Dynamic Architecture of a Protein Kinase

Christopher L. McClendon, Alexandr P. Kornev, Michael K. Gilson, Susan S. Taylor

Supporting Information

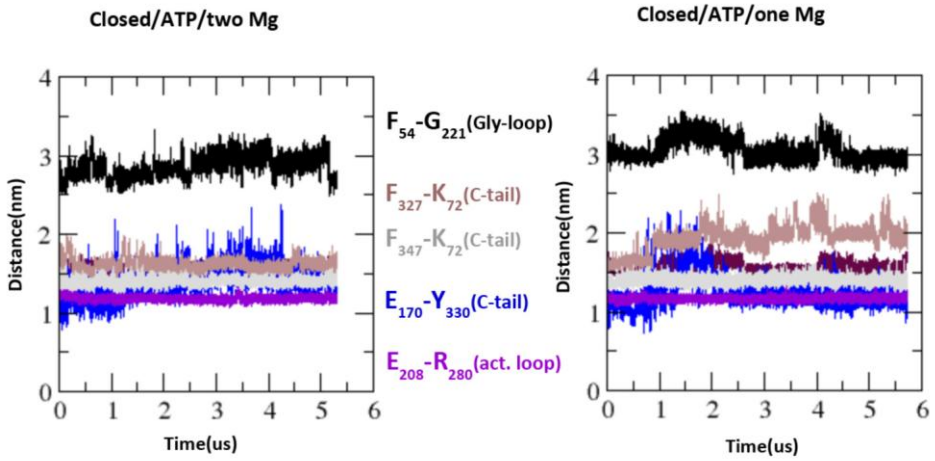

Figure S1. Selected inter-residue Cα distances show stability of the activation loop, Mg^{2+}-dependent flexibility of different regions of the C-tail, and opening/closing transitions of the Gly-rich loop in both liganded conditions, with one Mg^{2+} giving longer and more pronounced openings.

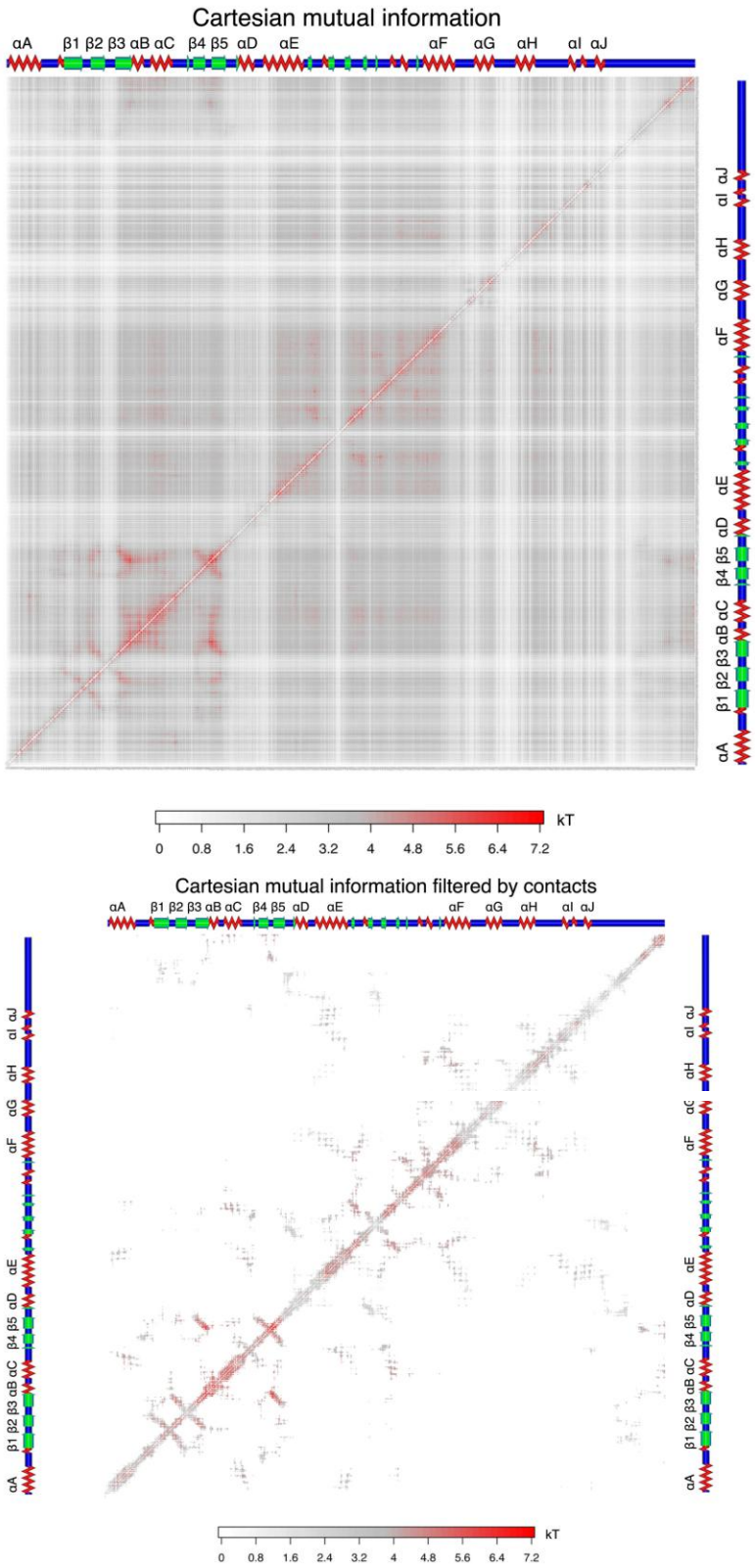

Figure S2. Heat plot of mutual informations between residues of PKA in the closed/ATP/twoMg state. Above: unfiltered data. Below: data filtered to include couplings only between atoms within 10Å of each other for 75% of the simulation.

