

MIT Open Access Articles

*Hierarchical Organization Endows the
Kinase Domain with Regulatory Plasticity*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Creixell, Pau et al. "Hierarchical Organization Endows the Kinase Domain with Regulatory Plasticity." *Cell Systems* 7 (2018): 371-383 © 2018 The Author(s)

As Published: 10.1016/j.cels.2018.08.008

Publisher: Elsevier BV

Persistent URL: <https://hdl.handle.net/1721.1/124648>

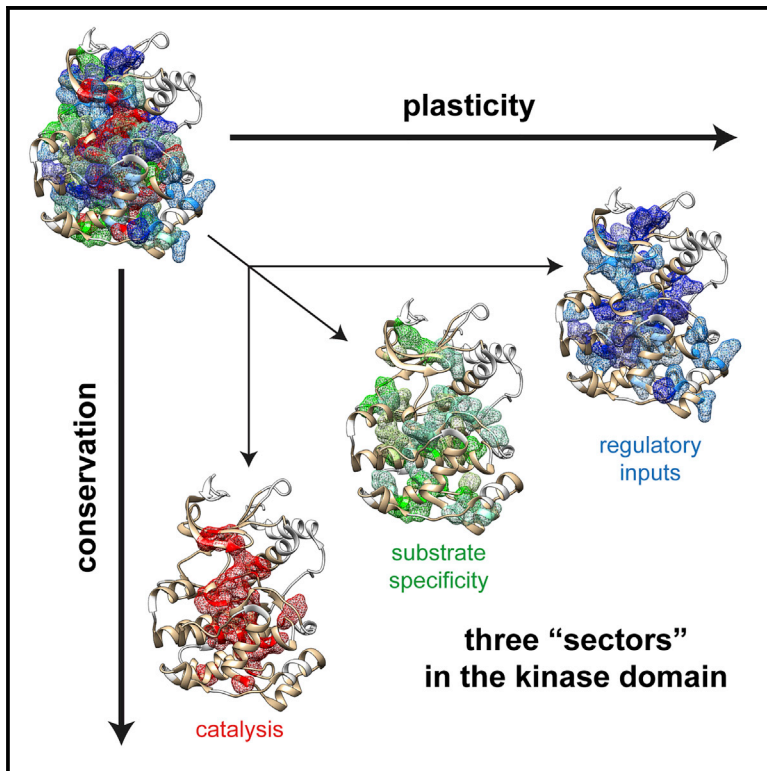
Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



Hierarchical Organization Endows the Kinase Domain with Regulatory Plasticity

Graphical Abstract



Authors

Pau Creixell, Jai P. Pandey, Antonio Palmeri, ..., Rama Ranganathan, David Pincus, Michael B. Yaffe

Correspondence

pincus@wi.mit.edu (D.P.), myaffe@mit.edu (M.B.Y.)

In Brief

Creixell et al. describe three quasi-independent “sectors”—groups of evolutionarily coupled amino acids—in the kinase domain that determine catalytic activity, substrate specificity, and regulatory interactions. The sectors are differentially utilized by subgroups of kinases and may help explain how the kinase domain evolved diverse regulatory inputs and substrate outputs.

Highlights

- Co-evolution analysis of the human kinome reveals three sectors in the kinase domain
- The three sectors encode for catalysis, substrate specificity, and regulation
- Comparative co-evolution analysis reveals sector remodeling across kinase subfamilies
- Sector positions are enriched for cancer mutations and bind allosteric inhibitors



Hierarchical Organization Endows the Kinase Domain with Regulatory Plasticity

Pau Creixell,^{1,2,3} Jai P. Pandey,⁴ Antonio Palmeri,^{5,9} Moitrayee Bhattacharyya,⁶ Marc Creixell,^{1,2,3} Rama Ranganathan,⁷ David Pincus,^{4,*} and Michael B. Yaffe^{1,2,3,8,10,*}

¹Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Center for Precision Cancer Medicine, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Departments of Biology and Bioengineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

⁵Viale Lido 94, 96012 Avola, Italy

⁶Department of Molecular and Cell Biology and California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720, USA

⁷Center for Physics of Evolving Systems, Department of Biochemistry and Molecular Biology, Institute for Molecular Engineering, University of Chicago, Chicago, IL 60637, USA

⁸Department of Surgery, Beth Israel Deaconess Medical Center, Divisions of Acute Care Surgery, Trauma, and Critical Care and Surgical Oncology, Harvard Medical School, Boston 02215, USA

⁹Present address: Celgene Institute for Translational Research Europe, 41003 Seville, Spain

¹⁰Lead Contact

*Correspondence: pincus@wi.mit.edu (D.P.), myaffe@mit.edu (M.B.Y.)

<https://doi.org/10.1016/j.cels.2018.08.008>

SUMMARY

The functional diversity of kinases enables specificity in cellular signal transduction. Yet how more than 500 members of the human kinome specifically receive regulatory inputs and convey information to appropriate substrates—all while using the common signaling output of phosphorylation—remains enigmatic. Here, we perform statistical co-evolution analysis, mutational scanning, and quantitative live-cell assays to reveal a hierarchical organization of the kinase domain that facilitates the orthogonal evolution of regulatory inputs and substrate outputs while maintaining catalytic function. We find that three quasi-independent “sectors”—groups of evolutionarily coupled residues—represent functional units in the kinase domain that encode for catalytic activity, substrate specificity, and regulation. Sector positions impact both disease and pharmacology: the catalytic sector is significantly enriched for somatic cancer mutations, and residues in the regulatory sector interact with allosteric kinase inhibitors. We propose that this functional architecture endows the kinase domain with inherent regulatory plasticity.

INTRODUCTION

The ability of cells to specifically respond to a wide variety of environmental cues is made possible by the capacity of signaling proteins to form both insulated and overlapping information processing networks. Protein kinases are critical nodes in these networks due to their ability to transmit information via phosphorylation, which can modify the activity, localization, interactions,

stability, and other functions of substrate proteins (Hanks and Hunter, 1995; Manning et al., 2002; Nolen et al., 2004; Fleuren et al., 2016). As such, kinases have diversified into more than 540 distinct proteins within the human proteome (Manning et al., 2002). While, by definition, all kinases share the core function of phosphorylating substrates, the evident specificity of signaling pathways indicates that kinases have evolved divergent substrate recognition capabilities and regulatory mechanisms. How these evolving kinase domain family members accomplished the balancing act of maintaining catalytic function while accommodating a diverse range of novel substrates and regulatory inputs remains a mystery.

Three features of protein kinases make solving this mystery particularly worthwhile. First, the kinase domain is the domain most often found encoded in cancer genes (Greenman et al., 2007). Second, there remains a major unmet need to develop allosteric drugs that perturb specific kinases. These drugs will have to take advantage of differences in substrate specificity and regulation rather than acting as ATP mimetics, which often results in off-target effects (Fleuren et al., 2016; Davis et al., 2011; Anastassiadis et al., 2011). Finally, from a synthetic biology perspective, the kinase domain represents a highly plastic molecular machine that should be able to be programmed to dynamically convert a broad range of molecular inputs to a diverse array of orthogonal outputs (Brunet and Pouyssegur, 1996; Roybal et al., 2016; Morsut et al., 2016).

Here, we sought to uncover the functional architecture of the kinase domain by identifying groups of co-evolving residues in the kinase domain, both at the whole-kinome level and within well-defined kinase subgroups. We find that there are three quasi-independent protein “sectors” in the kinase domain that appear to encode distinct functions: catalysis, substrate specificity, and regulation. The most conserved sector encodes the fundamental catalytic function required by all kinases, while the sectors that encode substrate recognition and regulatory inputs show progressively less conservation and more subgroup



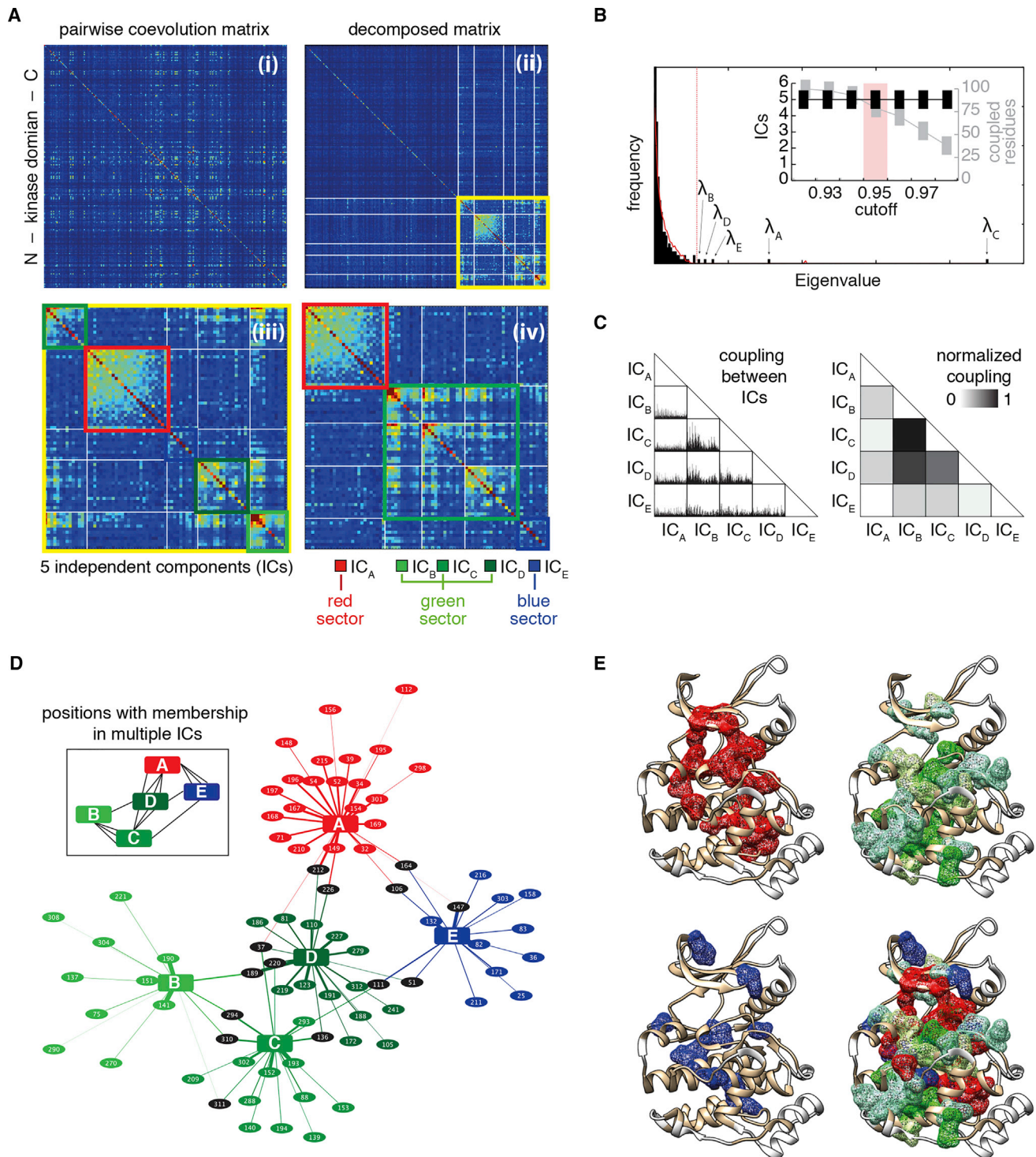


Figure 1. Statistical Coupling Analysis (SCA) of the Human Kinome Reveals Groups of Co-evolving Amino Acid Positions in the Kinase Domain

(A) SCA of the human kinome is used to calculate coupling between every pair of residues and every pair of positions in the alignment. The resulting high-dimensional matrix is further compressed into a two-dimensional coupling matrix, where the value at every position represents the degree of coupling between every pair of positions in the domain (i). From this coupling matrix, significantly coupled residues are identified by spectral decomposition and comparison to a randomized alignment in which residues are randomly reshuffled at each position in the alignment, thereby maintaining conservation while losing coupling (ii). Next, independent component analysis (ICA) allows the identification of “independent components” (ICs) or clusters of residues that are more coupled among

(legend continued on next page)

specificity. Mutational analysis coupled to quantitative reporter assays revealed strong functional enrichment for the putative regulatory sector positions and validated subgroup-specific sector positions. Finally, we find that sector residues are disrupted by cancer mutations and targeted by allosteric kinase inhibitors. Our results suggest a hierarchical organization of the kinase domain by which substrate recognition and regulatory inputs can be readily altered over evolution and tuned by mutations and inhibitors.

