

Chapter 1

Quantifying allosteric communication via both concerted structural changes and conformational disorder

The work in this chapter is published in: Singh, S., and Bowman, G.R., Quantifying allosteric communication via both concerted structural changes and conformational disorder with CARDS. Journal of Chemical Theory and Computation. 13:1507-1517, 2017. PMID: 28282132. [?]

1.1 Abstract

Allosteric (i.e. long-range) communication within proteins is crucial for many biological processes, such as the activation of signaling cascades in response to specific stimuli. However, the physical basis for this communication remains unclear. Existing computational methods for identifying allostery focus on the role of concerted structural changes, but recent experimental work demonstrates that disorder is also an important factor. Here, we introduce the Correlation of All Rotameric and Dynamical States (CARDS) framework for quantifying correlations between both the structure and disorder of different regions of a protein. To quantify disorder, we draw inspiration from methods for quantifying “dynamic heterogeneity” from chemical physics to classify segments of a dihedral’s time evolution as being in either ordered or disordered regimes. To demonstrate the utility of this approach, we apply CARDS to the Catabolite Activator Protein (CAP), a transcriptional activator that is regulated by Cyclic Adenosine MonoPhosphate (cAMP) binding. We find that CARDS captures allosteric communication between the two cAMP-Binding Domains (CBDs). Importantly, CARDS reveals that this coupling is dominated by disorder-mediated correlations, consistent with NMR experiments that establish allosteric coupling between the CBDs occurs without a concerted structural change. CARDS also recapitulates an enhanced role for disorder in the communication between the DNA-Binding Domains (DBDs) and CBDs in the S62F variant of CAP. Finally, we demonstrate that using CARDS to find communication hotspots identifies regions of CAP that are in allosteric communication without foreknowledge of their identities. Therefore, we expect CARDS to be of great utility for both understanding and predicting allostery.

1.2 Introduction

Despite its fundamental importance, understanding of the physical mechanisms of allosteric communication remains incomplete. For example, significant effort has gone into studying how G-Protein Coupled Receptors (GPCRs) allosterically transmit extracellular signals to intracellular binding partners [?]. However, understanding of this process is still insufficient for the routine design of drugs that allosterically modulate GPCRs [?].

Models of allostery have typically focused on concerted structural changes [?]. For example, the classic induced fit model postulates that ligand binding to one subunit of a protein causes a conformational change in other subunits

[?, ?]. The conformational selection model also focuses on structural changes, positing that ligand binding to one subunit stabilizes an alternative (but pre-existing) structure of other subunits [?]. Extensive work establishes there is often a role for conformational selection [?], though there is clearly a continuum between extreme versions of induced-fit and conformational selection [?, ?, ?]. This conclusion implies allostery can be inferred from proteins' equilibrium fluctuations even in the absence of an allosteric perturbation. Inspired by this implication, numerous methods have been developed to infer allosteric communication from structural fluctuations observed in molecular simulations [?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?].

While concerted structural changes are clearly important for allostery, there is mounting evidence that conformational disorder has an important role to play, and can even lead to allosteric communication in the absence of concerted structural transitions [?, ?, ?, ?, ?, ?, ?]. The importance of allostery without conformational change first appeared in a model where ligand binding perturbs the entropy of a distant site rather than its preferred structure [?]. Since then, NMR and ITC experiments on Catabolite Activator Protein (CAP) have established allosteric communication without conformational change exists in nature [?, ?, ?]. CAP is a homodimeric transcription factor whose DNA-binding affinity increases upon binding of cAMP to the cAMP-Binding Domains (CBDs) [?, ?]. In wild-type CAP, cAMP binding allosterically induces the DNA-Binding Domains (DBDs) to swivel around the central hinge region into a DNA-binding conformation (Fig. 1.1) [?, ?]. There is also negative coupling between the two CBDs [?, ?, ?]. A combination of NMR and ITC measurements reveal that binding of cAMP to one CBD reduces the cAMP-binding affinity of the second CBD without changing its structure [?, ?]. Additional experiments reveal that binding of cAMP activates the S62F variant of CAP without causing a conformational change in the hinge or DBDs [?, ?].

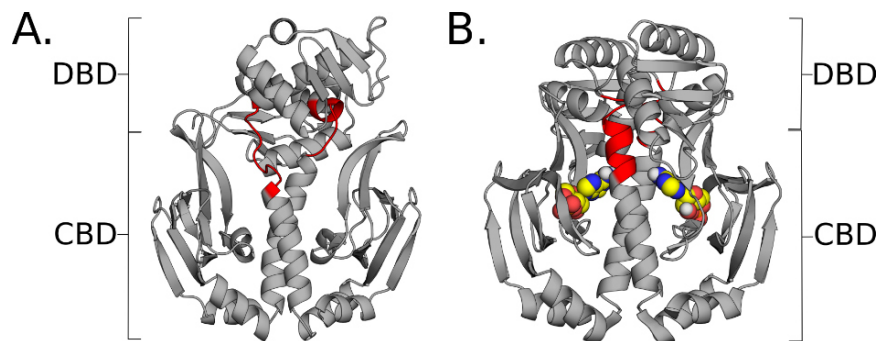


Figure 1.1: Structures of CAP in apo and holo forms. **A.** Structure of apo-CAP, with cAMP-binding domains (CBDs) and DNA-binding domains (DBDs) indicated by brackets. Residues that make up the hinge region are colored in red. **B.** Structure of holo-CAP with cAMP ligands (spheres) bound to the CBD regions. The same hinge residues are colored in red.