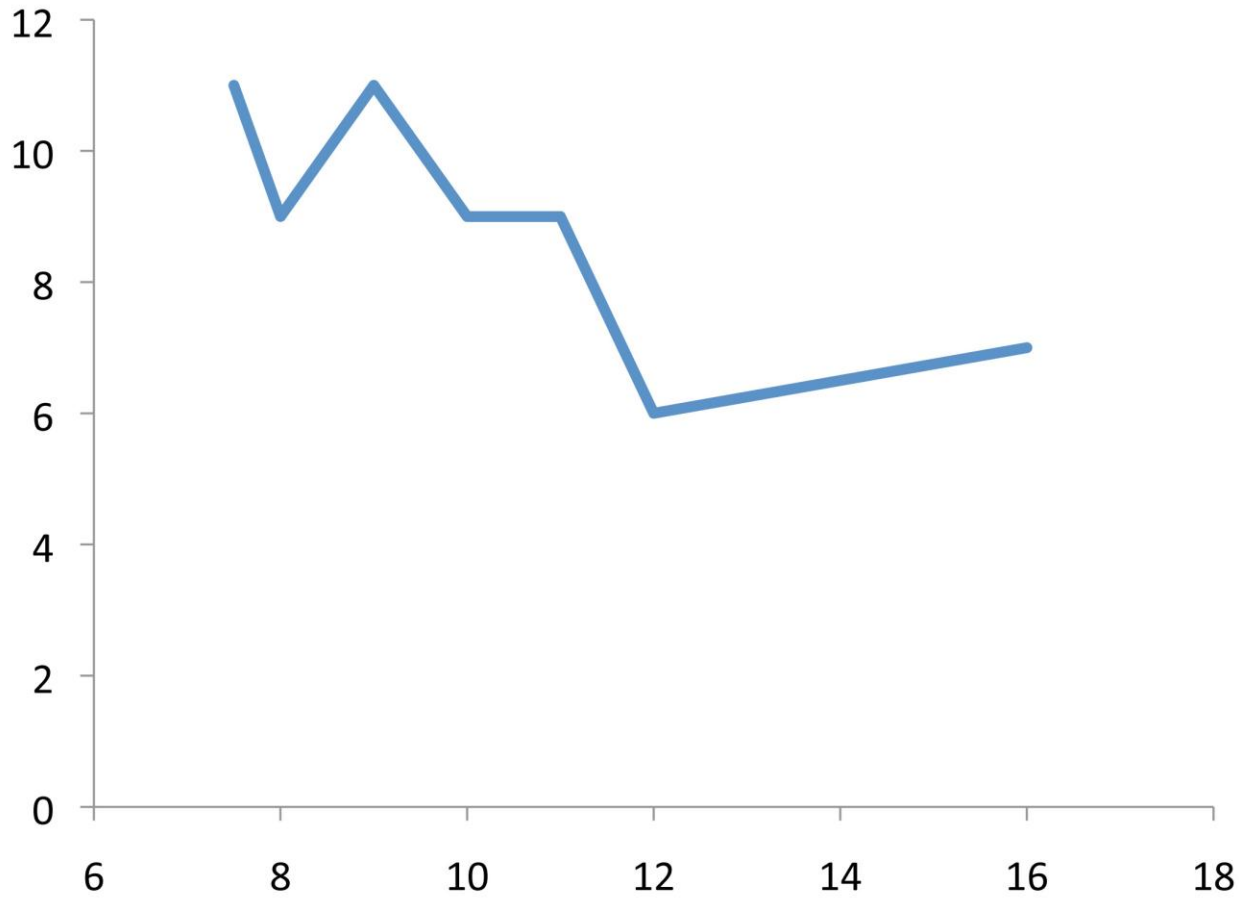

Figure S3. Number of communities in the active form of PKA as a function of the residue-residue cutoff distance used to prune the initial mutual information network. See text for details.

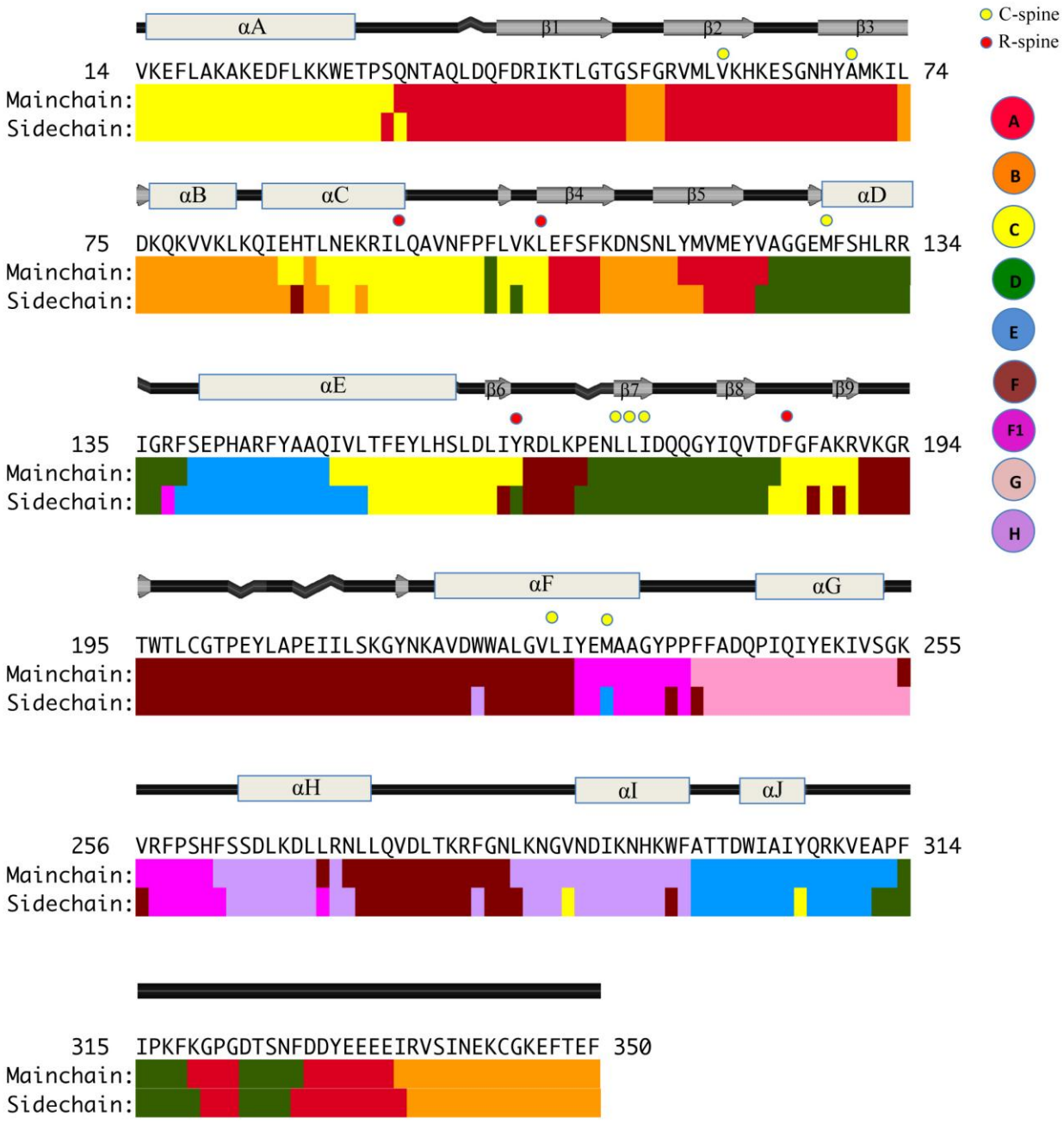

Figure S4. Community memberships of main chain and side chain elements of PKA. Bars below the primary sequence are colored according to community, and colored dots above the sequence indicate the C-spine (yellow) and R-spine (red) residues. The secondary structure of PKA is shown for reference. In a number of locations, the mainchain and sidechain belong to different communities; these we denote as bridging residues. For glycine, the carbonyl oxygen is used in lieu of a side chain.