RESULTS

Three Groups of Co-evolving Residues Constitute Distinct Sectors within the Kinase Domain

To gain insight into the hardwired functional organization of the kinase domain, we applied statistical coupling analysis (SCA) to the human kinome (Halabi et al., 2009; Rivoire et al., 2016). To this end, from the original 540 human protein kinases, we first filtered out kinases with large insertions and deletions and generated a multiple sequence alignment containing 389 human kinases. After correcting for over-representation of sequences with high identity, this yielded 265 “effective sequences” (Figure S1A). SCA identified 79 residues in the kinase domain that show significant coupling across the alignment (Figure 1A). Matrix decomposition revealed five independent components (ICs) with correlated coupling patterns (Figures 1A, 1B, and S1B–S1F). Importantly, the same number of ICs was obtained across multiple independent runs of SCA, making these five ICs a robust feature of the alignment and independent of the stringency of the significance cutoff (Figure 1B). Upon reorganization of the decomposed matrix, we observed extensive inter-IC coupling between three of the ICs but relative isolation of the other two ICs (Figures 1A and 1C). 65 out of the 79 positions show unique membership in a single IC, while 14 contributed to more than one IC (Figure 1D). For subsequent analysis, we assigned these ambiguous positions to the IC with the highest coupling. Finally, we merged the highly coupled ICs and arrived at three quasi-independent groups of positions that we term the “red,” “green,” and “blue” sectors (Figure 1A). The physical locations of the sector positions are mapped on the structure of human extracellular signal-regulated kinase 2 (ERK2) (Figure 1E).

The human kinome is subdivided into seven canonical kinase family subgroups—AGC, CAMK, CMGC, STE, CK1, TKL, and TK—classified on the basis of function, sequence and structural similarity, evolutionary history, and substrate chemistry (Mann-

ing et al., 2002). To understand how the organization of the kinase domain diversified to enable subgroup-level specialization, we developed a variation of SCA termed comparative coupling analysis (CCA). CCA differs from SCA in that it uses information from protein family subgroups to predict similarities and differences in the sector-forming residues at the subgroup and the superfamily levels. In this way, CCA may provide insight into both the conserved and evolving functions the sectors encode (Figures 2A–2C and S2). The main advantage of CCA comes from the fact that by generating group-specific alignments, residues that do not appear to co-evolve throughout the superfamily, do indeed co-evolve within a specific group of kinases (Figure S2B). Given that, like SCA, CCA requires large and diverse sequence alignments, the method is particularly suited to study large and well-sampled protein superfamilies.

We applied CCA to the kinase superfamily and subgroups, restricting our analyses to homologs of the 405 human kinases that belong to the seven canonical kinase family subgroups. Using sequences from 15 different species ranging from *Homo sapiens* to *Giardia*, we identified 4,867 total kinases with clear membership in the seven kinase subgroups and aligned them using the human kinase domains as a reference (Table S1; see STAR Methods). We applied CCA to identify sets of amino acid positions that co-evolve within each kinase subgroup but not necessarily across the whole kinome. We mapped the independent components defined by CCA within each subgroup onto the three sectors defined at the kinome-wide level to determine how the networks of co-evolving positions are remodeled to generate subgroup identity. This analysis revealed global similarities as well as both critical and subtle differences in the subsets of the positions that form the sectors and in the degree of conservation of the sectors among the seven subgroups (Figures 2B, 2C, and S3; Tables S2 and S3). Indeed, while the red sector is compositionally conserved, meaning that the same residues largely comprise the red sector in all subgroups, the green and blue sectors show progressively less compositional conservation (Figures 2C and 2D). To investigate potential evolutionary differences between the three protein sectors, we estimated the number of non-synonymous substitutions that occurred in residues forming the different protein sectors (see STAR Methods) (Yang and Nielsen, 2000). The number of non-synonymous substitutions was significantly different for the three sectors, suggesting that they are distinct evolving units resulting from diverging evolutionary pressures (Figure 2E).

themselves than with the other residues (iii). Finally, in cases where significant coupling still exists between several “independent components,” they are considered part of the same protein sector (iv).

(B) Eigenspectrum for the coupling matrix derived from the protein kinase alignment (in black bars) compared to the one derived from the reshuffled protein kinase alignment (red solid line), providing the basis for the significant eigenmodes identified—that is, those greater than the second eigenvalue from the reshuffled alignment (as marked by the vertical dashed line). Based on this, these five independent components (ICs) are taken forward for subsequent analysis. To determine the number of residues significantly contributing to the different ICs an empirical statistical distribution is fitted to the contributions of each residue to each IC and a stringency “cutoff” is applied above which residues are considered to be part of the specific IC (Figure S1). Inset: after running SCA independently seven times with varying stringency cutoff values for the membership of residues to a given IC, while the number of ICs remains naturally robust, the number of coupled residues decreases as the cutoff value increases. As previously, the default value for this stringency cutoff is set at >95% of the CDF (Rivoire et al., 2016).

(C) Left: degree of coupling between different independent components (inter-IC coupling) quantified by plotting the amount of coupling between every pair of residues belonging to every pair of ICs. Right: normalized inter-IC coupling displayed as the integrated coupling signal for all residues within every pair of ICs.

(D) Significant contributions of all 79 coupled positions to each independent component (IC) with line thicknesses corresponding to the residue contribution to each IC. 14 positions significantly contribute to multiple ICs. Inset: the number of shared positions contributing to multiple ICs is schematized.

(E) The resulting red, green, and blue sectors plotted onto the structure of human ERK2 (PDB: 4QTE).

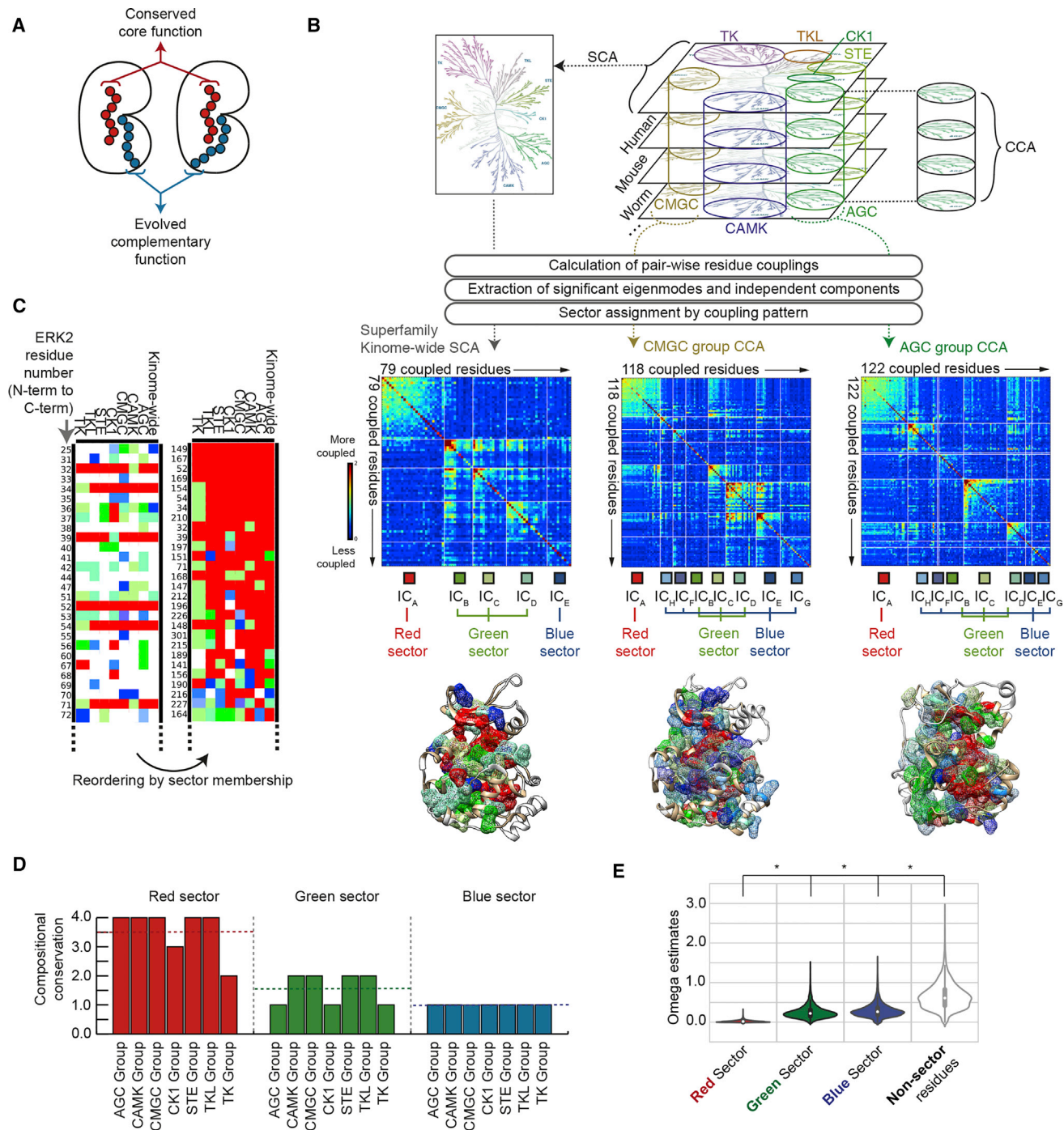


Figure 2. Comparative Coupling Analysis of the Kinase Domain

(A) Conceptual representation of amino acid residues driving specific molecular functions allows the distinction between those residues involved in *core* functions (such as catalysis, shown in red), which will be highly conserved in all instances of that domain, and those residues involved in *specialized* functions (like regulation, shown in blue) with higher freedom to evolve.