While the importance of disorder is gaining widespread acceptance, the field lacks systematic methods for identifying allosteric communication in the absence of conformational change. For example, NMR has yet to uncover how these signals are transmitted. COREX/BEST [?, ?, ?], other coarse-grained models [?], and normal modes [?] provide valuable insights but miss essential subtleties, such as important side-chain motions. Using molecular dynamics simulations to measure the mutual information between the orientations of different dihedral angles captures the reduction in uncertainty (measured with an entropy metric) about the structure of one dihedral given knowledge of another [?, ?, ?], but does not capture phenomena like changes in the rotameric state of one dihedral increasing the conformational heterogeneity of a distant site. Approaches for inferring allosteric coupling from sequence covariation [?, ?] are agnostic to the mechanism underlying this communication and cannot explain how it occurs. A method for identifying timing correlations has promise for capturing disorder-mediated coupling [?]. For example, application of this approach to side-chain degrees of freedom in CAP successfully identified hotspots for allosteric communication. It also demonstrated that disorder-mediated correlations give rise to long-range communication, while purely structural correlations are limited to short-ranged communication. However, timing correlations do not integrate structural and dynamical correlations into a holistic measure of communication that can capture the continuum of possibilities between purely structural and purely disorder-mediated coupling.

Here, we introduce the CARDS (Correlation of all Rotameric and Dynamical States) methodology for quantifying the roles of both concerted structural changes and conformational disorder. CARDS is based on our observation that a single degree of freedom (e.g. a dihedral angle) can transition between “ordered” regimes, wherein it undergoes small fluctuations within a single structural state, and “disordered” regimes wherein it undergoes bursts of transitions between different structural states (Fig. 1.2). Similar “dynamic heterogeneity” [?, ?, ?] has been observed in the physics of glasses, where it has been shown that a single degree of freedom’s local environment can either facilitate dynamics by flattening out the effective free energy surface that degree of freedom experiences or freeze out dynamics [?, ?, ?]. CARDS identifies ordered and disordered regimes based on two kinetic signatures: the average time a degree of freedom persists within a structural state and the typical timescale for transitions between structural states. For many dihedrals, we observe that the typical time that elapses between structural transitions, which is dominated by segments of a trajectory in disordered regimes, is orders of magnitude smaller than the typical persistence time in a state. Based on these kinetic signatures, CARDS assigns segments of trajectories to dynamical states (i.e. ordered and disordered regimes). CARDS then quantifies correlations between the structural and dynamical states of different dihedrals. Specifically, we employ the mutual information to assess how much better one can predict the structural and dynamical states of one dihedral given knowledge of the structural and dynamical states of another dihedral. To demonstrate the utility of CARDS, we assess whether it can identify allosteric communication in the absence of concerted structural changes observed in CAP.

1.3 Theory and methods

CARDS captures all forms of correlated fluctuations, including concerted structural changes, correlations between the conformational disorder of different degrees of freedom, and correlations between the structure of one degree of freedom and the conformational disorder of another. As in other recent work, we focus on dihedral angles, as they are natural degrees of freedom for describing protein structure and dynamics and are easily decomposed into a small number of rotameric states [?, ?].

1.3.1 Molecular Dynamics simulations

We ran three 500 ns simulations with our previously published protocol [?] (see SI for details). Briefly, we placed PDB ID 4N9H [?] in a dodecahedron box and solvated it with TIP3P explicit solvent [?] extending 1 nm beyond the protein in any dimension. We used PyMOL [?] to mutate Ser62 to Phe to create a starting structure for the S62F variant. For each system, we ran simulations with the Gromacs software package [?] using the Amber03 force field [?]. This combination of software and parameters was selected because it has proven reliable in our past studies of both protein folding [?] and structural fluctuations within folded proteins [?, ?, ?]. As described in the Results section, our microsecond timescale simulations are sufficient to recapitulate much of what is known from experiments about allostery in CAP. The correlation coefficient between the couplings obtained from any individual simulation and the combination of three simulations is 0.64 ± 0.02 , suggesting that hundreds of nanoseconds of simulation may be sufficient to discover gross patterns of communication but are insufficient to obtain converged results. The correlation coefficient between the coupling obtained from any pair of simulations and the combination of three simulations is 0.79 ± 0.02 , suggesting that 1-1.5 microseconds of simulation give more reproducible results. Taken together with our past work, we conclude that a few microseconds of simulation are adequate to demonstrate the utility of our new method and, likely, to gain valuable insights into many systems.

