Title

Gauging Dynamics-driven Allostery Using a New Computational Tool: A CAP Case Study.

Journal

Journal of Molecular Biology, 436(2)

Authors

Weng, Jui-Hung

Maillard, Rodrigo

Taylor, Susan

et al.

Publication Date

2024-01-15

DOI

10.1016/j.jmb.2023.168395

Peer reviewed

HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2024 February 08.

Published in final edited form as:

J Mol Biol. 2024 January 15; 436(2): 168395. doi:10.1016/j.jmb.2023.168395.

Gauging Dynamics-driven Allostery Using a New Computational Tool: A CAP Case Study

Alexandr P. Kornev[1], **Jui-Hung Weng**[1], **Rodrigo A. Maillard**[2], **Susan S. Taylor**[1,3]

[1] -**Departmen of Pharmacology,** University of California San Diego, La Jolla, CA 92093, USA

[2] -**Department of Chemistry,** Georgetown University, Washington, DC 20007, USA

[3] -**Department of Chemistry and Biochemistry,** University of California San Diego, La Jolla, CA 92093, USA

Abstract

In this study, we utilize Protein Residue Networks (PRNs), constructed using Local Spatial Pattern (LSP) alignment, to explore the dynamic behavior of Catabolite Activator Protein (CAP) upon the sequential binding of cAMP. We employed the Degree Centrality of these PRNs to investigate protein dynamics on a sub-nanosecond time scale, hypothesizing that it would reflect changes in CAP's entropy related to its thermal motions. We show that the binding of the first cAMP led to an increase in stability in the Cyclic-Nucleotide Binding Domain A (CNBD-A) and destabilization in CNBD-B, agreeing with previous reports explaining the negative cooperativity of cAMP binding in terms of an entropy-driven allostery. LSP-based PRNs also allow for the study of Betweenness Centrality, another graph-theoretical characteristic of PRNs, providing insights into global residue connectivity within CAP. Using this approach, we were able to correctly identify amino acids that were shown to be critical in mediating allosteric interactions in CAP. The agreement between our studies and previous experimental reports validates our method, particularly with respect to the reliability of Degree Centrality as a proxy for entropy related to protein thermal dynamics. Because LSP-based PRNs can be easily extended to include dynamics of small organic molecules, polynucleotides, or other allosteric proteins, the methods presented here mark a significant advancement in the field, positioning them as vital tools for a fast, cost-effective, and accurate analysis of entropy-driven allostery and identification of allosteric hotspots.

Correspondence to Alexandr P. Kornev: akornev@ucsd.edu (A.P. Kornev).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI ChatGPT4 in order to improve readability of the document. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Alexandr P. Kornev: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rodrigo A. Maillard:** Writing – review & editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168395.

Keywords

Protein allostery; Conformational entropy; Degree centrality; Molecular dynamics; Catabolite Activator Protein

Introduction

Proteins are dynamic entities that exhibit a range of motions across multiple timescales. Notably, motions on the sub-nanosecond timescale significantly contribute to the conformational entropy of proteins, playing a crucial role in processes such as molecular recognition and allostery.[1,2] It is now acknowledged that the redistribution of these rapid internal dynamics serves as the underlying mechanism for the allosteric response in proteins, eliminating the need for predefined molecular pathways, a concept of “dynamics-driven allostery” proposed by Cooper and Dryden in 1984.[3]

The Catabolite Activator Protein (CAP) is a prime example of this phenomenon. In *Escherichia coli*, CAP binds to DNA and regulates transcription in response to elevated concentrations of cAMP.[4] Structurally, CAP is a dimer made of two identical chains, each one with a Cyclic-Nucleotide Binding Domain (CNBD) and a DNA Binding Domain (DBD) (Figure 1A). CAP belongs to a large family of cyclic nucleotide binding proteins that play important roles in biology, including protein kinases, guanine nucleotide-exchange factors and nucleotide-gated channels.[5,6] CNBDs are largely conserved with a well-defined β-barrel fold; however, CAP is a distinctively different protein as it functions only as a dimer of CNBDs that interact directly.[7] Two α-helices create a close interface between the monomers and engage with cAMP in the opposite dimer (Figure 1B). This establishes the allosteric communication between the CNBDs, resulting in a strong negative cooperativity of cAMP binding between the CNBDs of Cap.[8,9] Traditional structure-based models of allostery cannot explain the signaling observed in CAP, as the structures of CNBDs in CAP before and after cAMP binding are largely identical (RMSD = 2.4 Å for CBNDs and 3.1 Å for the whole protein).[10] Kalodimos and co-workers proposed that CAP allostery can be explained by a “dynamics-driven allostery” mechanism and confirmed this with Nuclear Magnetic Resonance (NMR).[9] They showed that binding of the first cAMP molecule leads to significant destabilization of the second CNBD. As a result of the increased mobility, the entropic penalty for the second cAMP is significantly higher, resulting in negative cooperativity. That was the first experimental confirmation of the dynamics-driven allostery mechanism. Since this pioneering study, multiple examples of dynamics-driven allostery have been identified, including bacterial ribonuclease,[11] PDZ domains,[12] *E.coli* dihydrofolate reductase,[13] zinc-dependent transcriptional repressor CzrA,[14] peptidyl-prolyl cis–trans isomerase Pin1,[15] human thymidylate synthase.[16] These studies not only underscore the relevance of dynamics-driven allostery in protein function but also indicate that the role of entropy in allostery is a much more widespread strategy used in nature.