(B) Unlike SCA, where kinases from all groups are considered at once, for comparative coupling analysis (CCA), subgroup-specific alignments are performed using orthologs from 15 divergent genomes where each of the seven classically defined protein kinase subgroups (Manning et al., 2002) (TK, TKL, CK1, CMGC, STE, CAMK, and AGC) is considered individually. This subgroup-specific analysis typically identifies both a larger number and a distinct set of coupled residues as shown for AGC and CMGC. A full representation of all CCA results for the each of the different kinase subgroups is shown in Figure S3.

(C) Representation of a fraction of all coupled residues within the kinase domain colored by their independent component and sector membership in the different groups ordered by primary sequence (left) or by sector membership after clustering (right). For a full representation of all coupled residues, see Figure S3.

(legend continued on next page)

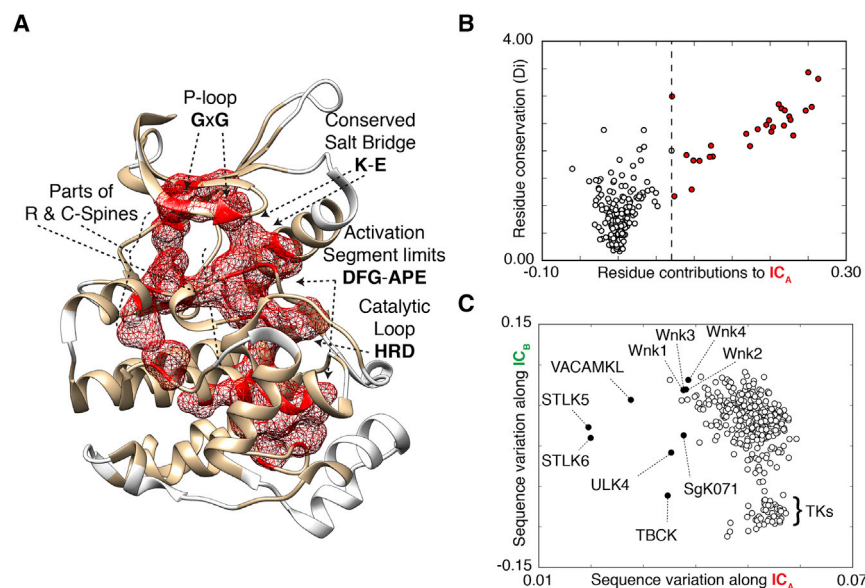


Figure 3. The Red Sector Drives Catalysis

(A) Depiction of residues contained within the red sector, illustrated on the structure of ERK2 (PDB: 4QTE). Residues that have been previously established as critical components of the catalytic core of the kinase domain are indicated.

(B) Comparison between the contribution of every residue within the kinase domain to the first independent component (the one constituting the red sector) and the degree of overall conservation of that same residue. Residue conservation is measured using the Kullback-Leibler relative entropy, D_i (see STAR Methods). The dashed line corresponds to the 95% of the CDF of an empirical statistical distribution fitted to the ICs (see STAR Methods).

(C) Scatterplot positioning protein kinases according to their sequence variation along the first and second independent components, as described in Figure 2B. The group of pseudo-kinases, which form the majority of kinases diverging along the first independent component (the red sector), are shown in filled black circles. A small subset of these pseudo-kinases, such as Wnk kinases (appropriately named With No K), has been

shown to maintain residual catalytic activity by evolving additional mutations that compensate for their deviations in key catalytic residues. Notably, the second independent component, part of the green sector, separates the tyrosine kinases (TKs). Sequence variation along an independent is defined and calculated by singular value decomposition (SVD) as described in the STAR Methods section.

The Red Sector Includes the Conserved Catalytic Core of the Kinase Domain

The red sector encompasses known kinase architectural determinants that are important for catalytic transfer of phosphate from ATP to a substrate hydroxyl group (Figures 3A and S4). These determinants include glycine residues in the P-loop, the DFG and APE motifs that delimit the activation segment, the catalytic loop HRD motif and parts of the catalytic and regulatory spines (Hanks and Hunter, 1995; Manning et al., 2002; Nolen et al., 2004; Kornev et al., 2006; Kornev et al., 2008), in addition to other residues that co-evolve with them and whose direct contribution to catalytic function has been less well recognized. Indeed, consistent with the core function of this sector, we observed a direct correlation between primary sequence conservation of a given residue and its contribution to the red sector (Pearson's correlation of 0.79, Figure 3B). However, this correlation that was not present for other sectors (Figure S4). The observation that key residues involved in catalysis were contained in the red sector was perplexing since strictly conserved residues that do not show co-variation with other amino acids should be invisible upon SCA and CCA analysis. Singular value decomposition (SVD) analysis revealed a group of known catalytically impaired pseudo-kinases (Zeqiraj and van Aalten, 2010) that were present in the full alignment helpfully serving as an out-group and providing sequence variations in the red sector (Figure 3C). Taken together, these results suggest that the red sector

positions delineate the deeply conserved catalytic core of the kinase domain.

Green Sector Composition Determines Substrate Specificity

The green sector displays intermediate compositional conservation (Figure 2D). The green sector is formed by residues that line and bracket the substrate binding site and includes known determinants of substrate specificity such as the P+1 loop and residues downstream of the HRD motif near the catalytic loop (Figure 4A) (Hanks and Hunter, 1995; Nolen et al., 2004; Creixell et al., 2015a). Indeed, in several kinase-substrate co-crystal structures, the green sector is the sector that makes the largest amount of direct substrate contact, as illustrated by the structure of AKT/PKB in complex with GSK-3 (Figure 4B) and the structure of PKA in complex with PKI (Figure S5). Further supporting a role in substrate recognition and specificity, SVD revealed that the three independent components that make up the green sector broadly separate kinases on the basis of their substrate specificity: tyrosine kinases (TKs) are clearly separated from the rest of the kinome, which comprises serine/threonine and dual-specificity kinases (Figures 4C and 3C). Moreover, SVD organized the non-TK kinome along a substrate specificity gradient, ranging from the proline-directed CMGC kinases toward basophilic-directed CAMK and AGC kinases (Figure 4C, upper right and lower left, respectively). As an orthogonal approach, we

(D) Quantification of compositional conservation for the different sectors as measured by the median number of other kinase subgroups for which a red sector residue (or green or blue) is also red (or green or blue).

(E) Estimation of the degree of negative selection for the different protein sectors identified within the kinase domain by calculating omega estimates corresponding to the number of synonymous and non-synonymous substitutions for the different residues, while correcting for multiple substitutions, transition/transversion rate biases, and base/codon frequency biases (Yang and Nielsen, 2000) (see STAR Methods). All groups of residues (i.e., red, green, blue, and non-sector residues) displayed significantly different omega estimates (Mann Whitney U, $p < 5.0 \times 10^{-65}$) as indicated by the asterisks.

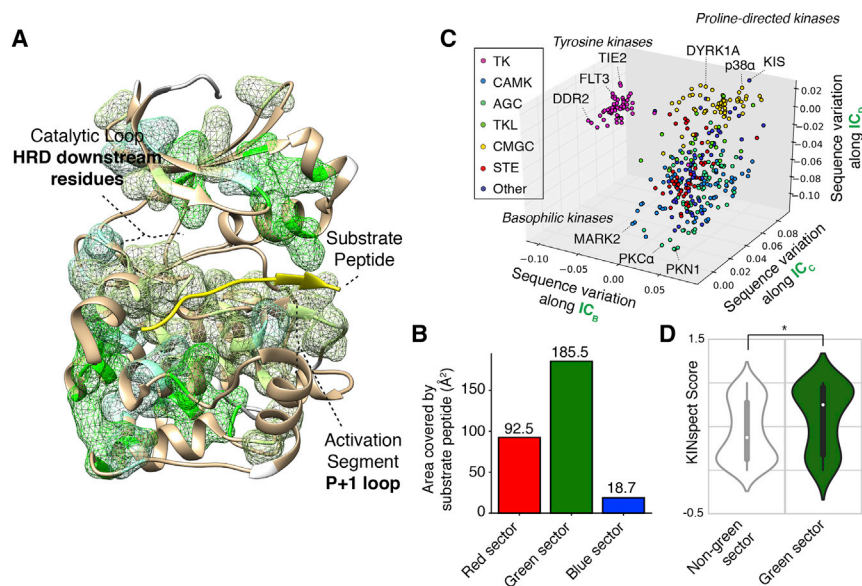


Figure 4. The Green Sector Encodes Substrate Peptide Specificity

(A) Illustration of residues contained within the green sector, superimposed on the structure of PKB/AKT in complex with a GSK-3 peptide substrate with sequence GRPRTTSFAE (PDB: 4EKK). Regions previously implicated as key determinants of substrate specificity are indicated.

(B) Barplot comparing the surface area of red, green, and blue sector residues that is buried by the peptide substrate in the structure of the PKB/AKT:GSK-3 substrate peptide complex, shown in (A). For this calculation, the solvent exposure of all residues is calculated in the presence and absence of the peptide substrate, and the differential exposure for the different sector residues is displayed.

(C) Scatterplot of sequence variation between all human kinases, relative to one another, of residues that form the second, third, and fourth independent components as described in Figure 2B. Each point indicates a particular kinase and is colored according to its major kinase group.

(D) Violin plot of the distribution of KINSpect scores, an orthogonal measure of the contribution of each

residue within the kinase domain toward substrate specificity (Creixell et al., 2015a), for residues belonging to the green sector compared to residues outside the green sector. The width of the violin at any particular KINSpect score indicates the number of residues that match that score. The difference between the green sector residues and other residues is significant as indicated by the asterisk (Mann Whitney U, $p = 0.005$).

compared the KINSpect score (Creixell et al., 2015a)—an established metric to quantify the likelihood that a residue has a role in substrate specificity—for each green sector residue across the kinase domain to the score for all non-green sector residues. Green sector residues showed a significantly higher KINSpect score than non-green sector residues (Mann Whitney U, $p = 0.005$, Figure 4D). Taken together, these results suggest that the green sector is composed of residues with functional roles in substrate recognition and specificity.

The Blue Sector Is Poised to Receive Regulatory Inputs

The most highly divergent sector between subgroups, the blue sector, appears to connect the active site and residues in the red and green sectors with more peripheral sites at the surface of the kinase domain (Figures 1E and 5A). This topology naturally suggests a role in transmitting regulatory inputs. Indeed, we found clear examples where regulatory interactions occur preferentially with blue sector sites, such as the binding site for cyclin A on CDK2 (Figures S6A and S6B) or an allosteric binding site on the yeast MAPK Fus3 for the scaffold protein Ste5 (Figures 5A–5C). Analysis of the network of local pairwise interactions between residues using “clique” and “community” graph theory approaches (Bhattacharyya and Vishveshwara, 2011; Bhattacharyya et al., 2013; Ghosh and Vishveshwara, 2008) (see STAR Methods) revealed a reorganization of the protein structure network upon Ste5 binding to Fus3 (Figures S6C–S6F). The unique clique residues in the apo Fus3 structure comprise a plurality of red sector residues (12 red sector residues, 9 green, and 7 blue). In contrast, the unique clique residues appearing in the Ste5-bound form of Fus3 contain a majority of blue sector residues (14 blue sector residues, 3 red, and 9 green), implicating them as putative mediators of allosteric regulation in Fus3 upon Ste5 binding. Taken together,

this evidence suggests that the blue sector is well poised to receive regulatory inputs.