1.3.2 Assignment of dihedrals to rotameric states

Dihedral angles were calculated with MDTraj [?] and assigned to discrete rotameric states (e.g. gauche+, gauche-, and trans for most χ angles and cis or trans states for backbone dihedrals). Transition-Based Assignment (TBA) is used to distinguish lasting transitions from transient fluctuations [?, ?, ?, ?]. TBA prevents over-counting of transitions (e.g. due to fluctuations at a barrier peak where a simulation repeatedly crosses a hard cutoff between rotameric states) by defining a core region within each rotameric state and buffer zones between them. A dihedral is only considered to have changed rotameric states if it transitions from the core of one state to the core of another state, passing completely

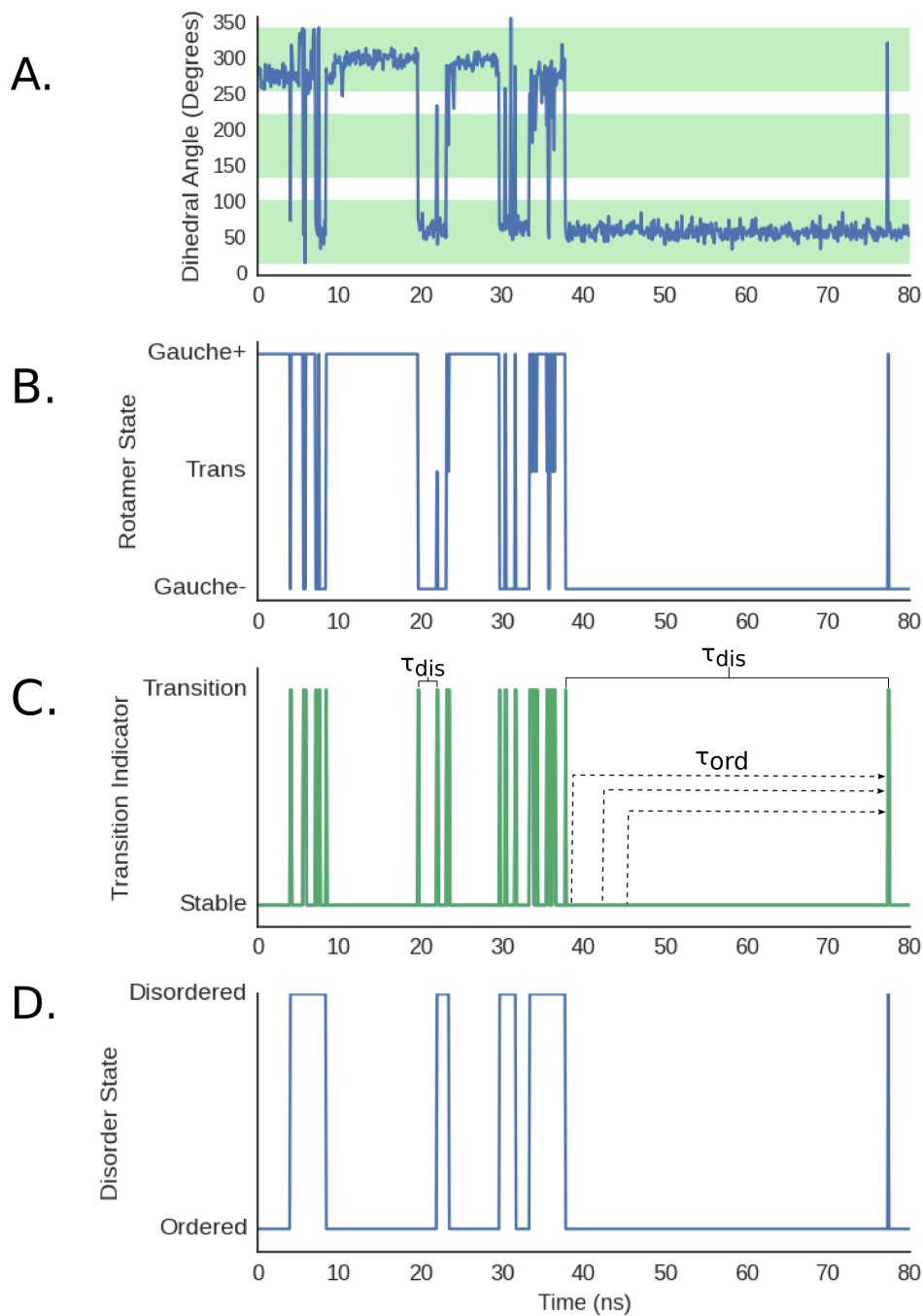


Figure 1.2: Workflow for identifying ordered and disordered regimes. **A.** An example of the time-evolution of a χ_1 angle. The core of each rotameric state is shaded green and the buffer zones between them are white. **B.** Assignment of each snapshot in the trajectory to rotameric states. **C.** Transition indicator function identifying where the dihedral changes states. Examples of ordered (τ_{ord}) and disordered (τ_{dis}) times are labeled. **D.** Assignment of each snapshot in the trajectory to dynamical states (e.g. ordered and disordered regimes).