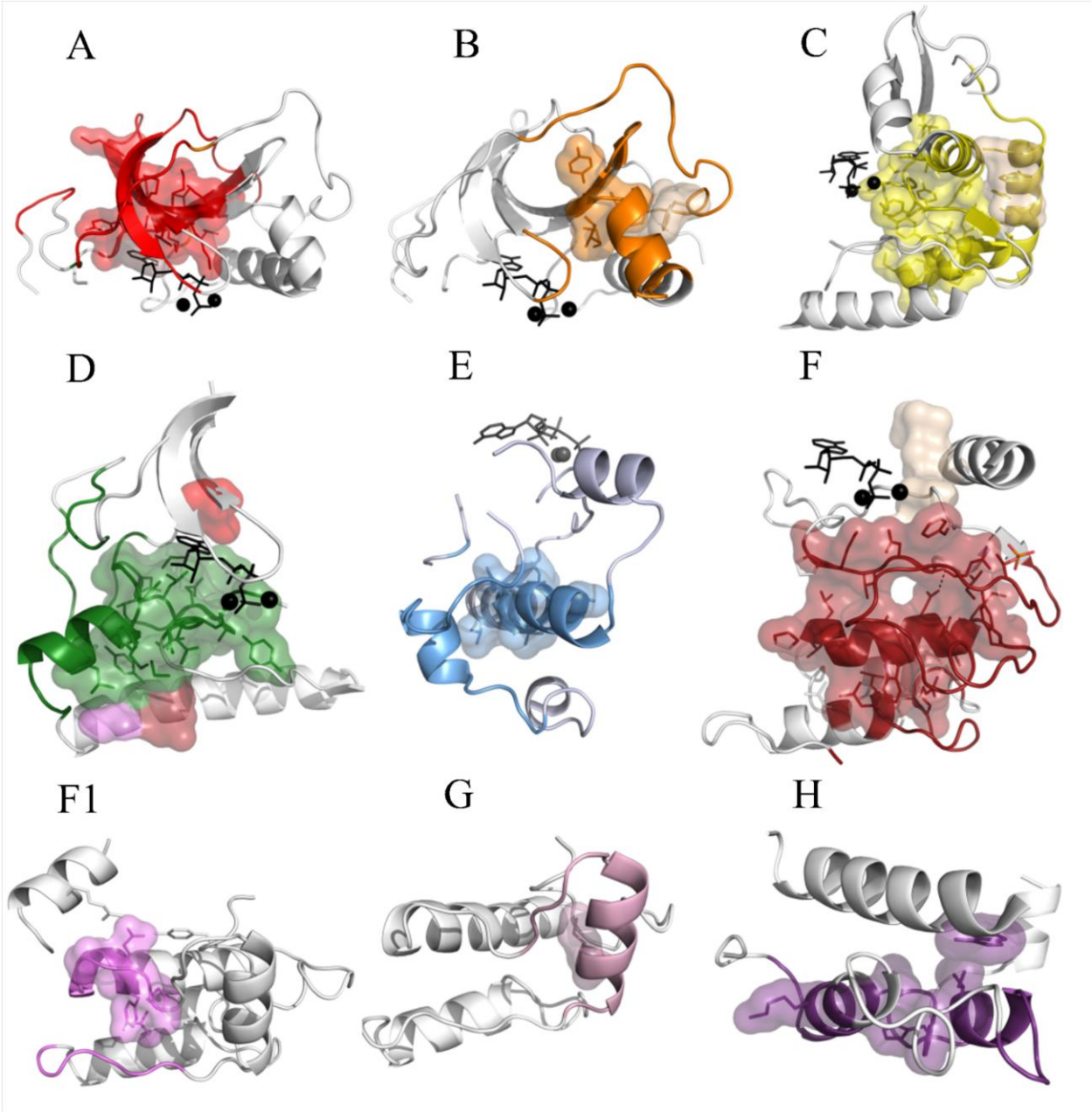

Figure S5. Community-forming residues (transparent surfaces) form central, connected networks of contacts that comprise a mostly hydrophobic core for each community. Communities are colored on a gray protein background, and community-forming residues are shown in surface representation with side chains depicted as sticks. Additional residues are shown in some cases as indicated below. (A) Community-forming residues in ComA comprise the center of the β-sheet. (B) The FxxF motif on the hydrophobic tail in ComB packs against community-forming residues at the tip of the β-sheet. (C) A cation-π stack, involving W30 in the αA-helix, packs against community-forming residues in ComC, which support the R-spine and αC-helix. The R-spine and αC-helix provide a hydrophobic core connecting the top of the αE helix, the activation loop, and the αA and αC helices. (D) Community-forming residues in ComD (green) comprise most of the C-spine, form the bottom of the ATP-binding site, and connect the C-spine to the R-spine at Y164. Other C-spine residues from ComA, ComF, and ComF1 are shown in surface representation and colored by community for reference. (E) Community-forming residues in ComE radiate out as spokes from the

αE-helix. (F) ComF contains a large number of community-forming residues that connect the activation loop and activation segment to the R-spine and F-helix, down to the AGC insert through a conserved salt bridge between E208-R280, both of which are community-forming residues. (F1) Community-forming residues in ComF1 connect the substrate-binding hotspot residue E230 to a relatively stable part of the αF-αG linker. (G) ComG has a single community-forming residue in the middle of the αG-helix. (H) Community-forming residues in ComH connect the αH-αI helices, and include the sidechain of W221 in the F-helix, which connects the αF-helix to the αH-helix through I291.

Annotation of PKA communities

Here we provide a deeper analysis of each community in the closed/ATP/two Mg^{2+} reference state, along with pertinent literature data and structural observations that support the proposed functional annotation of each community.

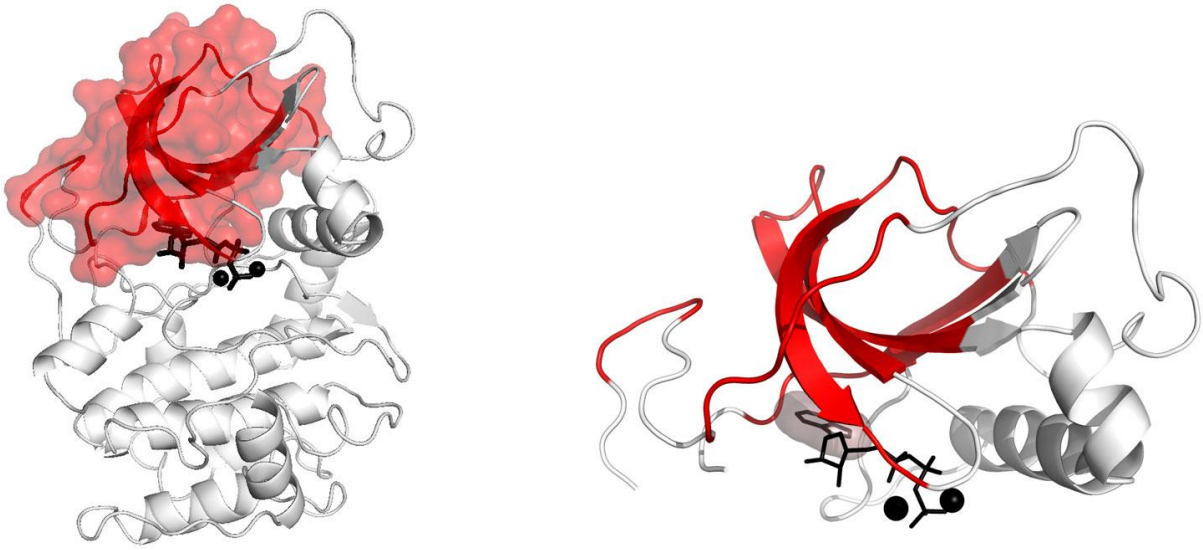

Fig. S6. ComA from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComA: ATP binding, assists substrate binding

ComA includes part of the N-lobe and part of the C-terminal tail, and sits atop the adenine ring of ATP. It contains L106 of the R-spine and recently-discovered R-spine "shell" residues V104, M118, and M120, whose hydrophobic nature has been shown to support the N-lobe part of the R-spine(1). ComA also includes V57 of the C-spine, which serves as the top of the sandwich between ComA, the adenine ring of ATP, and ComD. Also in this community are the FDDY motif of the C-terminal tail and the acidic patch that follows it; a number of these residues are critical for ATP binding or activity, particularly F327A, Y330A, and E333A(2). When F327 contacts the adenine ring of ATP, it closes one side of the ATP-binding pocket, constrains the remainder of the C-tail to engage with kinase core, and orients the Gly-rich loop for closure over the ATP to exclude solvent and promote

phosphotransfer(3). Meanwhile, Y330 coordinates a conserved water molecule as well as the P+1 Arg of substrates(4). Additionally, the E333A mutation in the acidic patch decreases Kemptide binding; the native E residue may function as part of an electrostatic docking site for this molecule. A mutation in yeast PKA corresponding to K47A showed a significant reduction in activity(5), presumably as this residue helps tether the N-lobe to the C-terminal tail through an electrostatic interaction with Asp329 in the FDDY motif of the C-terminal tail. This community also contains the sharp turn in the bulge of the C-terminal tail, which is structurally conserved in active structures of AGC kinases and presumably is important for its conformational flexibility or proper assembly of the end of the C-tail (Figure 1).