Mutational Analysis Validates the Functionality of the Blue Sector

To investigate the functionality of the blue sector experimentally, we performed a comprehensive mutational analysis on Fus3 and employed quantitative activity assays in live cells. Fus3 is specifically activated in response to mating pheromone (α factor) and coordinates cell-cycle arrest with the transcriptional and morphological responses required for mating (Malleshaiah et al., 2010; Bhattacharyya et al., 2006; Good et al., 2009). We alanine-scanned all 49 residues comprising the blue sector along with 49 non-sector positions evenly distributed along the Fus3 primary sequence. We integrated each of these mutants as the only copy of Fus3 in the genome and assayed for Fus3 activity in response to different concentrations of α factor using a fluorescent reporter of the mating pathway (Figure 5C; Tables S5 and S6). In our strain background, reporter output depends strictly on Fus3, and wild-type (WT) Fus3 produces a graded α factor dose response (Figure 5C). In this assay, 3 of 49 non-sector mutants were distinguishable from WT (1 loss-of-function [LoF] and 2 gain-of-function [GoF] phenotypes) (Figure 5D), as defined by being statistically different in at least two doses of mating pheromone. By contrast, 20 of 49 of the blue sector mutants had altered activity compared to WT (17 LoF, 3 GoF), revealing a significant enrichment of functional sites in the blue sector (Fisher’s test, $p = 2.09 \times 10^{-8}$, Figures 5D and 5E).

Blue Sector Mutants Phenocopy Disrupted Allosteric Regulation

The functionality of the blue sector suggests that it may be a conduit for regulatory inputs. If natural regulatory interactions

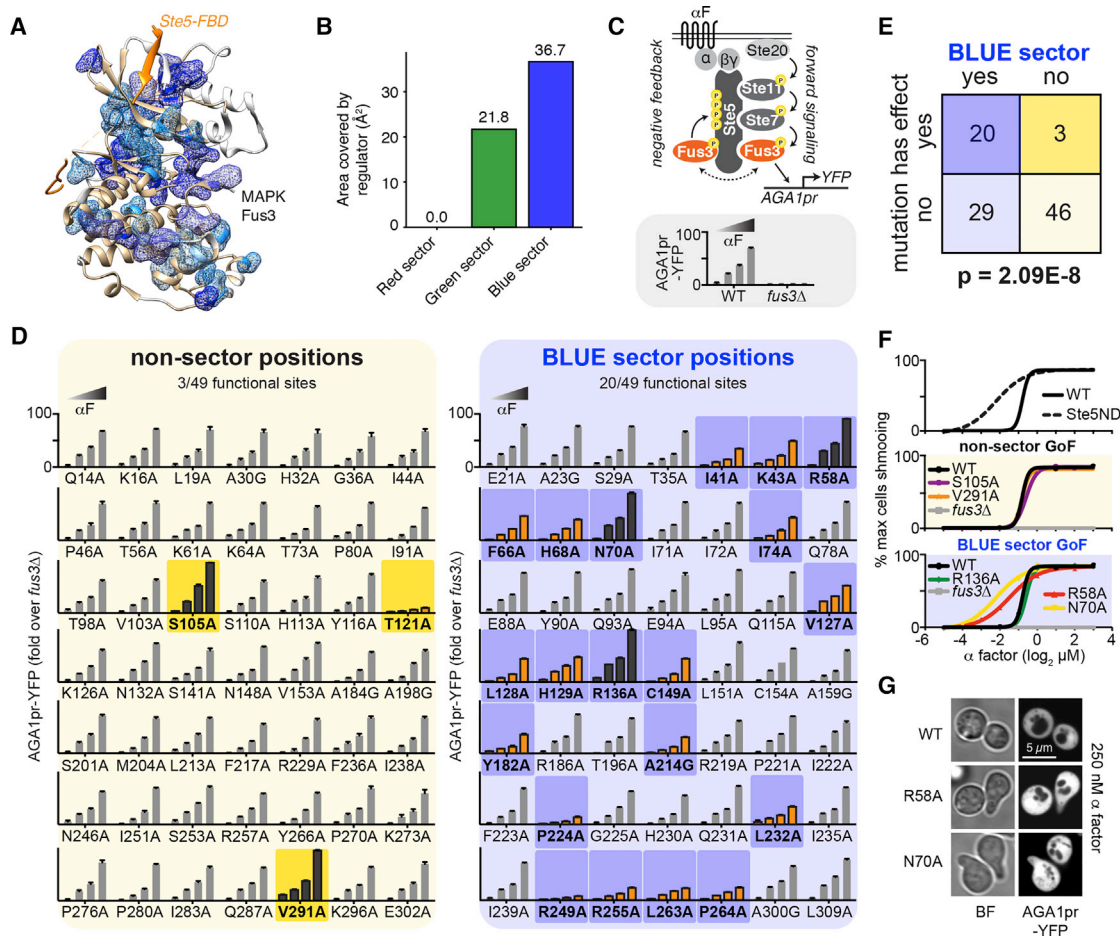


Figure 5. Mutational Analysis of Fus3 Reveals a Functionality for the Blue Sector in Mediating Regulatory Inputs

(A) Illustration of residues forming the blue sector, superimposed on the structure of the yeast MAP kinase Fus3 in complex with the Fus3-binding domain (FBD) of Ste5 colored in orange (PDB: 2F49).

(B) Barplot displaying the differential solvent-exposed area surfaces for red, green, and blue sector sites in Fus3 in the presence or absence of the FBD of Ste5, as described in the caption of Figure 4B.

(C) Schema of the yeast mating pathway, including the role of Fus3 and its allosteric regulation by Ste5-FBD to maintain the pathway in an inactive state. The downstream transcriptional reporter used in the functional assays for Fus3 activity shown in (D) is indicated (top). At the bottom, control experiments demonstrating the Fus3- and dose-dependency of the fluorescent signal by the reporter. Each value plotted is the mean of three biological replicates, and error bars show the standard deviation.

(D) On the left, fluorescence signal for the four doses of α -factor upon alanine scanning of 49 non-sector mutants, highlighting gain-of-function (GOF) mutants and loss-of-function (LOF) mutants in black and orange, respectively. Each value plotted is the mean of three biological replicates, and the error bars show standard deviation. On the right, fluorescence signal for the four doses of α -factor upon alanine scanning of 49 blue sector mutants, again highlighting GOF and LOF mutants in black and orange, respectively. Phenotypic mutants are highlighted by a darker background for the non-sector screen (yellow background) and the blue-sector screen (blue background).

(E) Statistical significance for the enrichment in phenotypic mutants when mutating sector sites as compared to non-sector sites (data from D).

(F) At the top, regulatory defects in the allosteric regulator Ste5, as illustrated for the case of a non-docking mutant, lead to more graded mating dose-responses as quantified by the percentage of "shmooring" cells. In the middle, non-sector GOF mutants display mating dose-responses that are comparable to wild-type. At the bottom, in contrast to GOF non-sector mutants, two out of the three GOF blue-sector mutants display more graded mating dose-responses phenocopying the effects observed for non-docking Ste5.

(G) Depiction of shmooring cells for the Fus3^{R58A} and Fus3^{N70A} in conditions that do not elicit shmooring in wild-type cells.

evolved to exploit the blue sector, then specific blue sector mutations should phenocopy the disruption of cognate regulation. To test this, we repurposed our GoF Fus3 mutants and performed additional functional assays. In addition to being regulated by its upstream MAPKK through canonical activation loop phosphorylation, Fus3 is regulated allosterically—both positively and negatively—by the scaffold protein Ste5 (Malle-

shaiah et al., 2010; Bhattacharyya et al., 2006; Good et al., 2009; Coyle et al., 2013). These dual modes of allosteric regulation allow the cell to simultaneously achieve a graded transcriptional response and a switch-like morphological response as a function of α factor concentration (Figure 5C). Several of the blue sector residues that have an impact on Fus3 upon mutagenesis appear to be junction residues between the Fus3-binding

domain (FBD) on Ste5 and the active site in Fus3 (Figures S6G and S6H). The FBD domain on Ste5 mediates the negative regulation of Fus3 required for the switch-like morphological transition that leads to formation of the mating projection or “shmoo” (Coyle et al., 2013). As such, disruption of the FBD results in graded shmooing across a dose response of α factor (Figure 5F) (Coyle et al., 2013). We hypothesized that some of the GoF Fus3 blue sector mutants may have lost the ability to be negatively regulated by Ste5. Consistent with this idea, while both of the non-sector GoF mutants retained switch-like shmoo responses, 2 of 3 of the blue sector GoF mutants showed graded shmooing (Figures 5F and 5G). Thus, mutation of blue sector residues recapitulated a phenotype associated with impaired allosteric regulation. Comparison of the cliques and communities found in Fus3 Ste5-bound and Ste7-bound structures (Figures S6I–S6K) highlighted how these two positions, N70 and R58, and their neighboring residues form inter-residue interactions between the effector binding site and the active site specifically when Fus3 binds its allosteric regulator, Ste5, but not upon binding with one of its (non-allosteric) substrates, Ste7.

CCA Reveals Private (Subgroup-Specific) Sector Residues

CCA not only identified the red, green, and blue sectors in each of the kinase subgroups, it also indicated that there are potentially important differences in the composition of the protein sector among the subgroups. In other words, whereas some sector sites are shared by several or all kinase subgroups, others are subgroup specific (Figure 2C). We hypothesized that these subgroup-specific sector residues, or “private” sites, would be enriched for functionality in members of the subgroup in question but not in out-groups. To test this, we first identified AGC and CMGC kinases as the two subgroups predicted to be the most functionally divergent (Figures 6A and S7). SVD revealed that the dissimilarity between AGC and CMGC sector composition was largely driven by 8 AGC-specific private sector sites that are non-sector sites in CMGC kinases (Figure 6B).

To experimentally test these AGC-specific sites, we performed mutational analysis on Pkc1, an essential AGC kinase in yeast (homolog of protein kinase C) required for cell wall integrity, and on Hog1, a CMGC kinase (yeast homolog of p38) required for adaptation to hyper-osmotic stress (Figure 6C). Mutation of 4 of 8 private AGC sector sites on Pkc1 resulted in phenotypic differences compared to WT: two of the sites were essential for viability, while two others altered levels of a fluorescent reporter upon exposure to the cell wall stressor zymolyase (Figure 6D). In contrast to Pkc1, Hog1 tolerated mutations at all 8 AGC-specific sector sites without altered induction of a fluorescent reporter upon addition of sorbitol (Figure 6D). By contrast, Hog1 cannot tolerate a red sector mutation (Figure 6D). Importantly, despite failing to complement for loss of WT Pkc1, the two inviable mutant proteins were expressed to WT levels and properly localized when co-expressed with WT Pkc1 (Figure 6E). These results demonstrate that the private AGC sector residues are specifically and significantly enriched for functionality in an AGC kinase (Fisher’s test, $p = 0.038$), validating the CCA predictions (Figure 6F).