through the buffer zone between cores (Fig. 1.2A and 1.2B). A dihedral that starts in one core, enters a buffer zone, and then returns to its initial core is said to have remained in the initial rotameric state. We define the core of a rotameric state as a region of width W centered between the boundaries between rotamers. We present results using a core width of 90° , but our results are robust to variations in the core width ranging from 60° to 90° (Fig. ??A)

1.3.3 Assignment of snapshots to dynamical states

CARDS assigns snapshots to ordered and disordered regimes based on two variables that describe the dynamics of the trajectory: the mean ordered time (τ_{ord}) and mean disordered time (τ_{dis}). An ordered time (τ_{ord}) is the time from any time-point to the next time where a transition occurs and a disordered time (τ_{dis}) is the time between two consecutive transitions (Fig. 1.2C). These times are called persistence and exchange times in the condensed matter physics literature [?, ?, ?]. For many dihedrals, we observe that $\langle \tau_{dis} \rangle \ll \langle \tau_{ord} \rangle$ because ($\langle \tau_{dis} \rangle$) is dominated by the short times between transitions in disordered regimes, whereas $\langle \tau_{ord} \rangle$ is heavily influenced by the lengthy times without any transitions in ordered regimes. To calculate these times, CARDS first identifies the time points where a dihedral transitions between two different rotameric states (Fig. 1.2C), referred to as the transition indicator function. The method then extract the complete set of τ_{dis} and τ_{ord} values in the trajectory. Next, CARDS classifies each segment of a trajectory between two consecutive transitions as being in an ordered or disordered regime based on whether the length of the segment t between the transitions is more consistent with the distribution of ordered or disordered times (Fig. 1.2D). We find that transitions within ordered and disordered regimes are roughly Poisson processes with different characteristic times ($\langle \tau_{ord} \rangle$ and $\langle \tau_{dis} \rangle$, respectively), see Fig. ?. Therefore, CARDS determines if a segment of a trajectory of length t between two consecutive transitions is more consistent with an ordered or disordered regime using the likelihood ratio (L):

$$L(t) = \frac{P_{dis}(t)}{P_{ord}(t)} = \frac{(\frac{1}{\langle \tau_{dis} \rangle})e^{-\frac{t}{\langle \tau_{dis} \rangle}}}{(\frac{1}{\langle \tau_{ord} \rangle})e^{-\frac{t}{\langle \tau_{ord} \rangle}}} \quad (1.1)$$

where P_{dis} is the probability the segment is disordered and P_{ord} is the probability it is ordered. Taking inspiration from the interpretation of Bayes factors [?], CARDS classifies a segment of a trajectory as being disordered if $L > 3.0$, otherwise the trajectory segment is classified as being in an ordered regime. Our results are robust to varying this cutoff from 1.5 to 5 (Fig. ??B).

1.3.4 Calculation of structural, disorder-mediated, and holistic correlations

The primary objective of CARDS is to calculate the total correlation between different dihedrals, including both their rotameric state and dynamical state (e.g. whether the dihedral is in an ordered or disordered regime at a given time). Towards this end, we define the holistic correlation (I_H) between dihedrals X and Y as

$$I_H(X, Y) = \overline{I_{ss}(X, Y)} + \overline{I_{dd}(X, Y)} + \overline{I_{ds}(X, Y)} + \overline{I_{sd}(X, Y)} \quad (1.2)$$

where $\overline{I_{ss}(X, Y)}$ is the normalized mutual information between the structure (i.e. rotameric state) of dihedral X and the structure of dihedral Y , $\overline{I_{sd}(X, Y)}$ is the normalized mutual information between the structure of dihedral X and the dynamical state of dihedral Y , $\overline{I_{ds}(X, Y)}$ is the normalized mutual information between the dynamical state of dihedral X and the structure of dihedral Y , and $\overline{I_{dd}(X, Y)}$ is the normalized mutual information between the dynamical state of dihedral X and the dynamical state of dihedral Y . The mutual information (I) is:

$$I(X, Y) = - \sum_{x \in X} \sum_{y \in Y} \frac{P(x, y)}{P(x)P(y)} \quad (1.3)$$

where $x \in X$ refers to the set of possible states that dihedral X can adopt, $p(x)$ is the probability that dihedral X adopts state x , and $p(x, y)$ is the joint probability that dihedral X adopts state x and dihedral Y adopts state y . We define the normalized mutual information ($I(X, Y)$) as

$$\overline{I(X, Y)} = \frac{I(X, Y)}{C(X, Y)} \quad (1.4)$$

where $C(X, Y)$ is the maximum possible mutual information between two dihedrals, called the channel capacity [?]. Using this normalized mutual information allows for a direct comparison between the different components of the holistic correlation by correcting for the fact that the largest possible mutual information between different types of dihedrals will vary based on how many different states there are. For example, a side-chain dihedral has three possible rotameric states but only two possible dynamical states, so structural correlations (I_{ss}) can be as large as $\log(3)$ while the correlations between dynamical states (I_{dd}) can only reach as high as $\log(2)$.