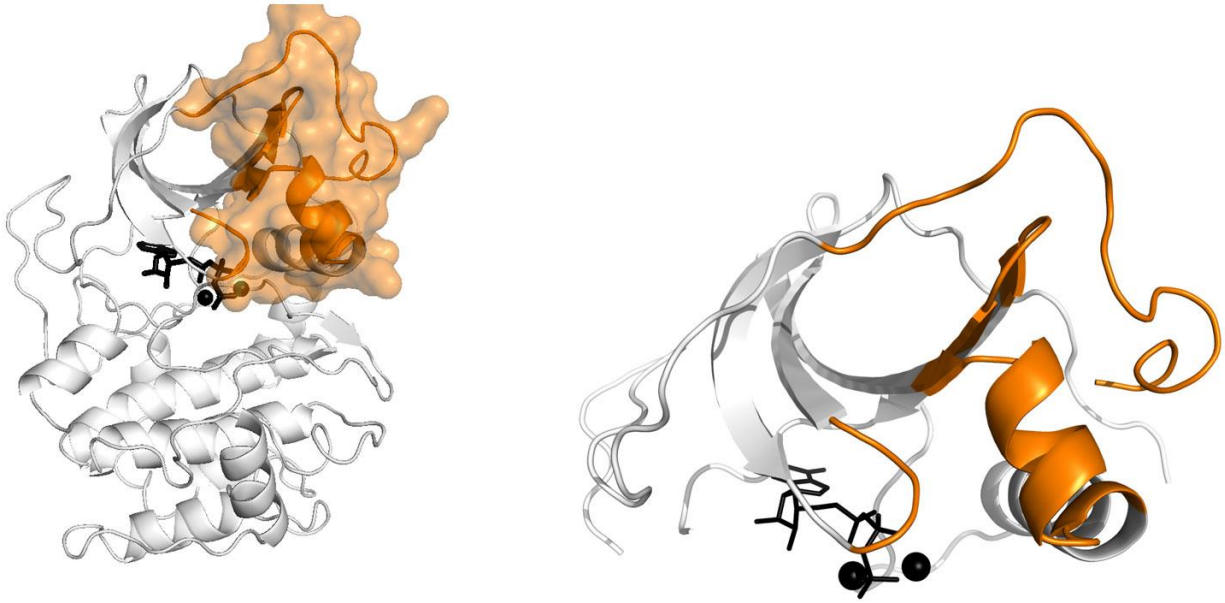

Fig. S7. ComB from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComB: Regulates position of the αC-helix

ComB contains part of the N-lobe, including the αB-helix and the tip of the Gly-rich loop; and the C-terminal part of the C-terminal tail, including the hydrophobic motif. Importantly, F54 of the Gly-rich loop makes hydrophobic contacts with L82 in the αB-helix, and, in the fully closed state, with Q84 in the αC-helix, which in turn hydrogen bonds to H87 in ComF. The hydrophobic motif at the end of the C-tail, conserved in the AGC kinase subfamily, fits into a pocket that is generally conserved among EPK structures(6), and is often an allosteric site where binding of peptide or a protein domain (i.e. Cdk2 or EGFR kinase) occurs and is transduced to adjust or stabilize the position of the αC-helix. In PDK1, small molecule activators and inhibitors have been discovered to bind to this same site(7); activators push the αC-helix towards the active site while inhibitors pull it away from the active site. ComB contains the C-terminal part of the C-terminal tail, whose Ala mutants at

K345A and E346A substantially raise the Km for Kemptide binding while mutations at F347 and F350 severely lower the catalytic Vmax(2). In the context of the communities, mutations at positions 345, 346, 347, and 350 can be understood to act at a distance to affect substrate binding through the contact between ComB and the C-terminal His of pseudosubstrate PKI, or to affect catalysis by adjusting ComC.

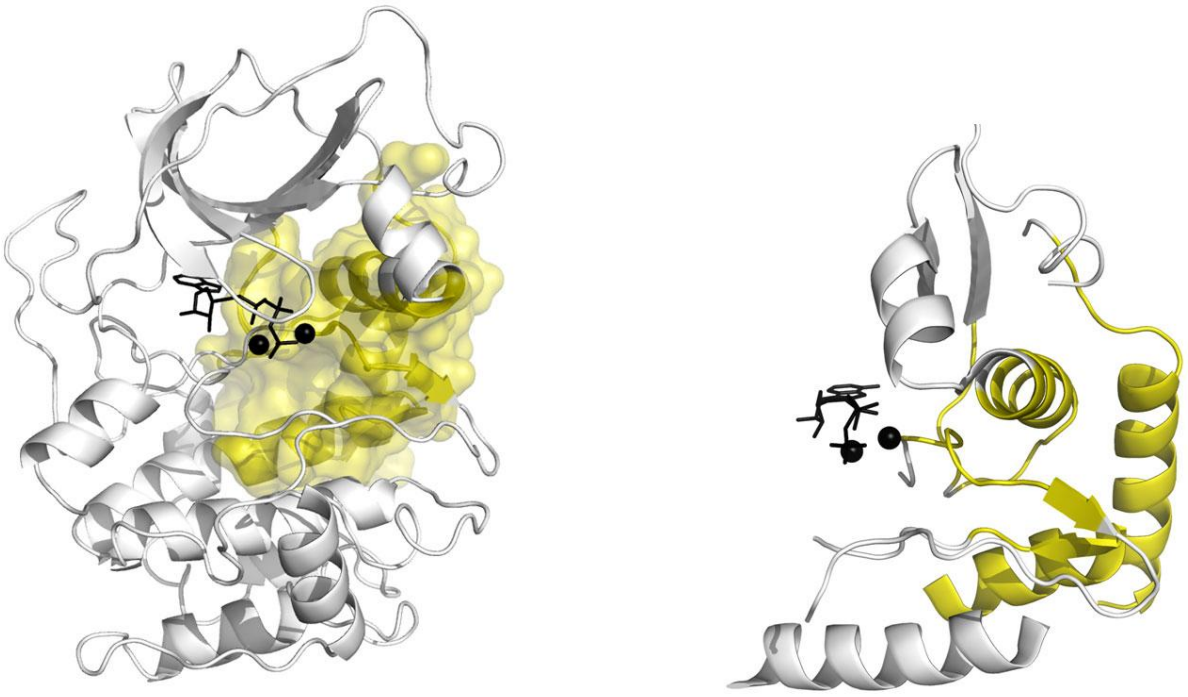

Fig. S8. ComC from the closed/ATP/two Mg^{2+} reference state, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComC: Regulatory, assembly of R-Spine

ComC includes most of the αA- and αC-helices, as well as the DFG-loop, the αA-β1 linker, most of the αC-β4 loop, and part of the αE-helix. Surprisingly, the αC-helix belongs to a separate community from the rest of the N-Lobe. Among EPKs, the αA-helix is unique to PKA, and it serves to stabilize the position of the αC-helix through W30. The position of the αC-helix is critical for kinase activity, as the salt bridge between E91 and K72(ComA) is important for activity(5). In numerous inactive structures of kinases, the position and orientation of the αC-helix are altered(1). ComC also contains the side-chain of D184 of the DFG loop, which coordinates an Mg^{2+} ion to ATP and is also conserved and important for activity. The DFG-loop of a number of other kinases (such as those that bind Type II inhibitors including Gleevec) transitions between DFG-in and DFG-out conformations; the DFG-out conformation often provides more selectivity for inhibitors than the active conformation. When PKA is phosphorylated on the activation loop, no DFG-out conformation has been observed, and the activation loop conformation is very stable, so that the activation loop phosphate is resistant to phosphatases. ComC also contains two

residues of the R-spine, L95 and F185, which help to provide a connection between the N-lobe and the C-lobe. Mutation of L95G/M120A or F185A yielded an inactive kinase(1). Another connection between the N-lobe and the C-lobe comes through the αC-β4 loop, which contains mostly residues from ComA. ComC also contains F102 of ComE, the side chain of V104 in ComD, and L106 in ComA.