Cancer Mutations and Allosteric Inhibitor Sites Preferentially Map onto Distinct Kinase Sectors

Finally, we assessed the relevance of kinase sectors to human disease. Since kinases are key targets of cancer mutations and therapeutic inhibitors (Fleuren et al., 2016), we cross-referenced the sector positions against known somatic cancer mutations (Table S4). After mapping 1,515,599 cancer somatic mutations onto canonical proteins, of which 14,860 mapped onto 13,152 sites within kinase domains, we observed a significant enrichment for cancer mutations at red sector sites (Wilcoxon test, $p = 2.4 \times 10^{-4}$; Figure 7B; STAR Methods section). Given that red sector residues comprise the catalytic core of the kinase domain and would thus be predicted to impair catalysis, we were surprised to discover red sector mutations not only in tumor suppressors but also in kinase oncogenes such as B-Raf and EGFR (Figures 7C and 7D). Importantly, in addition to mutations involving residues with well-established functions in catalysis (i.e., the DFG motif and Gly-rich loop), this mapping of cancer-associated mutations onto the red sector positions indicates potentially important roles for additional residues that have not been previously implicated in catalytic function. Moreover, even though their allosteric mechanism is unclear and appears to require the SH2 and SH3 domains for kinase inhibition (Zhang et al., 2010), we found that allosteric inhibitors of Abl bind at blue sector surface positions, underscoring the regulatory function of the blue sector (Figure 7E). Thus, distinct kinase sectors seem to be disrupted by mutations that promote cancer and targeted by small molecule allosteric inhibitors used to control it.

DISCUSSION

The discovery of distinctly folded domains in proteins led to the interpretation that domains represent units of evolution and function (Wetlauffer, 1973; Bork, 1991). However, it has become clear that there is a large degree of functional and evolutionary heterogeneity among the residues that form a domain. Here, we found that the kinase domain is organized in a functional hierarchy that allows a deeply conserved catalytic function to be differentially deployed over evolution to act on distinct sets of substrates and respond to specific regulatory inputs. By formally defining sets of residues likely to encode each of these molecular functionalities using SCA and CCA, it becomes tractable to study how these functions may have evolved or are perturbed in diseases such as cancer (Figure 7).

Previous studies using complementary computational approaches have identified networks of interacting residues or “communities” within the kinase domain based on molecular dynamics simulations (McClendon et al., 2014). Although our current analysis does not significantly correlate with functional sub-domains in the model kinase PKA defined by this analysis, the differences in the two approaches should be noted. While community analysis captures relevant structural and dynamic interactions between residues, our approach is rooted on co-evolution patterns spanning many protein kinases to identify a more stringent subset of functional residues within the kinase domain. The different results obtained using these orthogonal approaches demonstrate the need for an intermediate level of resolution between group averages and individual examples (Figure S2).

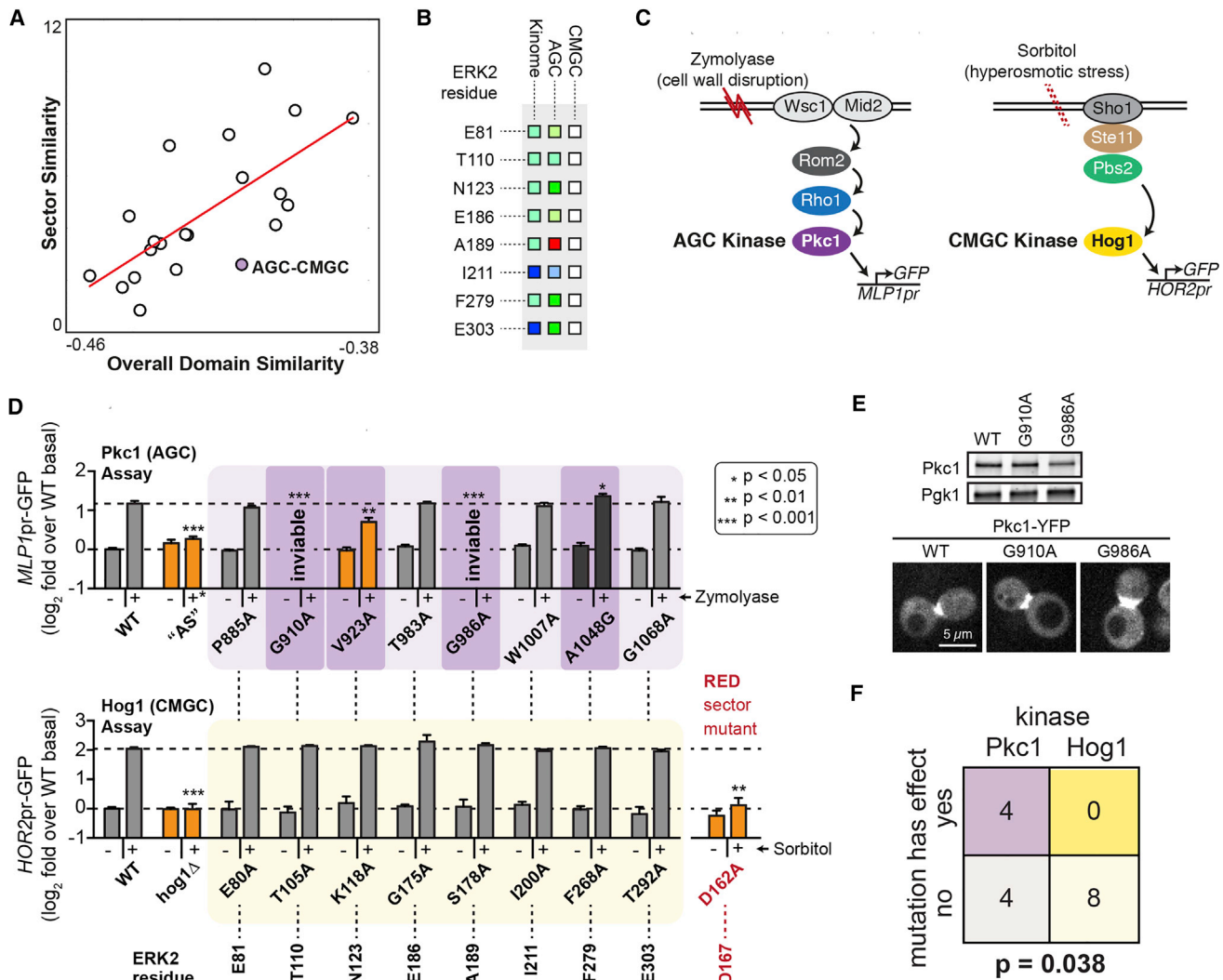


Figure 6. Private Sector Sites Encode Subfamily-Specific Functions

(A) Scatterplot including all pairs of kinase groups comparing their overall kinase domain similarity, as measured by overall BLOSUM distance (Henikoff and Henikoff, 1992), in the x axis to degree of overlap in their sectors in the y axis (see STAR Methods). Between the AGC and CMGC subgroups, the residues that define the sectors show limited overlap, despite a moderately high degree of overall similarity between the kinase domain sequences of AGC and CMGC family members.

(B) Sector memberships of the eight sites that are part of a sector in the kinome-wide SCA and in the AGC-specific CCA but not in CMGC-specific CCA. These sites are predicted to drive the functional divergence between AGC and CMGC kinases. The amino acid numbering shown corresponds to these sites in the representative structure used to map all CCA models, namely ERK2 (PDB: 4QTE) as described in the STAR Methods section.

(C) Left: functional assay for Pkc1 where the MLP1-driven downstream transcriptional reporter is activated upon the addition of zymolyase and the subsequent cell wall lysis. Right: functional assay for Hog1 where the HOR2-driven downstream transcriptional reporter is triggered by hyperosmotic stress resulting from the addition of sorbitol.

(D) Reporter signal for Pkc1 (top) and Hog1 (bottom) mutants upon mutating the eight AGC-specific private sector sites in both kinases. Zymolyase was added at 100 μ g/mL for 2 hr, and sorbitol was added at 0.8 M for 1 hr. Each value plotted is the mean of three biological replicates, and the error bars are the standard deviation. Analog-sensitive Pkc1 ("AS") in combination with the analog-specific inhibitor 1NM-PP1 (marked with +) and Hog1 deletion (*hog1* Δ) were used as positive controls for the two assays. A mutant in the aspartate of the DFG motif forming the red sector of Hog1 was used as point mutant control to confirm the Hog1 assay sensitivity to loss-of-function mutants of Hog1. Similarly as in Figure 5D, phenotypic mutants are highlighted by a darker background for Pkc1 screen (darker brown background). No phenotypic mutant was found for any of the eight Hog1 mutants (yellow background).

(E) Protein expression and cellular localization of wild-type Pkc1, Pkc1^{G910A}, and Pkc1^{G986A} as assayed by western blot and microscopy.

(F) Fisher's test results assessing the enrichment of phenotypic mutations in Pkc1, as compared to Hog1, upon mutation AGC-specific sector positions (data from Figure 5D).

In this study, we set out to understand how the kinase domain can maintain its catalytic function while evolving diversity and specialization of regulatory inputs and substrate out-

puts. Our results suggest that a subset of residues within the kinase domain display a significant degree of co-evolution, representing functional units hardwired into the structure of the