NMR is a leading method in the study of dynamics-driven allostery, providing precise information on protein dynamics at the residue level. However, NMR can be time-consuming, expensive, and offers limited atomistic detail. Computational methods present

a faster and more cost-effective alternative that provides detailed insights into protein dynamics. However, evaluating entropy in proteins computationally is nearly infeasible because of the astronomical number of conformational states in a system with thousands of degrees of freedom. The complexity of potential spatial arrangements and interactions exceeds the capabilities of current computational methods, making the task unattainable. Recognizing these challenges, researchers have explored various strategies to overcome them. For instance, Kumar and Jernigan employed Elastic Network Models to study protein fluctuations, demonstrating the allosteric transfer of these fluctuations to distant regions of proteins.[17] Similarly, Rodgers et al. studied Normal Modes in CAP, identifying residues that significantly impact these modes, suggesting a role for Normal Modes in dynamics-driven allostery.[18] Singh and Bowman took a different approach, using dihedral angles in CAP to quantify order–disorder transitions, and identified several residues that play important roles in allosteric communication.[19] Despite these advances, the development of a comprehensive computational method that allows for a fast and quantitative estimate of entropy, which would facilitate the study of entropy-driven allostery, remains an important problem.

In this study, we introduce a new computational method, Local Spatial Pattern (LSP) alignment, to evaluate conformational entropy in proteins. Based on graph theory, LSP alignment measures the conservation of Cα:Cβ vector orientations in space during molecular dynamics (MD) simulations. Initially developed for protein structure comparison, independent of their main chain configuration, LSP alignment led to the discovery of conserved hydrophobic ensembles in protein kinases that play a critical role in their function and regulation.[20,21] Recently, we applied this approach to MD simulations of Protein kinase A (PKA), constructing Protein Residue Networks (PRNs). Analysis of several centrality measures of these networks showed that using Degree Centrality and Betweenness centrality it is possible to identify residues crucial for PKA catalytic function and structural integrity.[22]

Using CAP as a case study, we show that Degree Centrality can also capture entropic changes to explain the negative cooperativity of cAMP binding. Additionally, by examining the Betweenness Centrality of these networks, we can identify global connectors in CAP. This analysis sheds light on allosteric communication at the residue level within the protein, highlighting critical mutations that coincide with the well-documented wealth of structural and experimental data on CAP. These insights further elucidate areas that can disrupt this communication. Our results not only conform to existing data but also introduce a robust quantitative tool for understanding local stability in proteins, enriching the analysis of dynamics-driven allostery and paving new paths for exploration within the field.

Results

Linking Degree Centrality to protein entropy

In our previous work, we explored several major centrality measures in Protein Residue Networks (PRNs) constructed using Local Spatial Pattern (LSP) alignment. We discovered that using Degree and Betweenness Centrality for these networks (DC and BC, respectively), it was possible to distinguish a small set of residues crucial for Protein Kinase A (PKA) from a functional and regulatory perspective, separating them from the rest of the protein.[22] We propose the use of DC as a tool to examine local stability of proteins, thereby

providing insight into their entropic state. Essentially, LSP-alignment assesses the stability of spatial patterns formed by residues represented by their Cα:Cβ vectors. In Figure 2, we illustrate that residues that maintain a stable spatial relationship with their adjacent residues during MD simulations correspond to nodes with higher DC values in the corresponding PRNs. This method is comparable to the traditional Root Mean Square Fluctuations (RMSF) — which measures mobility of atoms during MD simulation. However, there is a critical distinction: RMSF measures the fluctuation of individual atoms compared to their average positions, while LSP-based DC evaluates the fluctuation of residue vectors relative to the orientation of nearby residues. Thus, LSP gives a significantly more comprehensive picture of residue movements, enhancing our understanding of protein dynamics far beyond what RMSF can offer.

As demonstrated in Figure 2, residues with high DC are likely to move synchronously with their immediate neighbors, analogous to a semi-rigid body. From an entropic perspective, these cohesive groups may represent regions of the protein with low entropy. It is important to note that LSP-based DC only considers Cα:Cβ vectors and, thus, disregards entropic contributions from longer side chains or peptide bond rotations. Yet, our main suggestion is that areas of the protein with lower entropy often correspond with higher DC values, and vice versa.