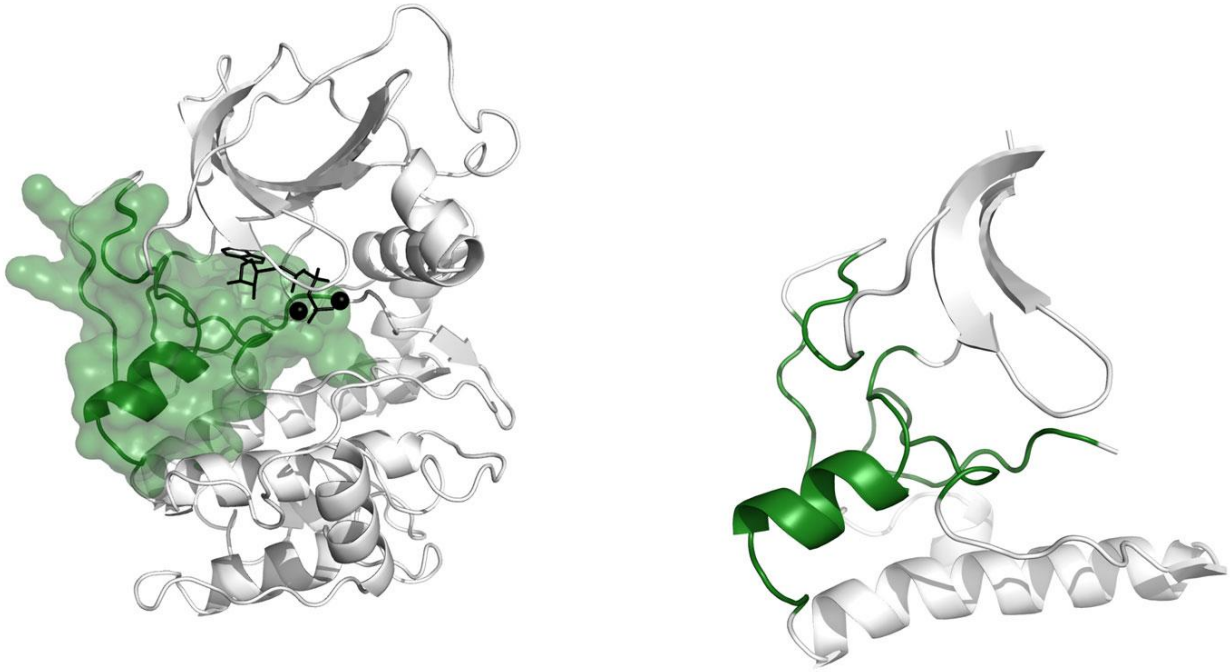

Fig. S9. ComD from the closed/ATP/two Mg^{2+} reference state, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComD: Catalytic, ATP-binding, assembly of the C-spine

ComD includes the bottom of the ATP-binding site, the bulge of the C-terminal tail, and the D-helix. A set of community-forming residues (Fig. 3D) comprise most of the C-spine, which connects the N-lobe and the C-lobe and is important for ATP-binding and catalytic activity. ComD also provides part of the catalytic loop, a number of ATP-binding site residues from below, as well as most of the active site tether of the C-terminal tail before the FDDY motif. Importantly, it contains N171, which coordinates a critical Mg^{2+} ion required for ATP binding even in low [Mg^{2+}] conditions.

ComD contains a part of the catalytic loop, including E170 and N171. The E170A mutation in yeast PKA showed substantial reduction in activity(5), and N171 is critical for coordination of the tighter-binding Mg^{2+} ion in the complex with ATP. The side chain of V104 from the αC-β4 loop defines a hydrophobic portion of the adenine ring binding pocket, and the V104G mutant has an activity only 10% of wildtype(1). ComD also contains a number of C-spine residues that structurally connect the adenine ring of ATP to the E- and F-helices. The side chain of Y146 of the E-helix connects ComD to ComE, providing hydrophobic support to the C-spine.

One significant mystery in AGC kinases is the role of the structurally conserved bulge in the C-terminal tail. It is in the active site tether region of the C-tail, but the bulge itself is outside the active site and does not interact with ATP. The function of this bulge can now be understood in part through the community map. Thus, ComD controls the position of this section of the C-terminal tail, which is involved in ATP-binding and is coupled to the ATP-binding and substrate-binding FDDY motif of ComA. Mutations K317A, K319A in the bulge decrease ATP binding (i.e., increase Km) by > 18-fold, while D323A is inactive(2). To keep the C-terminal tail closed over the N-lobe and plugging the ATP-binding site through F327, the active site tether must be in a loop and not be extended; D323 must apparently make an interaction that stabilizes the closed conformation, likely through a hydrogen bond to S325.

ComD also contributes to substrate binding. In yeast PKA, mutation of E168A/E171A showed only 2.3% of wildtype activity(5); in murine PKA these residues correspond to A124 and E127A in the β5-α4 linker (also called the hinge). In structures with bound substrates, E127A coordinates the P-3 Arg, though not as closely as a bidentate salt bridge.

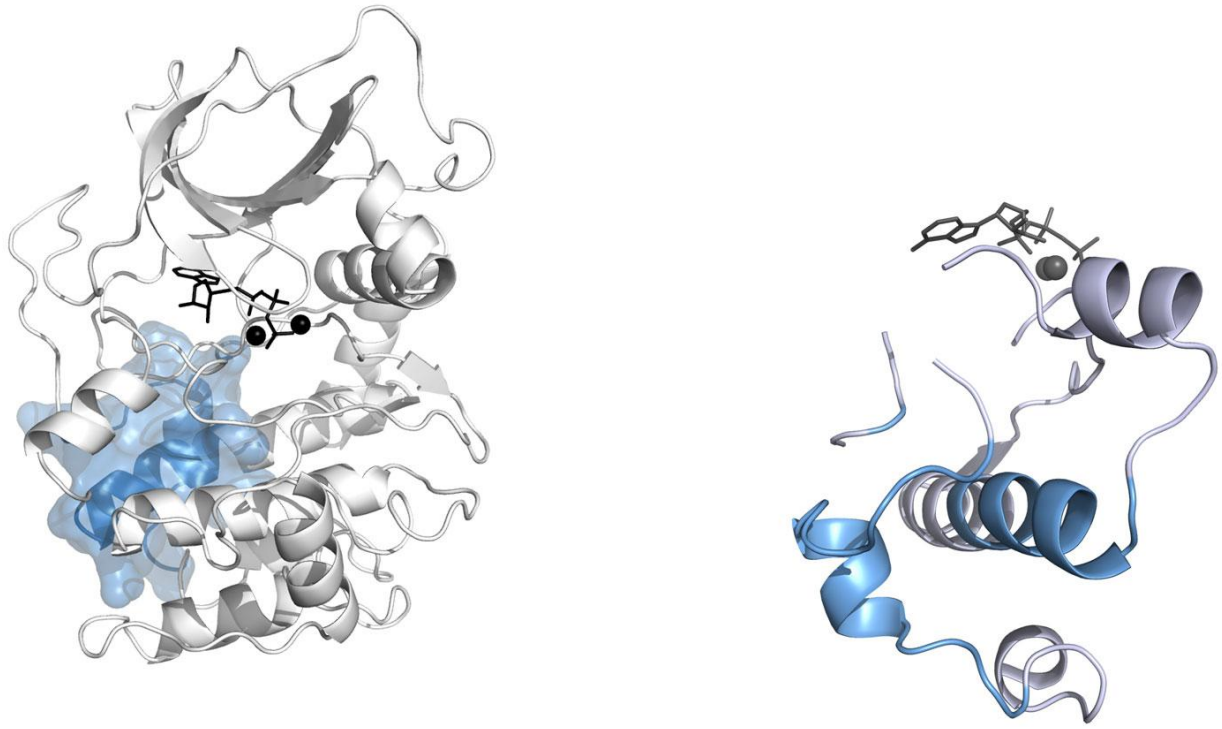

Fig. S10. ComE from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComE: Stabilizes C-spine

ComE contains part of the E-helix and the entire J-helix. Community-forming residues in ComE include A143 and residues 146-152 (except the side-chain of 146, which

belongs to ComD, and the main chains of 150-152, which belong to ComC). The E-helix but not the F-helix is present in Eukaryotic-like kinases(8) such as ChaK and PI3K, and it provides a stable structural support for the bottom of the C-spine. The J-helix anchors αC-β4 and is conserved in AGC kinases, but not in others(9). W302 extends to the myristoyl pocket, and F102 in αC-β4 loop interact with R308 in J-helix, which anchors αC-β4(9).