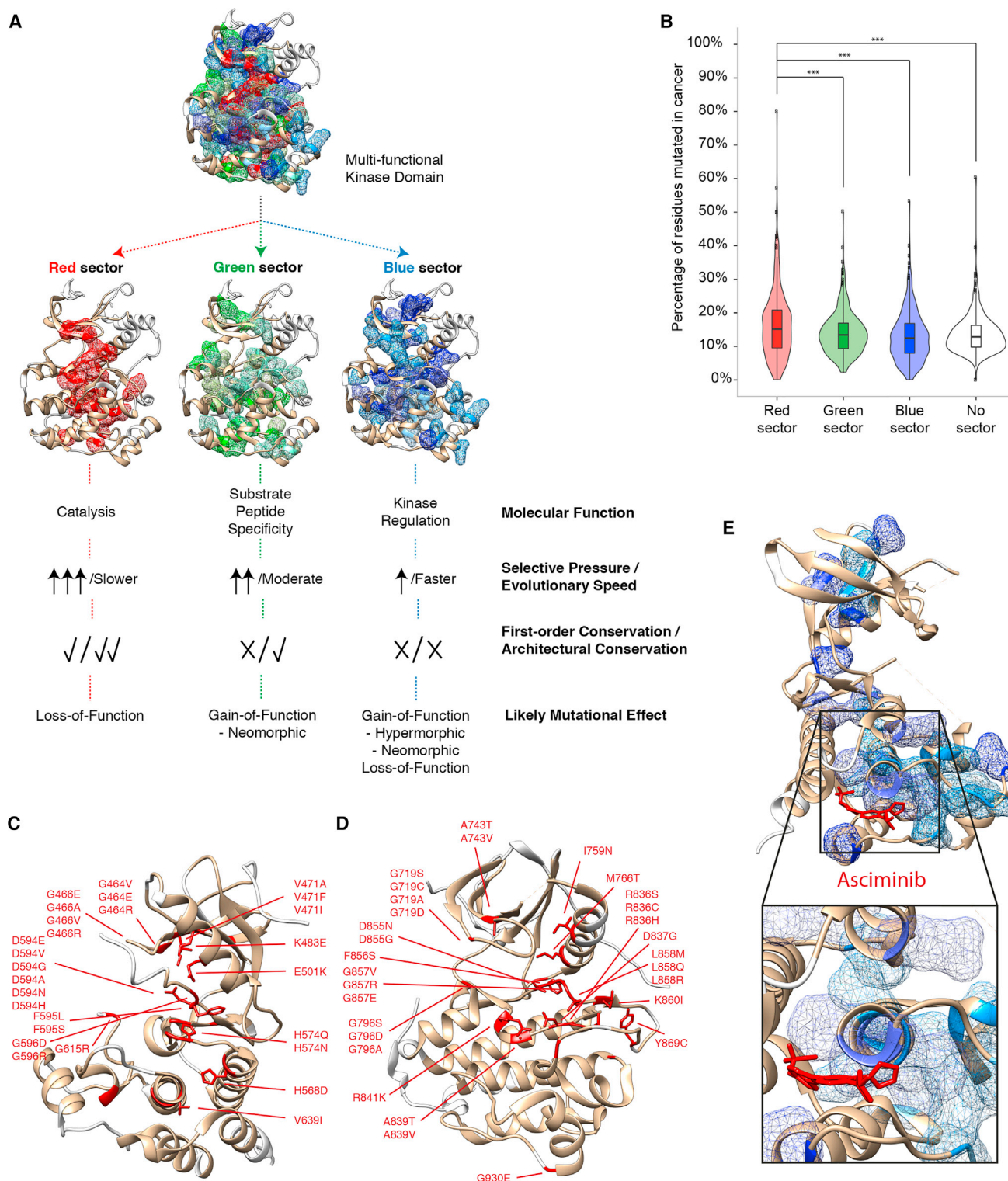


Figure 7. The Hierarchical Organization of the Kinase Domain Is Targeted by Somatic Cancer Mutations and Allosteric Inhibitors

(A) General model of the kinase domain as a highly heterogeneous multi-functional domain with sets of residues encoding distinct molecular functions and being constrained by different selective pressures and evolutionary speeds leading to differential conservation and effects upon mutation.

(B) Violin plots displaying the distribution of the percentage of residues belonging to the red, green, blue, and non-sector sites that are mutated in cancer, using data from the COSMIC repository (Forbes et al., 2011). Each data point represents a specific protein kinase, for which the percentage of residues in each sector that contains one or more cancer mutations was calculated. *** $p < 0.001$ as in Figure 6D.

(legend continued on next page)

kinase domain. Without *a priori* assumptions about the number of pseudo-independent co-evolutionary units or protein sectors that may be present within the kinase domain, spectral decomposition resolved the kinome-wide co-evolution matrix into five independent components (IC), and patterns of coupling between ICs allowed us to collapse these into three sectors (Figures 1 and S1). More generally, when applying SCA and CCA to other protein domains, the number of ICs and sectors is not necessarily the number of functional units; rather, the ICs and sectors provide an organizational framework for further investigation as described elsewhere (Rivoire et al., 2016). We would argue that the number of ICs determines the upper bound of discrete functionalities within a domain, and strong inter-IC coupling suggests variations on a theme rather than distinct functions. The complementary use of CCA to further analyze distinct protein groups within a protein superfamily then presents two main advantages: it expands the number of coupled residues that can be identified (e.g., compare Figures 1E and 5A), and it provides insights into how these themes are elaborated to allow for functional radiation and specialization (Figure 6).

In the kinase domain, the residues contributing to the catalytic core—the red sector sites—are limited in their ability to evolve due to their fundamental importance. Consequently, the red sector is characterized by a high degree of residue conservation and a resulting low degree of evolution (Figures 2D and 2E). By contrast, the green sector is formed by residues involved in the recognition of the substrate peptide, a molecular function that must—and indeed does—allow plasticity between kinase subgroups to accommodate varied substrates. Finally, the blue sector presents the widest degree of plasticity and divergence between kinase groups, consistent with different kinases evolving divergent regulatory mechanisms (Figure 7A). Accordingly, disease mutations can alter catalytic activity via the red sector, but perhaps more sinisterly, may modify substrate specificity or regulatory inputs to subvert and reroute signaling through alterations in the green and blue sector (Songyang et al., 1995; Creixell et al., 2015b).

There are two notable exceptions to the rule that the red sector is invariant. First, tyrosine kinases represent a functionally divergent subgroup, at least in part driven by their capacity to phosphorylate tyrosine residues instead of serine and/or threonine residues. While the red sector appears to be well defined and compact in all other kinase subgroups, tyrosine kinases present large divergence in all sectors including the red sector. Second, pseudo-kinases also show divergence in the red sector where they have accumulated mutations leading to their impaired catalytic ability. Thus, these exceptions can be easily interpreted and ultimately serve to further support the notion that the red sector drives catalysis.

Beyond the red sector, CCA highlights the functional and regulatory plasticity present in different protein kinases by revealing the variability in the blue sector among the kinase subgroups. At the same time, while CCA allows for this more granular understanding of the architecture of the kinase domain

by incorporating subgroup-level information, the *in silico* identification of functions that are idiosyncratic to individual kinases remains a significant challenge. This limitation arises from the need to capture enough sequences of sufficient diversity. Ultimately, the specific regulatory sites and interactions that control an individual kinase will need to be resolved using focused structural molecular dynamics and biochemical and molecular genetic approaches. Despite this caveat, our *in vivo* mutagenesis and functional screens suggest that quantitative experiments performed on a specific protein kinase can recapitulate the functional principles predicted *in silico* at the subgroup level. After validating the function of these sectors orthogonally, our models provide a means to identify trends and hypothesize mechanisms of action for disease-associated mutations or kinase inhibitors, which can be further tested in focused experiments.

Our current models remain blind to the possibility of individual residues performing multiple overlapping molecular functions, though we find that 14 of 79 coupled positions show membership in multiple ICs (Figure 1D). To facilitate interpretation and subsequent analysis, we forced these residues to have membership in only their IC with highest coupling. However, it is likely that several of these positions play overlapping roles in catalysis, substrate specificity, and/or kinase regulation. Similarly, while we define the three sectors as separate entities, there are clear differences in how related to one another the different sectors are. In particular, the blue sector—which connects putative allosteric sites on the kinase surface to the red and green sectors and the active site—includes residues that interface with the other sectors and may contribute to all three functionalities.

Finally, we observed a relationship between kinase sector positions and human disease progression and treatment. We found enrichment for somatic cancer mutations at red sector sites—not only in tumor-suppressor kinases but also in oncogenes (Figures 7C and 7D). While initially counterintuitive, since mutations in the red sector would be expected to impair catalytic activity, this finding suggests that there may be more examples of inactivating mutations leading to roles in trans-activation as has been reported for certain B-Raf mutations (Nieto et al., 2017; Yao et al., 2017; Wan et al., 2004; Haling et al., 2014). In addition, previous studies have suggested that the activation segment and other flexible regions of the kinase domain are hotspots for mutations in EGFR, B-Raf, and other protein kinases (Dixit et al., 2009; Paladino et al., 2015; Kiel et al., 2016). Mutations in these regions could modulate conformational changes and activation transitions without compromising structural stability of the kinase fold (Paladino et al., 2015; Kiel et al., 2016). Finally, it is exciting to consider that the blue sector residues that we have implicated in kinase regulation appear to serve as portals for allosteric inhibitors (Figure 7E). Targeting blue sector surface sites may lead to the development of next-generation allosteric modulators.

(C) Twenty-seven unique cancer somatic mutations perturbing red sector residues in B-Raf (PDB: 4MBJ).

(D) Thirty unique cancer somatic mutations perturbing red sector residues in EGFR (PDB: 2GS2).

(E) Tyrosine kinase (TK) blue sector residues contact the allosteric inhibitor asciminib in the co-crystal structure of the Abl:inhibitor complex (PDB: 5MO4).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Yeast Strains
- **METHOD DETAILS**
 - Plasmids
 - Site-Directed Mutagenesis
 - Cell Growth and Treatment with α Factor
 - Flow Cytometry
 - Confocal Microscopy
 - Western Blotting
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Statistical Coupling Analysis
 - Singular Value Decomposition and Mapping of Sequence Variations along Independent Components
 - Comparative Coupling Analysis
 - Estimation of Negative Selection (Omega Estimates)
 - Residue Conservation
 - Calculation of Area Covered by Substrate Peptide or Kinase Regulator in Kinase Structures
 - Construction of Protein Structure Network
 - Network Parameters
 - Measuring Similarity between Kinase Groups
 - Mapping of Somatic Cancer Mutations

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.cels.2018.08.008>.

ACKNOWLEDGMENTS

We would like to thank Brian A. Joughin, Daniel Lim, and all other members of the Yaffe Lab for discussion and critical input leading to this work. We are grateful to Z. Feder and J. Krakowiak for technical assistance and to the Whitehead Institute FACS facility and the Keck Microscopy facility. Since performing the work described, A.P. has become an employee of Celgene Research SL, part of Celgene Corporation, and declares no conflicts of interest. This work was supported by a Merck Postdoctoral Fellowship from the Helen Hay Whitney Foundation to P.C.; an NIH Early Independence Award DP5 OD017941 to D.P.; NIH grants CA226393 to P.C. and GM104047, ES015339, and ES028374 to M.B.Y.; the Charles and Marjorie Holloway Foundation (M.B.Y.); and the MIT Center for Precision Cancer Medicine (P.C., M.C., and M.B. Y.).

AUTHOR CONTRIBUTIONS

Conceptualization, P.C., D.P., R.R., and M.B.Y.; Methodology, P.C., D.P., and M.B.Y.; Investigation, P.C., J.P.P, D.P., A.P., M.B., and M.C.; Writing, P.C., D.P., and M.B.Y. with input from all authors; Supervision, P.C., D.P., and M.B.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 24, 2018

Revised: June 24, 2018

Accepted: August 13, 2018

Published: September 19, 2018

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Jeffrey et al. (1995) and Zheng et al. (1993).