Measuring Degree Centrality in CAP

To test the hypothesis that areas of a protein with low entropy often correspond to high DC values, we examined cAMP binding by the 9–137 homodimer of Catabolite Activator Protein (CAP) which includes the CNBDs and the αC dimerization helix and which was used by Popovych et al.[9] The NMR studies in this paper demonstrated that binding of the first molecule of cAMP to the cAMP-binding domain A (CNBD-A) leads to increased mobility in the second cAMP-binding domain CNBD-B. This results in a higher entropic penalty for cAMP binding, which explains the negative cooperativity of cAMP binding in CAP.

DC was calculated from a 410 ns MD simulation for each residue in the homodimer of CAP for three complexes: apo CAP (DC_{apo}), one cAMP bound to the A domain (DC_A), and two cAMP molecules bound (DC_{AB}) (Figure 3A). The general DC profiles followed the secondary structures of CAP in all three cases, indicating that DC is capable of capturing the expected higher stability (i.e., lower entropy) of well-structured α-helices and β-strands. However, several regions exhibited notable differences. To evaluate these differences, we calculated the changes in DC after cAMP binding to CNBD-A ($\Delta DC_{A\text{-}apo}$) and after the binding of two cAMP molecules ($\Delta DC_{AB\text{-}apo}$) (Figure 3B).

When two cAMP molecules were bound, we observed three main features: First, the values of $\Delta DC_{AB\text{-}apo}$ are positive in many regions of the protein, indicating greater stability (or lower entropy) in the cAMP-bound state compared to apo. The regions of high stability were β2, the β2-β3 loop, β4, β5, β6, the Phosphate Binding Cassette (PBC) and the C-terminal part of the βC helix, including the residues that are in direct contact with cAMP and residues that are distantly located, that is, long-range effects. Second, residues that are in direct contact with cAMP (Figure 1B) show the highest $\Delta DC_{AB\text{-}apo}$: G71, E72, S83, S128. A set

of hydrophobic residues that form stable contacts with cAMP directly or side chains of two main cAMP binding residues (E72 and R82) also increase their DC levels: I30, I51, L61, V131. Third, the regions and magnitudes of positive $\Delta DC_{AB\text{-}apo}$ are similar in both CNBDs, indicating symmetric allosteric effects due to cAMP binding.

However, when only CNBD-A had a bound cAMP molecule, we observed a more complex and asymmetric $\Delta DC_{A\text{-}apo}$ response between the two CNBDs in the homodimer. CNBD-A showed an increase in $\Delta DC_{A\text{-}apo}$ similar to that observed when two cAMP molecules are bound except at the C-terminal (residues 127–135) which has statistically significant negative values (Figure 3C). In contrast, CNBD-B displayed negative $\Delta DC_{A\text{-}apo}$ values throughout its structure (β4: residues 49–52, β5: residues 59–62, and PBC: residues 77–84) (Figure 3C). Negative $\Delta DC_{A\text{-}apo}$ in the CNBD-B values agree with the increase in local dynamics detected by NMR[9] and is consistent with the dynamics-driven allostery model. It also agrees with the result of the computational study by Rodgers et al. suggesting the role of L61 and V131 in allosteric signaling in CAP.[18]

Comparing Degree Centrality values with previous entropy estimates

In their previous work, Popovych et al.[9] offered estimates for changes in CAP entropy during sequential binding of cAMP molecules. Their isothermal titration calorimetry (ITC) measurements suggest an entropic penalty of 8.5 kcal/mol for the binding of the first cAMP molecule and a 4.6 kcal/-mol penalty for the second cAMP binding. In comparison, their estimates based on NMR-derived order parameters indicated entropic penalties of 3.2 kcal/mol and 18.1 kcal/mol for the first and second cAMP bindings, respectively. It is important to note that the estimates derived from NMR data are based on the assumption that the observed fluctuations are independent of other unobserved fluctuations, implying that these values represent an upper bound for the entropic penalties. In our work, we suggest that changes in DC associated with cAMP binding reflect local order/disorder states of the protein, and thus be related to its conformational entropy. When comparing our results with ΔS^2 (NMR order parameters) of the study by Popovych et al.,[9] we did not observe a strong correlation between ΔS^2 and ΔDC (Figure 4). However, most of the data points are located in the upper right quadrant of the plot, indicating that the second binding of cAMP led to an increase in both ΔDC and ΔS^2. Notable outliers, such as E72 and L73 of the PBC, are in the upper left quadrant. According to the ΔS^2 estimates they experienced a significant increase of dynamics upon second cAMP binding. At the same time, most neighboring residues in the PBC were reported to decrease their dynamics (G74, Q80, E81, S83) that correlates with our estimates of ΔDC.