In addition to the above, we define the disorder-mediated correlation (I_{DM}) as all forms of correlation between two dihedrals that rely, at least in part, on disorder ($\overline{I_{sd}} + \overline{I_{ds}} + \overline{I_{dd}}$). This construct is useful for assessing the importance of disorder relative to existing methods based purely on concerted structural changes (I_{ss}). We use bootstrapping to measure the uncertainty in our estimates of all the components of the holistic correlation to ensure any comparisons we make are statistically sound. Specifically, we draw 20 random samples of our trajectories, with replacement, and calculate the structural and disorder-mediated correlations between all pairs of residues. We conclude that disorder-mediated communication dominates if the average disorder-mediated communication minus the standard deviation across all our bootstrap samples is greater than the mean structural correlation plus the standard deviation.

1.3.5 Calculation of net communication to a target site

We are often interested in calculating how much influence a particular residue has over another site, such as an active site or ligand-binding site. To calculate the communication between a reference residue and some target site, we take the average mutual information between two sets of dihedrals: 1) all dihedrals in the reference residue and its nearest neighbors and 2) all dihedrals in the target site. We define the nearest neighbors of a reference residue as all residues with atoms that fall within 3Å of any atom in the reference residue. Varying this cutoff does not alter our results (Fig. ??C). Including both a reference residue and its nearest neighbors accounts for the fact that mutating the reference residue will directly change the environment of all neighboring residues.

1.3.6 Calculation of global communication

In addition to identifying residues that have strong correlations to a specific target site, it would also be valuable to identify residues that generally appear to play an important role in allosteric networks. Towards this end, we define the global communication strength of a residue as the sum of its holistic correlations to all other residues. For these calculations, we also include neighboring residues, as in our calculation of the net communication to a target site.

1.4 Results and discussion

1.4.1 Many dihedrals have the potential for disorder-mediated communication

For a dihedral to have ordered and disordered regimes, $\langle \tau_{ord} \rangle$ must be significantly larger than $\langle \tau_{dis} \rangle$. We reasoned that determining if $\langle \tau_{ord} \rangle \geq 3 \times \langle \tau_{dis} \rangle$ is a reasonable heuristic for identifying dihedrals with separable ordered and disordered regimes based on the likelihood ratio defined in Eq. 1.1. Dihedrals that do not meet this criterion are classified as entirely being in ordered regimes and, therefore, are only capable of having structural correlations with other dihedrals.

Based on the criterion defined above, we find that 556 of the 1584 dihedrals in CAP have separable ordered and disordered regimes and, therefore, are capable of disorder-mediated communication with other dihedrals (Fig. ??). Mapping these dihedrals to the apo structure of CAP highlights a number of interesting patterns (Fig. 1.3). First of all, CARDS reveals that many side-chain dihedrals buried in CAP’s core are capable of disorder-mediated communication. This finding helps to rectify the apparent contradiction between the common physical intuition that proteins’ cores should be rigid due to their tight packing and the observation that there is substantial conformational heterogeneity within proteins’ cores [?, ?]. That is, dihedrals within a protein’s core are commonly locked in a single rotameric state for extended periods of time but rare fluctuations create room for conformational changes. Backbone dihedrals that are capable of disorder-mediated communication tend to reside on the protein’s surface. Notably, a number of these backbone dihedrals are in β -sheets that contact cAMP. However, there are also backbone dihedrals within the core that are capable of disorder-mediated communication. For example, we find backbone dihedrals within the central hinge

region that are capable of disorder-mediated communication. This observation is noteworthy because the hinge region undergoes a large conformational change upon activation of CAP (Fig. 1.1) [?, ?, ?]. We find that similar patterns emerge when we vary the cutoff for determining whether a dihedral has separable ordered and disordered regimes (Fig. ??).