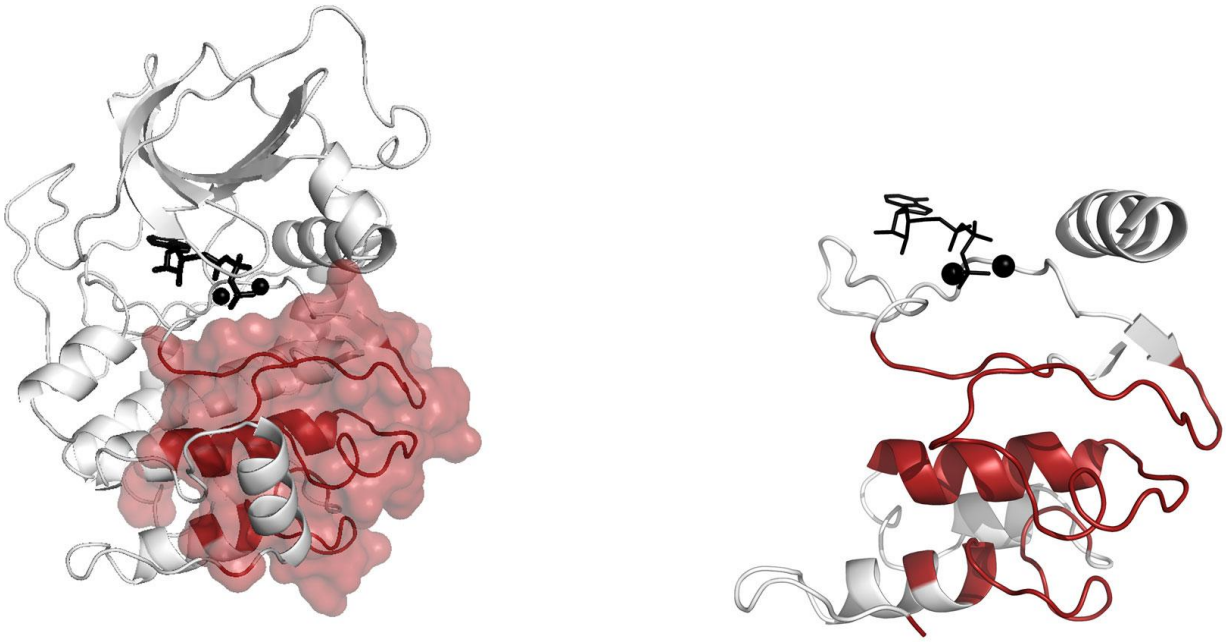

Fig. S11. Com F from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComF: Activation and substrate binding

ComF includes the activation loop, which is important for regulation of protein kinases, most often by phosphorylation; and most of the F-helix, which provides a helical scaffold for the R-spine (Fig. 7). A set of community-forming residues (red and yellow surface in Fig. 7) connects the αC-helix, activation loop, F-helix, and H-helix, with phospho-T197 (yellow in Fig. 7) as an important organizing center. A number of mutants in this community show decreased activity. Thus, a charged to alanine single mutants in yeast PKA have activity less than 15% of widltype(5). These mutants correspond to the following PKA residues in ComF: R165A, D166A, K168A, D184A, E208A, R280A. Importantly, D116A and K168 are in the catalytic loop (165-171). The E203A mutant has decreased substrate binding, and the Y204A mutant (APE motif) displays substantially reduced substrate binding, increased dynamics by H/D exchange, and thermodynamically decoupling of ATP binding from substrate binding. The E208/R280 salt bridge is extremely stable, and E208A or R280A mutations increase flexibility of the C-lobe by H/D exchange, and decrease activity.

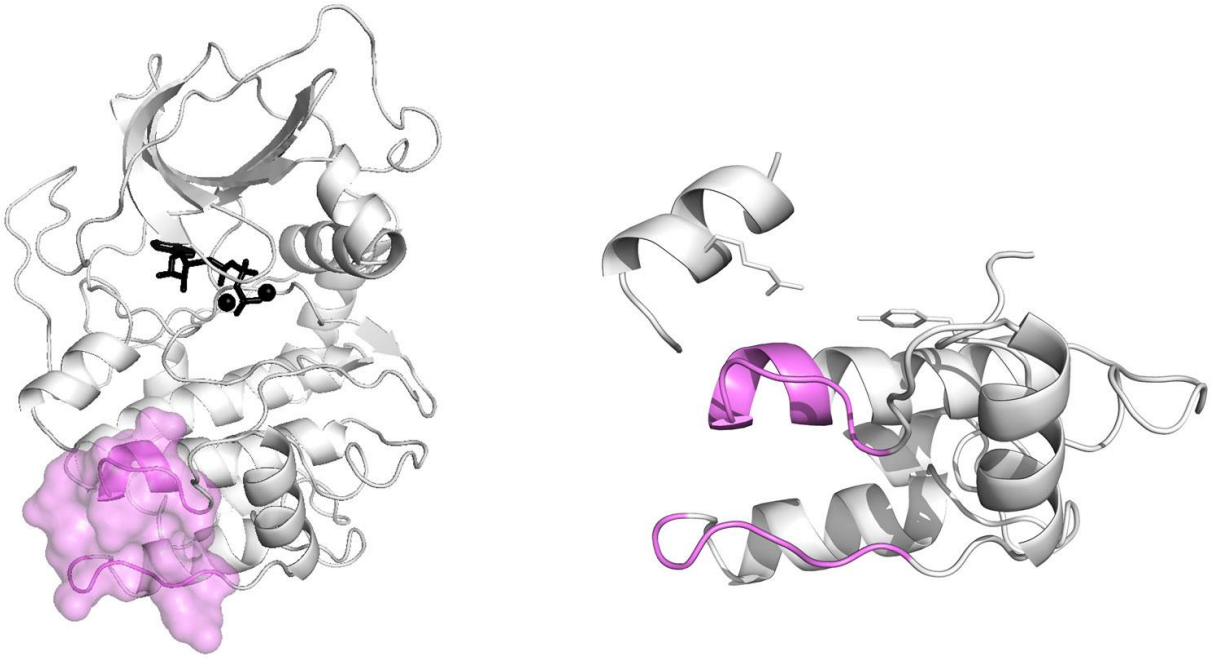

Fig. S12. ComF1 from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white. ComF1 presents an electrostatic hotspot for substrate binding and forms part of an important electrostatic node along with R133 from ComD and Y204 from ComF.

ComF1: Substrate binding

ComF1 includes E230, which is near the base of the C-spine and forms a salt bridge with Arg on substrates. Its name, F1, derives from our observation that, in the non-active PKA structures simulated here, nearly the entire F-helix resides in ComF, and ComF1 is highly coupled to ComF in the closed/ATP/twoMg^{2+} system. In yeast PKA, E230A has activity less than 15% of wildtype. E230Q forms a stable open, apo conformation(10) and has severely reduced substrate binding(11). E230 is thus a hotspot for substrate binding, and is held in place by Y204 on ComF and through the C-spine, as mutation of Y204A in ComF significantly reduces substrate binding.