As NMR experiments measure the mobility of the peptide bond amides and our method assesses mobility of the Cα:Cβ vectors, we repeated our DC calculations using Nitrogen-Oxygen (N:O) vectors that reflect the mobility of the main chain. As shown in Figure 4, the results for both side chain and main chain DC were rather close ($R^2 > 0.81$), with few exceptions, such as L73, S83 and Q80 of the PBC. Additionally, a graph of the DC difference between the side chain and main chain shows values close to or around zero (Figure S1) in the three forms of CAP (apo, bound to one or two cAMP) and were consistent for both CNBDs (Figure S2). Altogether, our estimates of entropic properties of CAP follow

the same trend as the numbers obtained by NMR, and are consistent with the proposed mechanism of the entropy-driven allostery in CAP. The notable difference between the ITC measurements of entropy, the NMR-based estimates, and our evaluation using the graph-theoretical approach calls for cautious interpretation, suggesting a potential for further refinement in understanding the entropic contributions during cAMP binding in CAP. While quantifying conformational entropy from Degree Centrality values in our Protein Residue Networks is a complex challenge that extends beyond the scope of this study, our primary goal is to establish that DC values positively correlate with conformational entropy. This correlation indicates that DC can serve as a practical estimator of the entropic characteristics of proteins, providing a basis for future research to further explore and refine this approach.

Measuring Betweenness Centrality in CAP

We then studied Betweenness centrality (BC) in the LSP-based PRNs (BC), which can provide valuable information on global connectivity within the protein molecule.[22] The main points of connectivity in the CAP homodimer were at β3, β6, and the αC helix (Figure 5A). The changes in BC after binding to one and two cAMP molecules are shown in Figure 5B. Residues with positive ΔBC are not essential connectors in the apo state, but become significant points of contact after binding of cAMP. As expected, the new connection points are located at the PBC that makes direct contacts with cAMP, such as L61, G71, E72, and S83. On the contrary, residues such as Y41 (β3), D68, F69 (β6), and L116 (αC) are vital connectors in apo CAP, as they are in the center of the β-barrel but become less prominent upon cAMP binding. After binding of a single cAMP to CNBD-A, significant changes were observed in BC around the second cAMP binding site, most notably in S128A, L61B and D68B (Figure 5C).

To gain deeper insights into the observed changes in DC and BC, we visualized the corresponding PRNs using the ForceAtlas2 layout algorithm implemented in the Gephi[23] network visualization software (Figure 6). This algorithm treats the weights assigned to the edges as an attractive force that balances the imposed repulsion of the nodes. In LSP-based PRNs, compact, highly interconnected nodes correspond to more ordered regions of the protein, where Cα:Cβ vectors move cohesively and preserve their mutual orientations. In general, the PRN for CAP with two bound cAMP molecules appeared much more compact compared to that for the apo state. When cAMP was bound only to CNBD-A, the PRN became visibly asymmetrical. This is consistent with the concept that cAMP binding introduces new contacts that coordinate the thermal motions of the surrounding residues and stabilize their Cα:Cβ vector orientations, as measured by LSP alignment. Although the crystallographic structures of these proteins are superimposable, the LSP-based PRNs offer a way to visualize the local stability of the residues, serving as a dynamic map of the molecule.

When cAMP binds to a CNBD, it introduces new robust connections between L61 and S83 from the β-barrel and R123 and S128 from the αC helices. These connections are virtually absent in the apo state, where the main line of connectivity lies along the Y40-F69-L116-M120 line. This explains the significant BC changes in the single cAMP-bound CAP, where L61 in the CNBD-B and S128 in the CNBD-A become even less of a connector than in the

apo CAP. In contrast, D68 becomes a much more significant connector in CNBD-B when cAMP is bound to CNBD-A, making it the most prominent outlier in the $\Delta BC_{A\text{-}apo}$ graph (Figure 5B). Noticeably, mutant D68G had been reported to affect CAP activity without changing the affinity of CAP to DNA.[24] Our results suggest that such a mutation would affect connectivity in CNBD-B upon the first cAMP binding, making it even more mobile and thus increasing the entropic penalty for the second cAMP binding. This would lead to a decrease in the sensitivity of CAP to cAMP.

Discussion

In this study, we introduced a new computational approach to evaluate conformational entropy in proteins using Local Spatial Pattern (LSP) alignment. The method uses short, approximately 10 ns-long, MD simulations. The conformations derived from these simulations are analyzed using LSP alignment to detect stable patterns formed by the residues. Based on this analysis, we construct a Protein Residue Network (PRN) where the edge weights reflect the mutual stability of the two residues. We argue that Degree Centrality in these PRNs (DC) should reflect local stability of residues (Figure 2) and thus can be associated with the entropy component that is known to drive the entropy-driven allostery in proteins.[3]

To test this hypothesis, we created LSP-based PRNs for the catabolite activator protein (CAP) and calculated Degree Centrality (DC) for these networks. CAP is a well-studied protein that binds two cAMP molecules with strong negative cooperativity. It is driven by significant changes in entropy and therefore can be used as a benchmark for methods that attempt to gauge entropic properties in proteins.[9] According to our hypothesis, after the first cAMP molecule binding to the A-domain, the DC values have to go up for this domain while going down for the B-domain. That was precisely what we observed in the experiment (Figure 3). Our results show that the detected values of DC are very robust and allow for the calculation of statistical significance of these values.