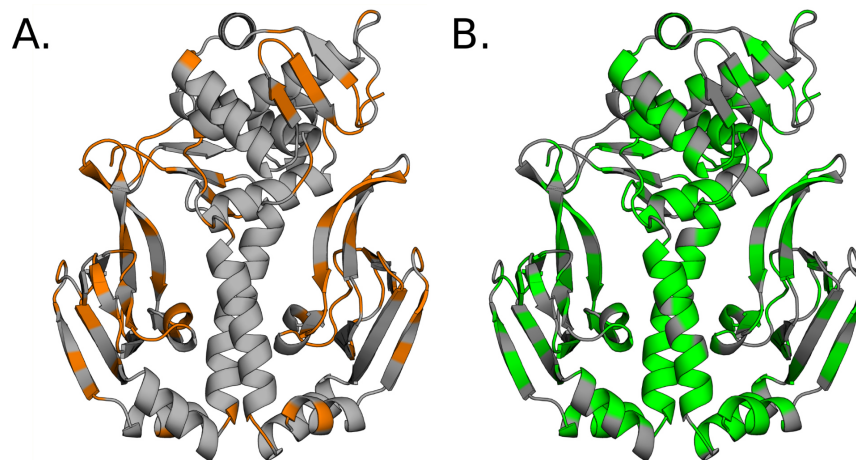


Figure 1.3: Residues whose dihedrals are capable of disorder-mediated communication. **A.** Residues with at least one backbone dihedral that is capable of disorder-mediated communication (orange). **B.** Residues with at least one side-chain dihedral that is capable of disorder-mediated communication (green).

1.4.2 Disorder-mediated correlations dominate communication between the CBDs

Given experimental evidence for allosteric communication between the CBDs without a concerted structural change [?], we expect the disorder-mediated component of the holistic correlation between these sites to be larger than the purely structural component. To test this prediction, we simulated apo CAP for 1.5 μ s and calculated the net communication of every residue to one of the cAMP-binding sites. Specifically, we defined the target site as all residues with heavy atoms within 6Å of one of the two cAMP molecules in the holo crystal structure (PDB ID 1CGP) [?]. The residues in this target site are 30, 36, 49, 61-62, 64, 69-86, and 99 from chain A and residues 122-129 from chain B.

As predicted, CARDS successfully identifies that there is communication between the two CBDs. Fig. 1.4A shows the holistic correlations to a single cAMP-binding site. Unsurprisingly, the residues with the strongest correlations to this set of residues reside within the same CBD. However, there is also strong communication between the target site and residues lining the other cAMP-binding pocket. There are also strong correlations on the central hinge region and the interface between the CBDs and DBDs that may be responsible for allosteric coupling between these domains.

To determine the relative importance of disorder-mediated communication and purely structural correlations, we broke the holistic communication into structural and disorder-mediated components. Furthermore, we used bootstrapping to estimate the uncertainty in each of these components.

Identifying residues where disorder-mediated communication to the target cAMP-binding site is larger than purely structural correlations automatically identifies a number of residues in the second cAMP-binding site (Fig. 1.4B, Fig. ??). This finding demonstrates that CARDS recapitulates the experimental finding that communication between the two CBDs does not primarily occur through concerted structural changes [?]. CARDS also identifies disorder-dominated communication within a single CBD. Furthermore, the Pearson correlation coefficient between structural and disorder-mediated communication is 0.44. This result indicates there are some similarities between patterns of allosteric coupling that could be observed by focusing entirely on structural correlations and those identified by CARDS, but that considering disorder provides additional information.

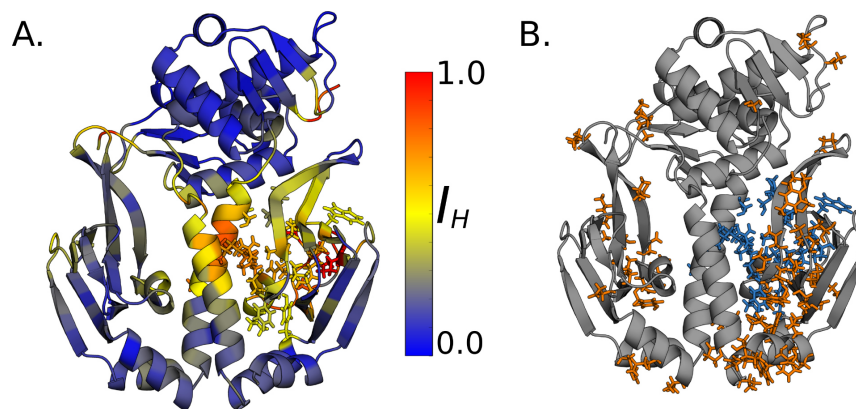


Figure 1.4: Communication to a single CBD. **A.** Holistic mutual information (I_H) of each residue to the residues lining a single cAMP-binding pocket (shown in sticks). **B.** Residues whose communication to the cAMP-binding pocket is dominated by disorder-mediated communication (orange sticks). Residues lining the target cAMP-binding pocket are in blue sticks.

1.4.3 Disorder-mediated communication is enhanced in the S62F variant

The S62F variant of CAP is still activated by cAMP binding [?, ?, ?, ?]. However, NMR studies have revealed that the conformation of the DBDs does not change upon cAMP binding. Rather, NMR and ITC experiments suggest an important role for conformational entropy, with the DBDs only changing conformation in the presence of both cAMP and DNA [?, ?]. Therefore, we expect an increase in disorder-mediated communication between the CBDs and DBDs in the S62F variant, compared to wild-type CAP.

To determine the effect of the S62F variant, we also ran 1.5 μ s of simulation of this variant. Then we calculated the holistic communication to a single cAMP-binding site, as described for wild-type CAP.