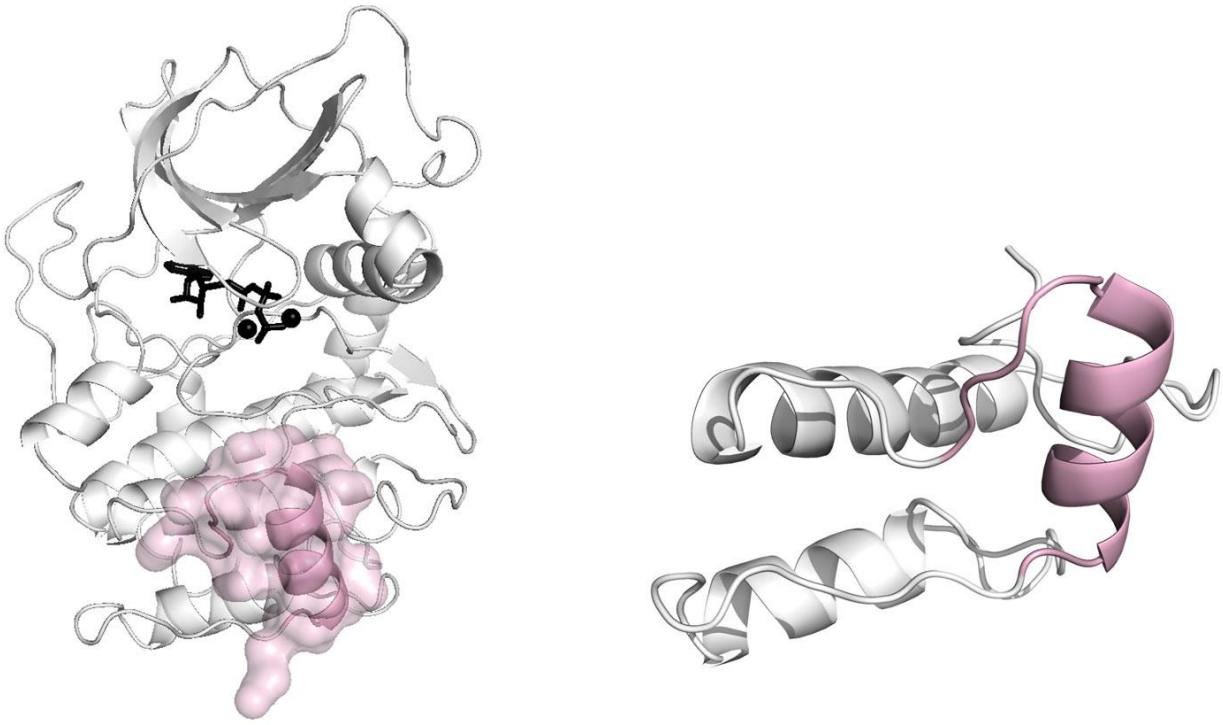

Fig. S13. ComG from the closed/ATP/two Mg^{2+} reference system, shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComG: Regulatory subunit and substrate binding

ComG encompasses the G-helix, which was observed to move towards and away from the body of the C-lobe during our simulations. The G-helix packs against the first cyclic nucleotide binding domain of the RIIβ regulatory subunit in the RIIβ-PKA tetrametic holoenzyme(12). Particularly high root-mean-square backbone flexibilities are observed in the turn before the G-helix and in the top of the G-helix (Figure 7). The G-helix also interacts with the P-1 Arg of substrate, especially through Q242 and P243.

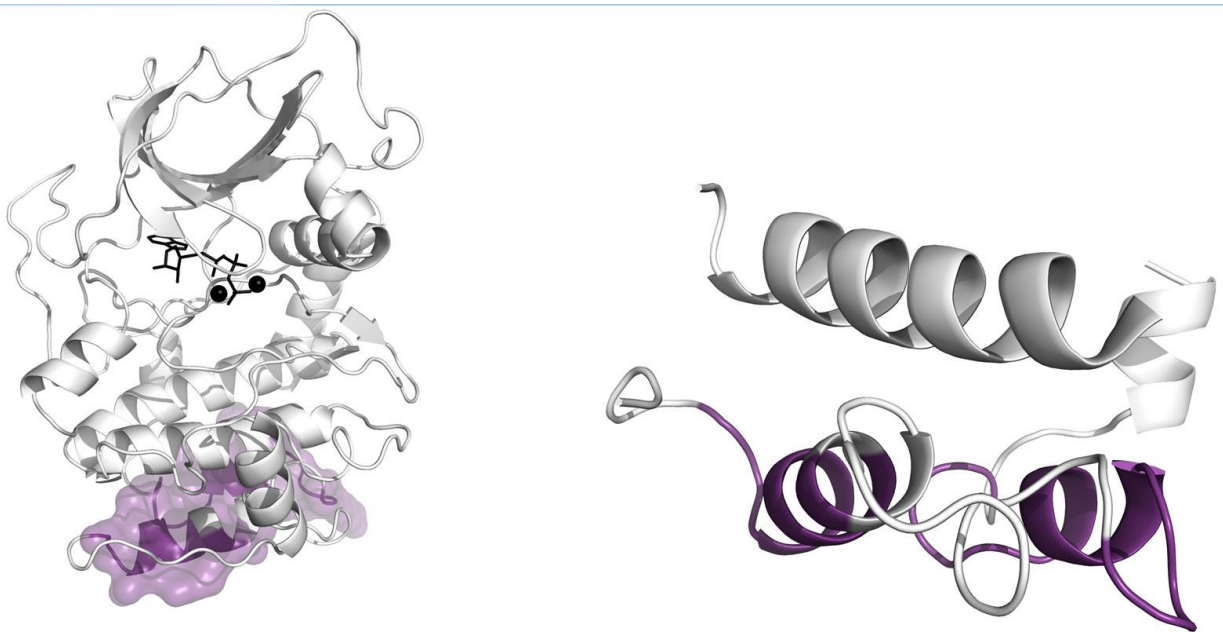

Fig. S14. ComH from the closed/ATP/two Mg^{2+} reference system. shown in transparent surface (left) and cartoon (right) representations, with the remainder of PKA shown in white.

ComH: Docking site

ComH contains residues near the R-subunit binding site. These exhibit slow dynamics in NMR on a different timescale than opening/closing(13). Deminoff et. al identified a 33-residue minimal fragment of yeast PKA (Tpk1) that contains a number of residues from ComH and binds to substrate Cdc25(14), defining this region as an allosteric substrate docking site. Minimal fragments of Cdc25 and Yak1 also bind to Tpk1.

Alternative clustering of the filtered mutual information matrix

To see if an entirely different approach to clustering the mutual information matrix might yield a comparable number of communities, we also performed hierarchical clustering of the mutual information matrix as in ref. 31 in the main text. We took the Cartesian mutual information matrix filtered at 10 Angstroms (i.e. the matrix that was provided to the Girvan-Newman clustering algorithm to generate the community map, Fig. S2, right). We then performed hierarchical clustering using a Euclidean metric, as in ref. 31 in the main text. Hierarchical clustering does not create sharp boundaries between clusters, so instead we cut the dedrogram horizontally using the blue horizontal line. This cut of the dendrogram yields eight clusters after manual merging of two highly-coupled clusters at the bottom right, with seven well-defined clusters and the eighth cluster containing the remainder of the data. Clusters are demarcated by blue vertical lines. Manually identifying an additional cluster in this dendrogram by inspection using a slightly lower cut yields another cluster for a total of nine clusters, with eight of them well-defined. And while these

clusters from the dendrogram may not necessarily correspond to the same groupings present in the community map, as the clustering method is completely different, these extremely different clustering methods can yield a similar number of clusters, suggesting that our community map with nine communities has a reasonable number of clusters.