Our approach diverges sharply from previous computational work on allostery-related protein entropy studies, which typically focused on slow, large global motions of proteins. While traditional methods, such as Normal Mode Analysis (NMA), conducted with microseconds of molecular dynamics (MD), have successfully characterized overall protein dynamics,[25] the prevailing emphasis on slow motions may overlook other essential aspects of protein behavior. Although slow global motions are undeniably vital for protein functions, they are less relevant for the study of fast motions in sub-nanosecond timescales, which constitute a significant portion of entropy.[1,2] The success of LSP-based PRNs presented in this work illustrates the potential advantages of prioritizing local stability over global dynamics when addressing entropy-related problems. In our study, we examined six 10 ns intervals, with the first three taken consecutively and the others spaced at 200–210 ns, 300–310 ns, and 400–410 ns to evaluate possible drifts in centrality values. We demonstrated that the centrality calculations were highly convergent across all three states of the CAP throughout the trajectory. This finding aligns with our previous results on LSP-based PRNs of PKA,[22] where we found that centrality values converge at around 10 ns. Here, we further

established that this consistency is true for at least several hundred nanoseconds. Whether this convergence will persist over longer time frames remains an open question.

Although our approach based on LSP is markedly different from traditional methodologies based on NMA, we observed a notable overlap with the previous study by Rodgers et al. They found that residues L61 and V131 (from the opposite monomer) affect slow normal modes in CAP.[18] These residues are in direct contact with each other and with S128, which binds to the adenine ring of cAMP (Figure 1B). This interaction may explain why these specific residues were found to be so influential. In our study, we discovered that L61 and V131 are among the residues that show the most significant destabilization of the CNBD-B binding site upon binding of the first cAMP to the CNBD-A. Consequently, it is logical that *in silico* mutations of these residues would lead to a global effect on CAP, as evident in changes to the normal modes. However, our method offers more comprehensive and statistically significant quantitative data, along with a clear physical interpretation of the allosteric phenomenon.

The calculation of DC in LSP-based PRNs enables us to gauge a specific aspect of entropy related to thermal motions that elucidate the entropy-driven allostery in CAP. However, the creation of these PRNs offers an additional advantage as they can be analyzed using standard graph-theoretical methods. Among these, the calculation of Betweenness Centrality (BC) which defines global connectivity within a network.[26,27] Nodes with high BC serve as connectors, bridging densely interconnected communities within the graph. As previously demonstrated in our analysis of PRNs in PKA,[22] employing both DC and BC effectively distinguished residues vital to PKA's structure and function. Through the calculation of BC, we identified known connector residues in the CAP homodimer: E72, S83, L116, R123 and S128 (Figure 5). These findings align with established knowledge about CAP, particularly the role of these residues at the interface between CNBD and cAMP or between two CNBDs, adding quantitative support to the existing understanding. The identification of D68 from the β6 strand was somewhat unexpected, as it emerged as a major connector in CNBD-B that was most affected by the first binding of cAMP to CNBD-A, indicating its role in the propagation of allosteric signaling within CAP. It seems that this residue plays a critical role in the connectivity between the central helices and the beta barrel, a finding that is consistent with previously reported mutational data.[24]

This study serves as an assessment of the ability of LSP-based PRNs to replicate or approximate what has been previously achieved through NMR to understand the CAP allostery.[9] The results presented in Figure 3 confirm that this is indeed the case, thereby establishing LSP-alignment as a viable tool to study entropy-driven allostery. Two significant advantages of the LSP- based approach are worth highlighting. First, as a graph-theory-based method, it provides additional ways to analyze protein dynamic maps, such as calculating centralities. Here, we demonstrate that the calculation of Betweenness Centrality provides valuable and meaningful information (Figure 5), confirming our previous report on PKA.[22] Second, our approach is not limited to individual residues, but can also include various non-protein entities, such as polynucleotides, lipids, or ligands, such as ATP in our previous work on PKA or cAMP in this study. The versatility of our approach lies in its ability to construct network graphs from any elements that can be defined as vectors.

For example, in this research, we used Nitrogen:Oxygen vectors instead of Cα:Cβ vectors to focus on the movements of the main chain, simply by adjusting the vector definitions in our configuration file. Furthermore, these vectors can represent functionally significant groups in amino acids, phosphate groups, or other post-translational modifiers. Our method is adaptable to any protein whose structure can be analyzed through Molecular Dynamics simulations, without limitations on the type of protein. This flexibility greatly expands the capabilities of the method to a broad array of protein systems that are regulated allosterically.

Although this computational method is relatively efficient, it has certain limitations regarding the size of the system. Our current protocol involves an all-to-all comparison of 100 structures from a 10 ns trajectory, leading to nearly 5000 LSP-alignments. While a single alignment on a standard computer takes under a minute, the total analysis for a protein with fewer than 500 residues may take several days. As the size expands to 2000 residues, the required time also increases, taking several hundred hours and demanding more powerful computing resources. Enhancing the efficiency of the algorithm is part of our future development goals.