As expected, we observe significant increases in disorder-mediated communication between the target CBD and the DBDs. There are particularly large increases in disorder-mediated correlations in regions of known importance for CAP activation, such as the central hinge region and along the interfaces between the CBDs and DBDs (Fig. 1.5A). At the same time, there are some decreases in disorder-mediated communication within the CBDs. There are also changes in purely structural correlations. These changes often follow the same qualitative trends as the changes in disorder-mediated communication. However, the magnitudes of any increases in purely structural correlations are considerably smaller than the increases in disorder-mediated correlations. Furthermore, reductions in structural correlations are often larger than any decreases in disorder-mediated communication.

1.4.4 Side-chain-side-chain and backbone-side-chain correlations dominate allosteric communication in CAP

To begin understanding the relative importance of different types of degrees of freedom, we plotted the matrix of correlations between every pair of dihedrals (Fig. 1.6A). The upper triangle represents purely structural correlations, while the lower triangle represents purely disorder-mediated correlations. Side-chain and backbone dihedrals are also grouped together to enable comparisons between the relative strengths of backbone-backbone, backbone-side-chain and side-chain-side-chain correlations. Inspection of the matrix of all pairwise correlations immediately reveals that side-chain-side-chain correlations dominate allosteric communication in CAP. This observation is consistent with previous reports that side-chain degrees of freedom are more variable than the backbone [?, ?, ?, ?, ?, ?, ?, ?]. Backbone-backbone correlations are far rarer, and we find that disorder-mediated correlations between backbone dihedrals are more common than structural correlations. We also find a small number of backbone dihedrals that appear to be hubs of communication that are coupled to the side-chains of a large fraction of the residues in CAP. These hubs appear as lines in the backbone-side-chain quadrants of the matrix in Fig. 1.6A. Mapping the strongest (top 5%) backbone hubs to the structure reveals that they cluster in two regions: the phosphate-binding cassette (PBC) of each CBD and the

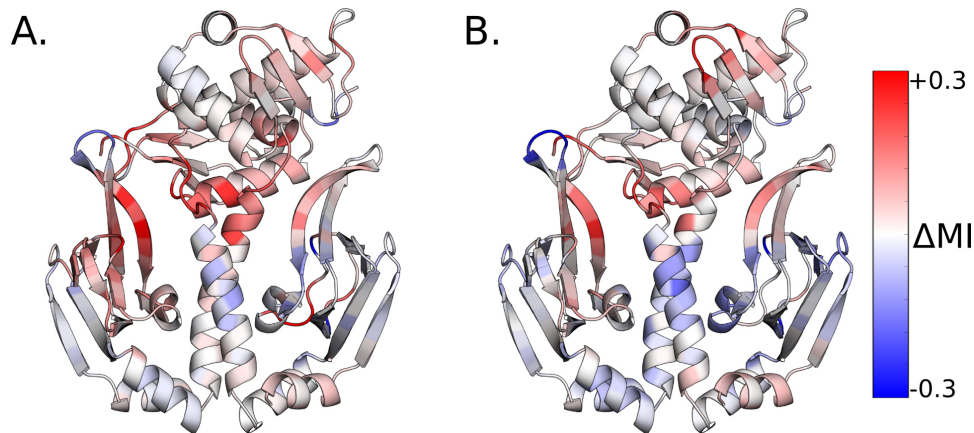


Figure 1.5: Change in coupling to a single CBD pocket upon the S62F mutation. **A.** Change in disorder-mediated communication of each residue to the single cAMP-binding pocket depicted in Fig. 1.4. **B.** Change in structural communication of each residue to the same cAMP-binding pocket. Red indicates increased communication in S62F compared to wild-type, and blue indicates decreased communication.

interface between the CBDs and DBDs, including the central hinge (Fig. 1.6B). Using an even stricter cutoff (top 2%) only identifies residues along the CBD-DBD interface and within the central hinge (Fig. ??), further emphasizing their importance in the allosteric network. This result suggests that perturbations to these functionally important regions (e.g. cAMP binding) can influence the behavior of the entire protein, and vice versa.