REFERENCES

- Adamcsek, B., Palla, G., Farkas, I.J., Derényi, I., and Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021–1023.
- Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H., and Peterson, J.R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* 29, 1039–1045.
- Bhattacharyya, M., Bhat, C.R., and Vishveshwara, S. (2013). An automated approach to network features of protein structure ensembles. *Protein Sci.* 22, 1399–1416.
- Bhattacharyya, M., Ghosh, S., and Vishveshwara, S. (2016). Protein structure and function: looking through the Network of Side-Chain Interactions. *Curr. Protein Pept. Sci.* 17, 4–25.
- Bhattacharyya, M., and Vishveshwara, S. (2011). Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry* 50, 6225–6236.
- Bhattacharyya, R.P., Reményi, A., Good, M.C., Bashor, C.J., Falick, A.M., and Lim, W.A. (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* 311, 822–826.
- Bork, P. (1991). Shuffled domains in extracellular proteins. *FEBS Lett.* 286, 47–54.
- Brunet, A., and Pouyssegur, J. (1996). Identification of MAP kinase domains by redirecting stress signals into growth factor responses. *Science* 272, 1652–1655.
- Chaikuad, A., Tacconi, E.M., Zimmer, J., Liang, Y., Gray, N.S., Tarsounas, M., and Knapp, S. (2014). A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. *Nat. Chem. Biol.* 10, 853–860.
- Coyle, S.M., Flores, J., and Lim, W.A. (2013). Exploitation of latent allostery enables the evolution of new modes of MAP kinase regulation. *Cell* 154, 875–887.
- Creixell, P., Palmeri, A., Miller, C.J., Lou, H.J., Santini, C.C., Nielsen, M., Turk, B.E., and Linding, R. (2015a). Unmasking determinants of specificity in the human kinome. *Cell* 163, 187–201.
- Creixell, P., Schoof, C.D., Longden, J., Miller, C.J., Lou, H.J., Perryman, L., Cox, T.R., Zivanovic, N., Palmeri, A., et al. (2015b). Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* 163, 202–217.
- Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., and Zarrinkar, P.P. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 29, 1046–1051.
- Dixit, A., Yi, L., Gowthaman, R., Torkamani, A., Schork, N.J., and Verkhivker, G.M. (2009). Sequence and structure signatures of cancer mutation hotspots in protein kinases. *PLoS One* 4, e7485.
- Flouren, E.D.G., Zhang, L., Wu, J., and Daly, R.J. (2016). The kinome ‘at large’ in cancer. *Nat. Rev. Cancer* 16, 83–98.
- Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., et al. (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 39, D945–D950.
- Ghosh, A., and Vishveshwara, S. (2008). Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry* 47, 11398–11407.
- Good, M., Tang, G., Singleton, J., Reményi, A., and Lim, W.A. (2009). The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. *Cell* 136, 1085–1097.
- Greenman, C., Stephens, P., Smith, R., Dalgleish, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.

- Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138, 774–786.
- Haling, J.R., Sudhamsu, J., Yen, I., Sideris, S., Sandoval, W., Phung, W., Bravo, B.J., Giannetti, A.M., Peck, A., Masselot, A., et al. (2014). Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell* 26, 402–413.
- Hanks, S.K., and Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* 9, 576–596.
- Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89, 10915–10919.
- Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., and Pavletich, N.P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376, 313–320.
- Kannan, N., and Vishveshwara, S. (1999). Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* 292, 441–464.
- Katoh, K., and Frith, M.C. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* 28, 3144–3146.
- Kiel, C., Benisty, H., Lloréns-Rico, V., and Serrano, L. (2016). The yin-yang of kinase activation and unfolding explains the peculiarity of Val600 in the activation segment of BRAF. *Elife* 5, e12814.
- Kornev, A.P., Haste, N.M., Taylor, S.S., and Eyck, L.F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. USA* 103, 17783–17788.
- Kornev, A.P., Taylor, S.S., and Ten Eyck, L.F. (2008). A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. USA* 105, 14377–14382.
- Lee, S.J., Han, B.G., Cho, J.W., Choi, J.S., Lee, J., Song, H.J., Koh, J.S., and Lee, B.I. (2013). Crystal structure of Pim1 kinase in complex with a pyrido[4,3-D]pyrimidine derivative suggests a unique binding mode. *PLoS One* 8, e70358.
- Longtine, M.S., McKenzie, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953–961.
- Malleshaiah, M.K., Shahrezaei, V., Swain, P.S., and Michnick, S.W. (2010). The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. *Nature* 465, 101–105.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298, 1912–1934.
- McClendon, C.L., Kornev, A.P., Gilson, M.K., and Taylor, S.S. (2014). Dynamic architecture of a protein kinase. *Proc. Natl. Acad. Sci. USA* 111, E4623–E4631.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biol.* 17, 122.
- Morsut, L., Roybal, K.T., Xiong, X., Gordley, R.M., Coyle, S.M., Thomson, M., and Lim, W.A. (2016). Engineering customized cell sensing and response behaviors using synthetic Notch receptors. *Cell* 164, 780–791.
- Newhouse, B.J., Wenglow, S., Grina, J., Laird, E.R., Voegtli, W.C., Ren, L., Ahrendt, K., Buckmelter, A., Gloor, S.L., Klopfenstein, N., et al. (2013). Imidazo [4,5-b]pyridine inhibitors of B-Raf kinase. *Bioorg. Med. Chem. Lett.* 23, 5896–5899.
- Ni, L., Li, S., Yu, J., Min, J., Brautigam, C.A., Tomchick, D.R., Pan, D., and Luo, X. (2013). Structural basis for autoactivation of human Mst2 kinase and its regulation by RASSF5. *Structure* 21, 1757–1768.
- Nieto, P., Ambrogio, C., Esteban-Burgos, L., Gómez-López, G., Blasco, M.T., Yao, Z., Marais, R., Rosen, N., Chiarle, R., Pisano, D.G., et al. (2017). A Braf kinase-inactive mutant induces lung adenocarcinoma. *Nature* 548, 239–243.
- Nolen, B., Taylor, S., and Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* 15, 661–675.
- Paladino, A., Morra, G., and Colombo, G. (2015). Structural stability and flexibility direct the selection of activating mutations in epidermal growth factor receptor kinase. *J. Chem. Inf. Model.* 55, 1377–1387.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Reményi, A., Good, M.C., Bhattacharyya, R.P., and Lim, W.A. (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the Yeast MAPK Network. *Mol. Cell* 20, 951–962.
- Rivoire, O., Reynolds, K.A., and Ranganathan, R. (2016). Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* 12, e1004817.
- Roybal, K.T., Rupp, L.J., Morsut, L., Walker, W.J., McNally, K.A., Park, J.S., and Lim, W.A. (2016). Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell* 164, 770–779.
- Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289, 1938–1942.
- Songyang, Z., Carraway, K.L., III, Eck, M.J., Harrison, S.C., Feldman, R.A., Mohammadi, M., Schlessinger, J., Hubbard, S.R., Smith, D.P., Eng, C., et al. (1995). Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* 373, 536–539.
- Wan, P.T.C., Garnett, M.J., Roe, S.M., Lee, S., Niculescu-Duvaz, D., Good, V.M., Jones, C.M., Marshall, C.J., Springer, C.J., Barford, D., et al. (2004). Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-Raf. *Cell* 116, 855–867.
- Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* 70, 697–701.
- Xu, B., and Yang, Z. (2013). PAMLX: a graphical user interface for PAML. *Mol. Biol. Evol.* 30, 2723–2724.
- Xue, Y., Wan, P.T., Hillertz, P., Schweikart, F., Zhao, Y., Wissler, L., and Dekker, N. (2013). X-ray structural analysis of tau-tubulin kinase 1 and its interactions with small molecular inhibitors. *ChemMedChem* 8, 1846–1854.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43.
- Yao, Z., Yaeger, R., Rodrik-Outmezguine, V.S., Tao, A., Torres, N.M., Chang, M.T., Drosten, M., Zhao, H., Cecchi, F., Hembrough, T., et al. (2017). Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. *Nature* 548, 234–238.
- Zeqiraj, E., and van Aalten, D.M. (2010). Pseudokinases—remnants of evolution or key allosteric regulators? *Curr. Opin. Struct. Biol.* 20, 772–781.
- Zhang, J., Adrián, F.J., Jahnke, W., Cowan-Jacob, S.W., Li, A.G., Iacob, R.E., Sim, T., Powers, J., Dierks, C., Sun, F., et al. (2010). Targeting bcr-abl by combining allosteric with ATP-binding-site inhibitors. *Nature* 463, 501–506.
- Zheng, J., Trafny, E.A., Knighton, D.R., Xuong, N., Taylor, S.S., Ten Eyck, L.F., and Sowadski, J.M. (1993). 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 49, 362–365.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-GFP (recognizes Venus YFP)	Sigma-Aldrich	SAB4301138
Mouse monoclonal anti-Pgk1	Abcam	ab113687
Experimental Models: Organisms/Strains		
See Table S2 for yeast strains	N/A	N/A
Software and Algorithms		
pySCA	Rivoire et al., 2016	http://reynoldsk.github.io/pySCA/
PAMLX	Xu and Yang, 2013	http://abacus.gene.ucl.ac.uk/software/paml.html
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/download.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Michael B. Yaffe (myaffe@mit.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Yeast Strains

Yeast strains used in this work are described in [Table S6](#). All strains are in the W303 genetic background. Gene deletions were performed by one-step PCR as described ([Longtine et al., 1998](#)). All mutants were integrated into yeast genome as a single copy expressed from their endogenous promoter.

METHOD DETAILS

Plasmids

Plasmids used in this work are described in [Table S5](#).

Site-Directed Mutagenesis

Site-directed mutagenesis was performed with QuickChange according to the manufacturer's directions (Agilent).

Cell Growth and Treatment with α Factor

All cells were grown in synthetic complete media with dextrose (SDC). Three single colonies from each strain bearing the AGA1pr-YFP reporter were inoculated in 1 ml SDC in 2 ml 96-well deep well plates and serially diluted 1:5 three times. Plates were incubated overnight at 30°C. In the morning, cells from the row that had been diluted 1:25 were typically found to have OD₆₀₀ ~0.5. These cells were diluted 1:5 in 4 rows of a 96 well U-bottom micro-titer plate in a total volume of 180 μ l and incubated for 1 hr at 30°C. In each row, cells were treated with different concentrations of α factor: 0, 0.01, 0.1 and 1 μ M (10x stocks of α factor were prepared and 20 μ l were added to 180 μ l cells). Treated cells were incubated for 4 hrs at 30°C before translation was stopped by addition of 50 μ g/ml cycloheximide. Cells were incubated for an additional hr at 30°C to allow time for fluorophores to mature.

Flow Cytometry

The AGA1pr-YFP, HOR2pr-GFP and MLP1pr-GFP reporters were measured by flow cytometry by sampling 10 μ l of each sample using a BD LSRFortessa equipped with a 96-well plate high-throughput sampler. Data were left ungated and FlowJo was used to calculate median YFP/GFP fluorescence. Bar graphs show the average of the median of the three independent colonies that were assayed, and error bars are the standard deviation.

Confocal Microscopy

96 well glass bottom plates were coated with 100 $\mu\text{g/ml}$ concanavalin A in water for 1 hr, washed three times with water and dried at room temperature. 80 μl of cells that had been treated with pheromone at the indicated concentrations for 3 hrs were diluted to $\text{OD}_{600} \sim 0.05$ and added to a coated well. Cells were allowed to settle and attach for 15 minutes, and unattached cells were removed and replaced with 80 μl SDC media. Imaging was performed at the W.M Keck Microscopy Facility at the Whitehead Institute using a Nikon Ti microscope equipped with a 100 \times , 1.49 NA objective lens, an Andor Revolution spinning disc confocal setup and an Andor EMCCD camera. Images were analyzed in ImageJ.