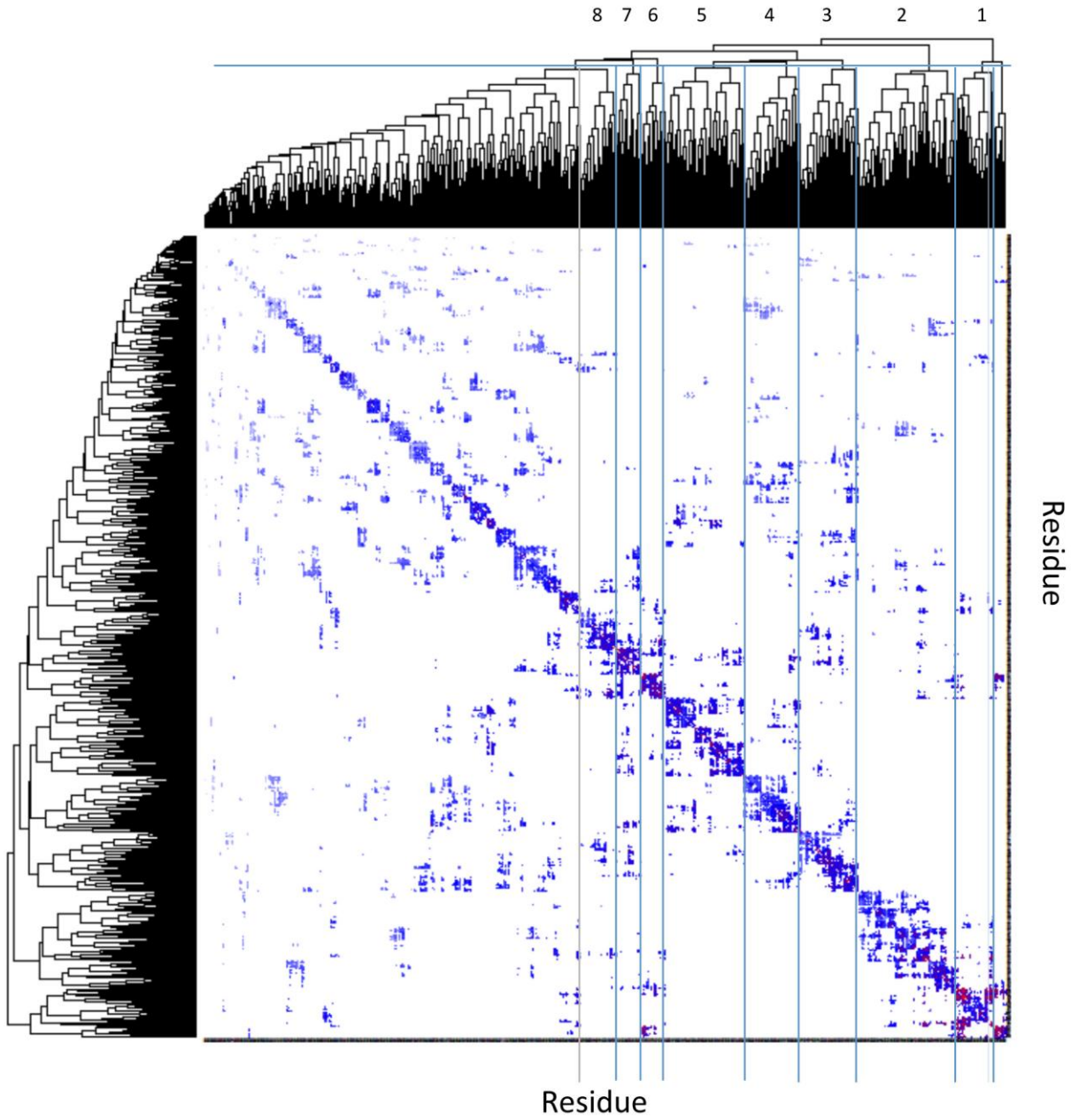

Fig. S15. Hierarchical clustering of the filtered mutual information matrix using a Euclidean distance metric yields eight well-defined and well-separated clusters, indicated by cuts to the dendrogram tree (blue horizontal line indicates cut level, blue vertical lines indicate cluster groupings). The strong intra-cluster couplings in the heatmap support the clustering, and suggest that at least eight strongly-connected groupings are present, and that the number of communities found in our community map (nine) is not an

unreasonable number given the connectedness of the matrix elements under the hierarchical clustering shown here.

Table SI. Communities in the active form of PKA. Name: name of community. N_{MC}: number of residue backbones in the community. N_{SC}: number of residue side-chains in the community.

Name	**N_{MC}**	**N_{SC}**	**Main Function(s)**
ComA	60	50	ATP binding, assists substrate binding
ComB	58	58	Regulates position of the αC-helix
ComC	38	41	Regulatory, assembly of R-Spine
ComD	58	62	Catalytic, ATP binding, and assembly of C-spine
ComE	28	29	Stabilizes C-spine
ComF	40	44	Promotes activity, activation
ComF1	15	15	Substrate binding
ComG	23	22	Regulatory subunit and substrate binding
ComH	16	15	Protein-protein interaction site

Table S2. Relative edge weights between communities provide a measure of coupling and connectivity between each pair of communities. Raw weighted edge betweenness values shown are multiplied by 10 for ease of comparison.

Community Pair		closed/ATP/two Mg	closed/ATP/one Mg	open/ATP/one Mg	open/apo
A	A1			0.32	0.33
A	B	1.41	1.23	1.19	1.65
A	D	1.79	0.10	0.93	0.04
A	D1		0.48	0.27	0.00
B	A1				0.47
B	F	0.57		0.003	
C	A	0.95	2.11	1.19	1.34
C	B	1.49	1.52	2.10	2.79
C	D	0.78	0.63	2.00	
C	D1		0.25		
C	E	0.95	1.66		2.01
C	F	2.33	2.57	2.30	2.40
C	H	0.27		0.05	
D	D1		0.65	0.43	0.31
D	F1	0.08		0.85	
D	H			0.48	
D1	A1			0.22	0.13
D1	E		0.34		
E	D	0.86	0.25		1.06
E	F1	0.01			
E	H	0.45	0.17		
F	D	0.94	0.81	1.03	1.01
F	E	0.38	0.96		1.45
F	F1	0.83		1.14	
F	G	0.97			
F	H	0.82	0.66	0.28	
F1	G	0.02			
F1	H	0.10		0.21	

Supplemental References

1. Meharena, H. S.*, et al.* (2013) Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS Biol* **11**,e1001680.
2. Batkin, M., Schvartz, I., & Shaltiel, S. (2000) Snapping of the carboxyl terminal tail of the catalytic subunit of PKA onto its core: characterization of the sites by mutagenesis. *Biochemistry* **39**,5366-5373.
3. Yang, J.*, et al.* (2009) Contribution of non-catalytic core residues to activity and regulation in protein kinase A. *The Journal of biological chemistry* **284**,6241-6248.
4. Shaltiel, S., Cox, S., & Taylor, S. S. (1998) Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. *P Natl Acad Sci USA* **95**,484-491.

5. Gibbs, C. S. & Zoller, M. J. (1991) Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. *The Journal of biological chemistry* **266**,8923-8931.
6. Thompson, E. E.*, et al.* (2009) Comparative surface geometry of the protein kinase family. *Protein Sci* **18**,2016-2026.
7. Sadowsky, J. D.*, et al.* (2011) Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *P Natl Acad Sci USA* **108**,6056-6061.
8. Scheeff, E. D. & Bourne, P. E. (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Computational Biology* **1**,e49.
9. Kannan, N., Haste, N., Taylor, S. S., & Neuwald, A. F. (2007) The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. *P Natl Acad Sci USA* **104**,1272-1277.
10. Wu, J.*, et al.* (2005) Crystal structure of the E230Q mutant of cAMP-dependent protein kinase reveals an unexpected apoenzyme conformation and an extended N-terminal A helix. *Protein science : a publication of the Protein Society* **14**,2871-2879.
11. Grant, B. D., Tsigelny, I., Adams, J. A., & Taylor, S. S. (1996) Examination of an active-site electrostatic node in the cAMP-dependent protein kinase catalytic subunit. *Protein science : a publication of the Protein Society* **5**,1316-1324.
12. Zhang, P.*, et al.* (2012) Structure and allostery of the PKA RIIbeta tetrameric holoenzyme. *Science* **335**,712-716.
13. Masterson, L. R.*, et al.* (2010) Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat Chem Biol* **6**,821-828.
14. Deminoff, S. J., Ramachandran, V., & Herman, P. K. (2009) Distal recognition sites in substrates are required for efficient phosphorylation by the cAMP-dependent protein kinase. *Genetics* **182**,529-539.