Despite these limitations, our method opens up a broad range of applications across various protein systems, allowing for rapid in silico analysis of dynamic maps in sub-nanosecond timescale. This capability facilitates the study of changes in thermal motions within proteins, which can be related to various factors such as mutations, binding with ligands, or interactions with other proteins. The significance of this approach becomes particularly relevant with the recent shift in the paradigm of structural biology, recognizing a more complex model of “sequence-structure-dyna mics-function.”[28,29] The ability to examine proteins that can change their dynamics without altering their tertiary structure is crucial, especially in the context of entropy-driven allostery, as elucidated by Cooper and Dryden.[3]

Materials and Methods

Molecular dynamics simulation

The simulation model was established using the dimeric structure of the Catabolite Gene Activator Protein (CAP) (PDB: 4hzf). Residues 10 to 137 from both chains were selected as the initial model. The full system was parameterized using LEaP in Amber.[30] The resulting model was solvated in a cubic box of TIP4P-EW and 150 mM KCl with a 10 Å buffer in AMBER tools. The energy minimization, heating, and equilibration steps were performed using AMBER16.[31] Systems were first minimized through 900 steps of hydrogen-only minimization, 2000 steps of solvent minimization, 2000 steps of side-chain minimization, and 5000 steps of all-atom minimization. The systems were then heated from 0 K to 100 K linearly over 250 ps with 2 fs time-steps and 5.0 kcal mol^{-1} $Å^{-2}$ position restraints on the protein backbone under constant volume. Langevin dynamics was used to control the temperature using a collision frequency of 1.0 ps^{-1}. Then the systems were heated from 100 K to 300 K linearly over 200 ps with 2 fs time-steps and 5.0 kcal mol^{-1} $Å^{-2}$ position restraints on the protein backbone under constant pressure. Constant pressure equilibration was performed with a 10 Å non-bonded cut-off and particle mesh Ewald was performed by 250 ps, with 5.0 kcal mol^{-1} $Å^{-2}$ position restrains on the protein backbone, followed by 250

ps of unrestrained equilibration. Finally, the simulation was equilibrated for 10 ns before the 10 ns production simulations. These simulations were conducted in triplicate for each construct for 10 ns using a GPU-enabled AMBER16. Additionally, for each condition, one simulation was extended to 410 ns.[32,33]

Construction of LSP-based PRNs

LSP alignment was performed using previously created software[20,21] adapted for molecular dynamics simulation as described earlier.[22] In brief: protein structures derived from the MD simulation were used to create Protein Residue Networks with nodes representing residues. Weighted edges between nodes were created if the distance between the corresponding Cα atoms was within 12 Å. The weights were represented by four numbers defining the mutual orientation of Cα:Cβ vectors of the three distances ($C_{\alpha 1}$-$C_{\alpha 2}$, $C_{\alpha 1}$-$C\beta_2$, $C_{\beta 1}$-$C_{\alpha 2}$) and the dihedral angle Θ ($C_{\beta 1}$-$C_{\alpha 1}$-$C_{\alpha 2}$-$C_{\beta 2}$) (Figure 2B). LSP-alignment of two graphs is presented as a graph with residues as nodes and links created only if all the three distances and the dihedral angles are similar, i.e., within predefined cutoff levels: $\Delta C_{\alpha 1 \alpha 2} < 0.2$ Å, $\Delta C_{\alpha 1 \beta 2} < 0.45$ Å, $\Delta\Theta < 10°$. Weights for the links are calculated using the following formula:

$$W = \frac{1}{4} * \left(\left(1 - \frac{\delta_{\alpha 1 \alpha 2}}{\Delta\, C_{\alpha 1 \alpha 2}}\right) + \left(1 - \frac{\delta_{\alpha 1 \beta 2}}{\Delta\, C_{\alpha 1 \beta 2}}\right) + \left(1 - \frac{\delta_{\beta 1 \alpha 2}}{\Delta\, C_{\alpha 1 \beta 2}}\right) + \left(1 - \frac{\delta\theta}{\Delta\, \theta}\right) \right)$$

where $\delta_{\alpha 1 \alpha 2}$, $\delta_{\alpha 1 \beta 2}$ and δΘ are the corresponding differences between two Cα:Cβ vectors. Weights for the not matching links are assigned zero values. For LSP-alignment of multiple structures, comparisons were made in all-to-all way. The resulting adjacency matrices were averaged and used for the subsequent analysis.

Normalized centralities were calculated using the igraph R library (version 1.4.2).[34] To calculate betweenness and closeness centralities weights were converted to distances using the following formula: $D = -\log W$. “Strength” function was used to calculate weighted Degree Centrality.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The work was supported by NIGMS, National Institute of Health (grant 5R35GM130389 to S.S.T. and and R15 GM135866 to R.A.M.).

Glossary

LSP-alignment (Local Spatial Pattern alignment)
A computational method for assessing the similarity of patterns formed by vectors in 3D space. For proteins, this typically involves the Cα:Cβ vectors of the residues.

PRN (Protein Residue Network)

A graph representation of residues in a protein. The nodes in this graph represent residues, while the links (or edges) represent relationships between these residues. The nature of these relationships can vary depending on the specific definition used for the PRN in question.