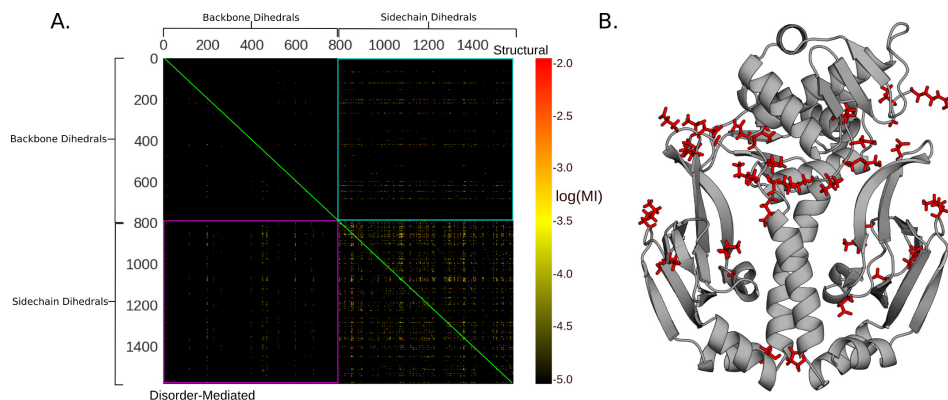


Figure 1.6: Hubs of backbone-side-chain communication in wild-type CAP. **A.** Mutual information between every pair of dihedrals in wild-type CAP. The upper triangle of the matrix (above the green diagonal) represents structural communication and the lower triangle (below the green diagonal) represents disorder-mediated communication. The pink and cyan squares encompass the regions of the matrix that capture backbone-side-chain communication that is mediated by disorder-mediated and structural coupling, respectively. **B.** Structure of wild type CAP highlighting the residues that are the strongest hubs (top 5%) of backbone-side-chain communication (red sticks).

1.4.5 Locating communication hotspots identifies key functional sites

The coincidence of backbone dihedrals that are hubs of communication and key functional sites suggests that CARDS may be capable of predicting the locations of such sites. Indeed, if evolution has selected for communication between particular sites, then one might expect residues in these sites to have stronger coupling to other regions of a protein than typical residues.

To detect strongly communicating residues, we ranked each residue based on the sum of its correlations to all other residues in the protein. This measure of global communication will highlight two types of hotspots: 1) hubs with correlations to many residues and 2) residues that have strong correlations to a few residues.

Coloring each residue in the *apo* structure according to its global communication highlights that the central hinge region and the helices between the two cAMP-binding sites are key mediators of allosteric communication in CAP (Fig. 1.7). There are also hotspot residues in other parts of the cAMP-binding sites and along the interfaces between the CBDs and DBDs. These are precisely the regions that were identified by assessing communication to a single cAMP-binding site, providing evidence that CARDS can indeed identify key functional sites without foreknowledge of their locations. This conclusion is further supported by the fact that the central hinge region has even stronger communication in the S62F variant (Fig. ??).

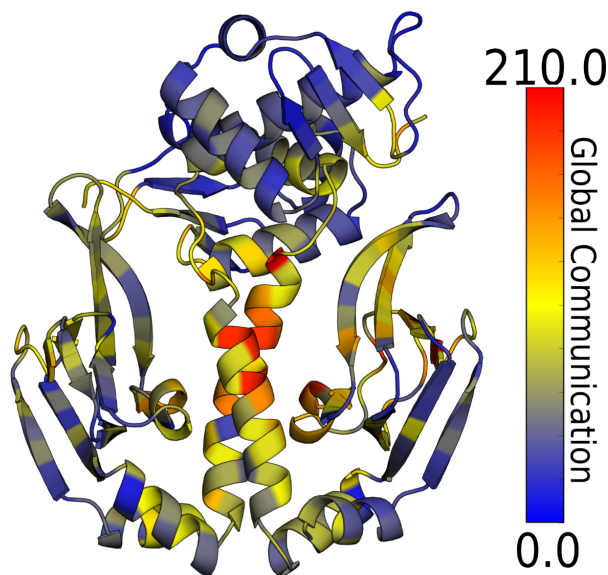


Figure 1.7: Global communication strength of each residue in *apo* CAP.

1.5 Conclusion

CARDS provides a means to integrate concerted structural changes and disorder-mediated correlations into a holistic view of allostery. Application of this approach to wild-type CAP and the S62F variant demonstrates the method's ability to identify allosteric coupling in the absence of concerted structural changes. Specifically, we showed that examining the coupling of every residue to a known cAMP binding site naturally highlights regions of the protein that are known to be impacted by cAMP binding, such as the second cAMP binding site and the central hinge region connecting the CBDs and DBDs. Decomposing the correlations between these sites into disorder-mediated and purely structural components demonstrates an important role for disorder-mediated coupling in the absence of concerted structural changes. Our global communication metric also provides a means to identify important functional sites without foreknowledge of their existence and locations. For example, this metric identifies the central hinge region—which undergoes the largest conformational change upon activation—and the cAMP binding pockets as important components of the allosteric network in CAP. Therefore, CARDS should be a powerful means to identify allosteric networks in systems that have not been studied as thoroughly as CAP. Taken together, we expect CARDS to be of great utility for understanding allostery in systems where it is already known to occur, as well as for predicting allostery in systems where it has yet to be observed.

1.6 Acknowledgements

This work was funded by NSF CAREER Award MCB-1552471. G.R.B. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation. We are also grateful to NVIDIA Corporation for the GTX Titan X used to run preliminary simulations. Finally, thanks to Kelsey C. Schuster and David Chandler, who, in collaboration with G.R.B., first demonstrated that proteins' side-chains have dynamic heterogeneity.