Western Blotting

Total protein was TCA purified from cells as described. 10 μl of each sample was loaded into 4-15% gradient SDS-PAGE gels (Bio-Rad). Gels were run at 25 mA for 45 minutes, and blotted to PVDF membrane at 225 mA for 40 minutes. After 1hr blocking in Li-Cor blocking buffer, membranes were incubated with anti-FLAG primary antibody (SIGMA, F3165) and/or anti-PGK (22C5D8) overnight at 4°C on a platform rotator (all 1:1000 dilutions in blocking buffer). Membranes were washed three times with TBST and probed by anti-mouse or anti-rabbit IR dye-conjugated IgG (Li-Cor, 926-32352, 1:10000 dilution). The fluorescent signal was detected with the Li-Cor/Odysey system.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Coupling Analysis

To perform SCA, we constructed an alignment including the kinase domain for protein kinases of all groups (kinome-wide alignment). Briefly, following alignment processing as detailed in previous work (Halabi et al., 2009; Rivoire et al., 2016), we used the python-based software package (pySCA) to compute a four-dimensional array with conservation-weighted covariance between all possible pairs of positions in the alignment and every possible amino acid residue within this pair of positions. By taking the magnitude (Frobenius norm) of the vector for all the amino-acids for a given pair of positions, this four-dimensional array was subsequently compressed into a two-dimensional coupling matrix. The value at each position in this two-dimensional coupling matrix represents the degree of coupling between the corresponding pair of positions in the domain. Significantly coupled residues are identified by spectral decomposition and comparison to a randomized alignment, where residues within a kinase position are reshuffled thereby maintaining conservation while losing coupling. Next, independent component analysis (ICA) allows the identification of “independent components” (ICs) or clusters of residues that are more coupled amongst themselves than with other residues. Positions contributing to each IC are defined by fitting an empirical statistical distribution to the ICs and selecting positions above a defined default cutoff (>95% of the CDF). For further analysis of these independent components, by using singular value decomposition (SVD) as described in the next section, we can evaluate which specific protein sequences and domain positions contribute the most to a specific independent component (Rivoire et al., 2016). Finally, in cases where extensive coupling still exist between several “independent components,” they are considered as part of the same protein sector, as discussed elsewhere (Rivoire et al., 2016). In the case of the kinase superfamily, the existing coupling between independent components IC_B , IC_C and IC_D led to them being considered part of a single sector, the green sector. Since the main purpose behind the development of Comparative Coupling Analysis (CCA) was to compare coupled residues between kinase groups, when considering the assignment of independent components into protein sectors in group-specific alignments, we aimed to maintain this assignment as similar as possible to the one used for the kinase superfamily.

Singular Value Decomposition and Mapping of Sequence Variations along Independent Components

While a more complete theoretical description of Singular Value Decomposition (SVD) in the context of SCA can be found elsewhere (Rivoire et al., 2016), here we provide a shorter description. Briefly, SVD allows to link co-evolving groups of amino acid residues (such as those forming an independent component, IC, or protein sector) to patterns of sequence divergence in the original alignment. As such, using SVD we can map each protein in the original alignment as a function of its sequence divergence to every other protein in the alignment. Even more, by restricting the mapping to specific ICs the obtained mapping reflects the sequence relationship of each protein to every other protein specifically as it relates to the amino acid residues forming that IC.

Comparative Coupling Analysis

Taking advantage of the seven standard kinase groups as classified on the basis of function, sequence and structural similarity, evolutionary history and broad substrate specificity (AGC, CAMK, CMGC, STE, CK1, TKL and TK) (Manning et al., 2002), we constructed group-specific alignments by restricting each alignment to protein kinases belonging to that group. In order to allow the comparison between groups and with the kinome-wide alignment, the alignments were constructed with Mafft (with its parameters –add and –keeplength) (Katoh and Frith, 2012) using as baseline an original alignment including all human eukaryotic protein kinases. A canonical representative structure for each group was chosen based on completeness of the structure solved and optimizing for the largest number of residues within the kinase domain being covered. The canonical representative structures used were PKC θ for AGC (PDB: 2JED), Pim1 for CAMK (PDB: 4JX3) (Lee et al., 2013), TTBK1 for CK1 (PDB: 4BTJ) (Xue et al., 2013), ERK2 for CMGC (PDB: 4QTE) (Chaikuad et al., 2014), MST2 for STE (PDB: 4LGD) (Ni et al., 2013), BRaf for TKL (PDB: 4MBJ) (Newhouse et al., 2013) and Abl for TK (PDB: 1FPU) (Schindler et al., 2000). Using each canonical representative structure and following the steps

described in the SCA section above, we calculated coupling matrices for each kinase group separately. Once coupling matrices were calculated for the different kinase groups, they were mapped back to the representative structure that covered the largest number of residues within the kinase domain, namely ERK2 (PDB: 4QTE) (Chaikuad et al., 2014). By cross-comparing with the sectors identified in the kinome-wide analysis and other groups, we predicted *in silico* functional similarities and differences between the kinase groups. Finally, we quantified the degree to which a residue predicted to be member of one sector in one kinase group tended to belong to the same sector in other kinase groups and used the median number of groups encoding the same sector as a general measure of compositional conservation. As introduced in the main text, CCA revealed that a fraction of the residues forming a specific independent component and sector in the kinome-wide alignment also formed similar independent components and sectors in other subgroups. In these cases, the same annotation scheme that had been used for the kinome-wide results (red, green and blue sectors) was used for the group-specific results.

Estimation of Negative Selection (Omega Estimates)

Using the YN00 program that is part of the PAMLX package (Yang, 2007 and PAML) with default parameters we estimated the number of synonymous and nonsynonymous substitutions for the different residues to be considered, while correcting for multiple substitutions, transition/transversion rate biases and base/codon frequency biases (Yang and Nielsen, 2000). These omega estimates are a measure of the amount of negative or positive selection that a specific protein or protein segment has gone through, with distributions around 1.0 indicating similar degrees of positive and negative selection and distributions below 1.0 indicating stronger negative selection). For our purposes, after obtaining cDNA for all CMGC kinases from KinBase (kinase.com/kinbase), we constructed a cDNA alignment from the CMGC-specific alignment, allowing us to map back the sector sites that each codon corresponds to for a large number of sites, and computed omega estimates for the three different sectors as well as for non-sector sites.

Residue Conservation

The conservation of amino acid residues independently of other positions is here measured by the Kullback-Leibler relative entropy, D_i . This measure compares the observed amino acid residue at a position to the background frequency of this amino acid from a non-redundant database of protein sequences.

Calculation of Area Covered by Substrate Peptide or Kinase Regulator in Kinase Structures

The solvent exposure of every residue in the kinase domain is calculated using the UCSF Chimera package (Pettersen et al., 2004) in the presence and absence of the substrate peptide or kinase regulator. In-house python scripts subsequently compare the solvent exposure calculated for both situations and calculate the area that is buried by the peptide or regulator in residues that form the red, green or blue sectors.

Construction of Protein Structure Network

Protein structure networks are constructed by considering the strength noncovalent interactions between sidechains, as evaluated from the normalized number of contacts between residues (Kannan and Vishveshwara, 1999). The details of the construction of such a network at a particular interaction cut-off (I_{\min}) and its implications have been previously described (Bhattacharyya and Vishveshwara, 2011; Bhattacharyya et al., 2016). Typically, an I_{\min} value between 2 - 3% is chosen to compute the interaction networks as discussed earlier (Bhattacharyya et al., 2013; Ghosh and Vishveshwara, 2008; Reményi et al., 2005). Here, the construction of protein structure network and subsequent analysis is carried out using the PSN-Ensemble software (Bhattacharyya et al., 2013). The high-resolution crystal structures of Fus3 in the apo state (PDB: 2B9F) or in the Ste5 (PDB: 2F49) or Ste7 (PDB: 2B9H) bound state were used to construct the corresponding protein structure networks for analysis.

Network Parameters

Various network parameters are used to examine the topological features of the protein structure network. A brief definition of the parameters used in this study and their physical significance is described below.

Cliques and Communities

A k -clique (also referred to as clique) is defined as a set of k nodes (amino acid residues), in which every node is connected to every other node. A k -clique-community (also referred to as clique-community) is an assemblage of k -cliques, sharing either $k-1$ or $k-2$ nodes (Figure S5C). Cliques and communities are percolating interaction units that indicate highly connected structural features within the protein structure network. These parameters are ideal for querying ligand induced conformation rearrangements in proteins (Ghosh and Vishveshwara, 2008; Adamcsek et al., 2006; Palla et al., 2005).

Junction Residues

Dijkstra's algorithm is used to compute the shortest paths (SPs) of communication in the structure network between all pairs of residues in two chosen domains, the Ste5-FBD and the Fus3 active site (residues 155-192) in this case. The residues which appear in more than 10% of these SPs are termed as junction nodes and are considered to be key mediators of allosteric communication between these domains in the protein (Bhattacharyya et al., 2013; Bhattacharyya et al., 2016).

Measuring Similarity between Kinase Groups

Pairwise kinase group similarity was measured by calculating residue-normalized BLOSUM distances for every residue within the kinase domain as described elsewhere (Creixell et al., 2015). The coupling difference between two kinase groups is calculated by measuring the LogWorth, $-\log_{10}(\text{p-value})$, where the p-value is calculated from a hypergeometric test comparing the number of shared sector sites given the size of the sectors in both groups. As a result from these calculations, higher LogWorth values correspond to higher coupling similarity between kinase groups. After identifying AGC-CMGC as the pair of kinase most divergent in their coupling given their kinase similarity, their divergence is further inspected using SVD and other standard methods previously described (Halabi et al., 2009; Rivoire et al., 2016).

Mapping of Somatic Cancer Mutations

Genomic coordinates (human genome version GRCh38.p7) for missense cancer somatic point-mutations were retrieved from COSMIC v79 (Forbes et al., 2011), and they were mapped to ENSEMBL canonical proteins, predicting the variants functional effect with the standalone perl script of the Ensembl Variant Effect Predictor, v87.18 (McLaren et al., 2016). A total of 1,515,599 of cancer somatic mutations were mapped to a canonical protein. To obtain, for all protein kinases, the kinase domain residues perturbed by somatic cancer mutations, all the variants that mapped to the kinase domain were aggregated by kinase residues. Only ENSEMBL canonical proteins with a 100% identical kinase domain sequence, with respect to the corresponding kinase domain sequence reported in KinBase, were considered further.

In order to define sectors for all protein kinases of all groups, using the kinome-wide alignment, the sector sites identified in the group representative kinases were mapped to the corresponding residues of all the other kinases within the groups. The kinase domain residues that in the kinome-wide alignment did not map to any residues of the corresponding group representatives, i.e. sequence insertions, due to the uncertainty in sector association, were excluded from the analysis. A total of 14,860 mutations were mapped to 13,152 sites within a kinase domain. The mutation percentage was calculated across all kinases, for all sectors, as the number of residues perturbed by somatic cancer mutations, divided by the number of residues in the sector. Wilcoxon signed-rank tests were performed to assess the significance of the difference, across all kinases, between the mutation percentage in the red sector compared to the blue, the green and the non-sector. Definitions for OG and TSG were obtained from a work reviewing the functional role of the kinome in cancer (Fleuren et al., 2016).