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol. Syst. Biol.* **14**, e8064 (2018).

224. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018).

225. Schraivogel, D. *et al.* High-speed fluorescence image-enabled cell sorting. *Science* **375**, 315–320 (2022).

226. Cappelletti, V. *et al.* Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* (2020) doi:10.1016/j.cell.2020.12.021.

227. Richards, A. L., Eckhardt, M. & Krogan, N. J. Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**, e8792 (2021).

228. Stynen, B. *et al.* Changes of Cell Biochemical States Are Revealed in Protein Homomeric Complex Dynamics. *Cell* **175**, 1418-1429.e9 (2018).

Appendix A. Titles and Legends of Data Files for Chapter 2

Supplementary Data 1 Genetic interaction (GI) data from the E-MAP screens.

This data file contains genetic interaction (GI) scores (S-scores) from the E-MAP screens of 56 *S. cerevisiae* strains (wild type and 55 Gsp1 point mutants).

Column definitions:

query allele name (Gsp1 mutant): point mutation (amino acid substitution) in the *S. cerevisiae* Gsp1 gene (query gene in the E-MAP screen, see Methods and Collins et al[127]).

query allele ORF: open reading frame ID, a unique database identifier of the query gene Gsp1 (from the *Saccharomyces* Genome Database, yeastgenome.org).

array allele: allele name, either a gene deletion or a gene DAmP[64] (array gene in the E-MAP screen, see Methods and Collins et al[127]).

array allele ORF: open reading frame ID, a unique database identifier of the array gene (from the *Saccharomyces* Genome Database, yeastgenome.org).

E-MAP S-score: genetic interaction (GI) score between the query and the array alleles. See References[63,64,127] for definition.

Supplementary Data 2 Pairwise Pearson correlations of profiles between SGA genes and Gsp1 point mutants, with associated p-values.

This data file contains the Pearson correlation coefficients and accompanying p-values for correlations between genetic interaction profiles of Gsp1 point mutants and the genetic interaction profiles of *S. cerevisiae* alleles from the CellMap SGA dataset published in [14].

Column definitions:

mutant: point mutation (amino acid substitution) in the *S. cerevisiae* Gsp1 gene (query gene in the E-MAP screen).

CellMAP_allele: *S. cerevisiae* gene allele (gene deletion) from the CellMap.[65,131]

yeast_gene: standard gene name (as defined in the Saccharomyces Genome Database, yeastgenome.org) of the CellMap allele.

Pearson correlation: Pearson correlation between the genetic interaction profile of a Gsp1 mutant and the CellMAP allele (from the CellMap dataset[65,131]).

greater p-value: p-value associated with the Pearson correlation (one-sided positive t-test).

greater FDR: greater p-value after correction by the FDR method.

greater Bonferroni: greater p-value after Bonferroni correction.

Supplementary Data 3 Affinity purification mass spectrometry (AP-MS) data reported as fold change and significance value, as well as a list of significant interaction hits.

This data file contains two tables. Table 1 contains the affinity purification mass spectrometry (AP-MS) data for Gsp1 point mutants. The data in the table are the output from MSstats[140] (see Methods) and report on the abundance of the pulled down protein, the log-transformed fold change of the abundance compared to the wild type with the appropriate 3xFLAG tag, and the accompanying FDR adjusted p-value. The data are provided for both the global and equalized median normalization methods available.

Table 1 column definitions:

sample: unique identifier of the *S. cerevisiae* strain. Contains information on the position of the 3xFLAG tag (N- or C-terminal) and the point mutation (amino acid substitution) in the *GSP1* gene.

terminus position of the 3xFLAG tag: N- or C-terminus position of the 3xFLAG tag.

Gsp1 mutant: amino acid substitution in the GSP1 gene in *S. cerevisiae.*

normalization method (equalized median or global standard of PPI list): normalization method used in MSstats (eqM is equalized median, gs is global standard).

Prey protein ORF: open reading frame ID, a unique database identifier of the pulled-down protein interaction partner in the AP-MS experiment with Gsp1 point mutants (ORF ID defined in the *Saccharomyces* Genome Database, yeastgenome.org).

Prey protein gene name: standard gene name of the pulled-down protein interaction partner (as defined in the *Saccharomyces* Genome Database, yeastgenome.org).

log2 (fold change): log-transformed fold change of the abundance of pulled-down interaction partner between the point mutant and the wild type Gsp1 sample with the corresponding 3xFLAG tag (N- or C-terminal).

FDR adjusted p-value: FDR adjusted p-value of the fold change of abundance, from MSstats.

abundance of pulled down protein: total abundance of the pulled-down protein interaction partner.

Table 2 contains the list of high-confidence interaction partners of Gsp1 from our AP-MS experiments (as determined by SAINTexpress,[139] see Methods).

Table 2 column definitions:

C-terminal 3xFLAG tag: list of high-confidence protein interaction partners identified with wild type or mutant Gsp1 with the C-terminal 3xFLAG tag.

N-terminal 3xFLAG tag: list of high-confidence protein interaction partners identified with wild type or mutant Gsp1 with the N-terminal 3xFLAG tag.

ORF: open reading frame ID, a unique database identifier of the gene (as defined in the *Saccharomyces* Genome Database, yeastgenome.org).

gene name: standard gene name (as defined in the Saccharomyces Genome Database, yeastgenome.org)

Supplementary Data 4 S. cerevisiae genes from the SGA data with significant positive correlations with Gsp1 mutants organized by biological functions into gene sets.

This data file provides a list of *S. cerevisiae* alleles from the SGA dataset whose GI profiles have significant correlations with the GI profiles of Gsp1 mutants (see Methods). The genes were manually grouped into "gene sets" based on their biological function (as annotated in the *Saccharomyces* Genome Database, yeastgenome.org).

Column definitions:

Allele in the SGA CellMAP: *S. cerevisiae* gene allele (gene deletion) from the CellMap.[65,131]

S. cerevisiae gene name: standard gene name of the CellMap allele (as defined in the *Saccharomyces* Genome Database, yeastgenome.org).

gene set: annotated gene set (genes grouped by their annotated biological function from the *Saccharomyces* Genome Database, yeastgenome.org, and updated annually).

Cluster from Fig. 4a (1-7, or expanded dataset, see Methods): cluster number (1-7) corresponding to the hierarchical clustering presented in **Figure 2.17A**, referred to as Fig. 4a in the published manuscript.

Appendix B. Titles and Legends of Data Files for Chapter 3

Supplementary Data 5 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).

Supplementary Data 6 Gsp1 ΔΔG 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 ΔΔG of the mutation, in Rosetta Energy Units (REU), scaled so that 1 REU ~ 1 kcal/mol based on the benchmark data.

Supplementary Data 7 Benchmark ΔΔG 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.

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

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

ddg_expt: experimental ΔΔG of the mutation, in kcal/mol.

ddg_calc: computed ΔΔG of the mutation, in unscaled Rosetta Energy Units (REU).

ddg_calc_adj: computed ΔΔG of the mutation, in REU, scaled down so that 1 REU ~ 1 kcal/mol.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Christopher Mathy
4BDC691D69E3453...
Author Signature

5/30/2022
Date