Degree Centrality
A metric for a node in a graph that denotes the number of links (or edges) connected to this node.

Weighted Degree Centrality
A variant of degree centrality, where the value for a node is the sum of the weights of links connected to it.

Betweenness Centrality
A metric for a node in a graph that quantifies the number of times that this node acts as a bridge on the shortest path between two other nodes.

CAP (Catabolite Activator Protein)
A transcription activator protein that is activated by cyclic AMP.

CNBD (Cyclic Nucleotide Binding Domain)
A conserved protein domain known for binding cyclic nucleotides, including cyclic AMP.

DBD
The DNA-binding domain present in CAP.

PBC (Phosphate Binding Cassette)
A conserved sequence of about 12 residues in CNBDs that serve as a major binder for cAMP phosphate.

References

1. Wand AJ, (2013). The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. Curr. Opin. Struct. Biol 23, 75–81. [PubMed: 23246280]
2. Wand AJ, Sharp KA, (2018). Measuring entropy in molecular recognition by proteins. Annu. Rev. Biophys 47, 41–61. [PubMed: 29345988]
3. Cooper A, Dryden DT, (1984). Allostery without conformational change. A plausible model. Eur. Biophys. J 11, 103–109. [PubMed: 6544679]
4. Harman JG, (2001). Allosteric regulation of the cAMP receptor protein. Biochim. Biophys. Acta-Protein Struct. Molec. Enzym 1547, 1–17.
5. Berman HM et al. , (2005). The cAMP binding domain: An ancient signaling module. PNAS 102, 45–50. [PubMed: 15618393]
6. Kannan N et al. , (2007). Evolution of allostery in the cyclic nucleotide binding module. Genome Biol. 8, R264. [PubMed: 18076763]
7. Kornev AP, Taylor SS, Ten Eyck LF, (2008). A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. PLoS Comput. Biol 4, e1000056. [PubMed: 18404204]
8. Heyduk E, Heyduk T, Lee JC, (1992). Intersubunit communications in Escherichia coli cyclic AMP receptor protein: studies of the ligand binding domain. Biochemistry 31, 3682–3688. [PubMed: 1314647]

9. Popovych N, Sun S, Ebright RH, Kalodimos CG, (2006). Dynamically driven protein allostery. Nature Struct. Mol. Biol 13, 831–838. [PubMed: 16906160]
10. Popovych N, Tzeng SR, Tonelli M, Ebright RH, Kalodimos CG, (2009). Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. PNAS 106, 6927–6932. [PubMed: 19359484]
11. Zhuravleva A et al. , (2007). Propagation of dynamic changes in barnase upon binding of barstar: an NMR and computational study. J. Mol. Biol 367, 1079–1092. [PubMed: 17306298]
12. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL, (2009). Hidden dynamic allostery in a PDZ domain. PNAS 106, 18249–18254. [PubMed: 19828436]
13. Boehr DD et al. , (2013). A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. Biochemistry 52, 4605–4619. [PubMed: 23758161]
14. Capdevila DA, Braymer JJ, Edmonds KA, Wu H, Giedroc DP, (2017). Entropy redistribution controls allostery in a metalloregulatory protein. PNAS 114, 4424–4429. [PubMed: 28348247]
15. Wang J et al. , (2017). Dynamic Allostery Modulates Catalytic Activity by Modifying the Hydrogen Bonding Network in the Catalytic Site of Human Pin1. Molecules 22
16. Bonin JP, Sapienza PJ, Lee AL, (2022). Dynamic allostery in substrate binding by human thymidylate synthase. Elife, 11.
17. Kumar A, Jernigan RL, (2021). Ligand binding introduces significant allosteric shifts in the locations of protein fluctuations. Front. Mol. Biosci 8, 733148 [PubMed: 34540902]
18. Rodgers TL et al. , (2013). Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors. PLoS Biol 11, e1001651. [PubMed: 24058293]
19. Singh S, Bowman GR, (2017). Quantifying allosteric communication via both concerted structural changes and conformational disorder with CARDS. J. Chem. Theory Comput 13, 1509–1517. [PubMed: 28282132]
20. Kornev AP, Haste NM, Taylor SS, Ten Eyck LF, (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. PNAS 103, 17783–17788. [PubMed: 17095602]
21. Kornev AP, Taylor SS, Ten Eyck LF, (2008). A helix scaffold for the assembly of active protein kinases. PNAS 105, 14377–14382. [PubMed: 18787129]
22. Kornev AP, Aoto PC, Taylor SS, (2022). Calculation of centralities in protein kinase A. PNAS 119
23. Jacomy M, Venturini T, Heymann S, Bastian M, (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 9, e98679. [PubMed: 24914678]
24. Zhang X, Schleif R, (1998). Catabolite gene activator protein mutations affecting activity of the araBAD promoter. J. Bacteriol 180, 195–200. [PubMed: 9440505]
25. Krieger JM, Doruker P, Scott AL, Perahia D, Bahar I, (2020). Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods. Curr. Opin. Struct. Biol 64, 34–41. [PubMed: 32622329]
26. Girvan M, Newman ME, (2002). Community structure in social and biological networks. PNAS 99, 7821–7826. [PubMed: 12060727]
27. Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A, (2013). Protein contact networks: an emerging paradigm in chemistry. Chem. Rev 113, 1598–1613. [PubMed: 23186336]
28. Kornev AP, Taylor SS, (2015). Dynamics-driven allostery in protein kinases. Trends Biochem. Sci 40, 628–647. [PubMed: 26481499]
29. Medina E, Latham DR, Sanabria H, (2021). Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies. Curr. Opin. Struct. Biol 66, 129–138. [PubMed: 33246199]
30. Case DA et al., (2016). AMBER University of California, San Francisco.
31. Homeyer N, Horn AH, Lanig H, Sticht H, (2006). AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. J. Mol. Model 12, 281–289. [PubMed: 16240095]

32. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC, (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. J. Chem. Theory Comput 9, 3878–3888. [PubMed: 26592383]

33. Le Grand S, Gotz AW, Walker RC, (2013). SPFP: speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun 184, 374–380.

34. Csardi G, Nepusz T, (2006). The igraph software package for complex network research. InterJornal Complex Systems 1695

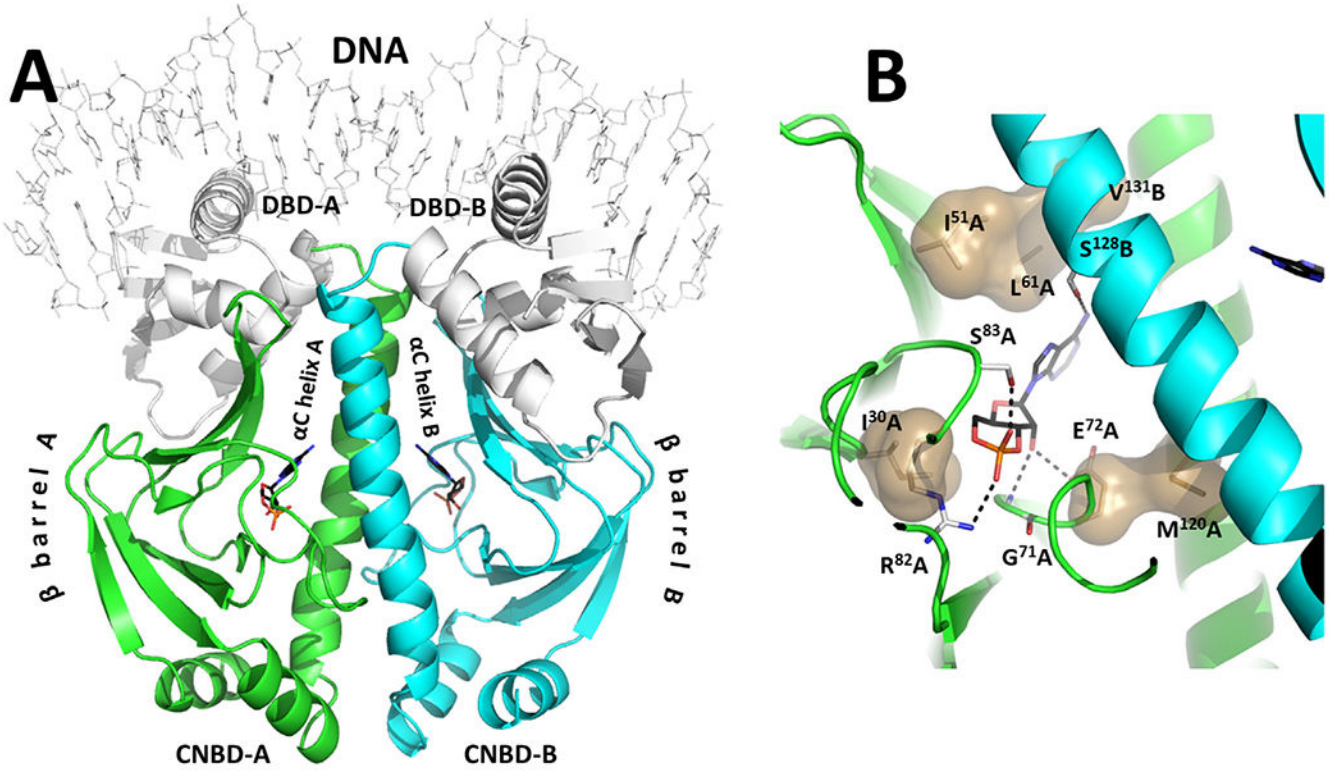

Figure 1. Catabolite Activator Protein (CAP) structure.

A) Overall view of CAP homodimer bound to a DNA molecule. The white cartoon represents N-terminal DNA binding domains (DBDs) that were not included in the analysis. Two cAMP molecules bind between the corresponding β-barrels and the central bundle of two long αC-helices. B) Closeup of the cAMP binding site with several key residues shown as sticks. Prominent hydrophobic contacts are shown as semitransparent surfaces.

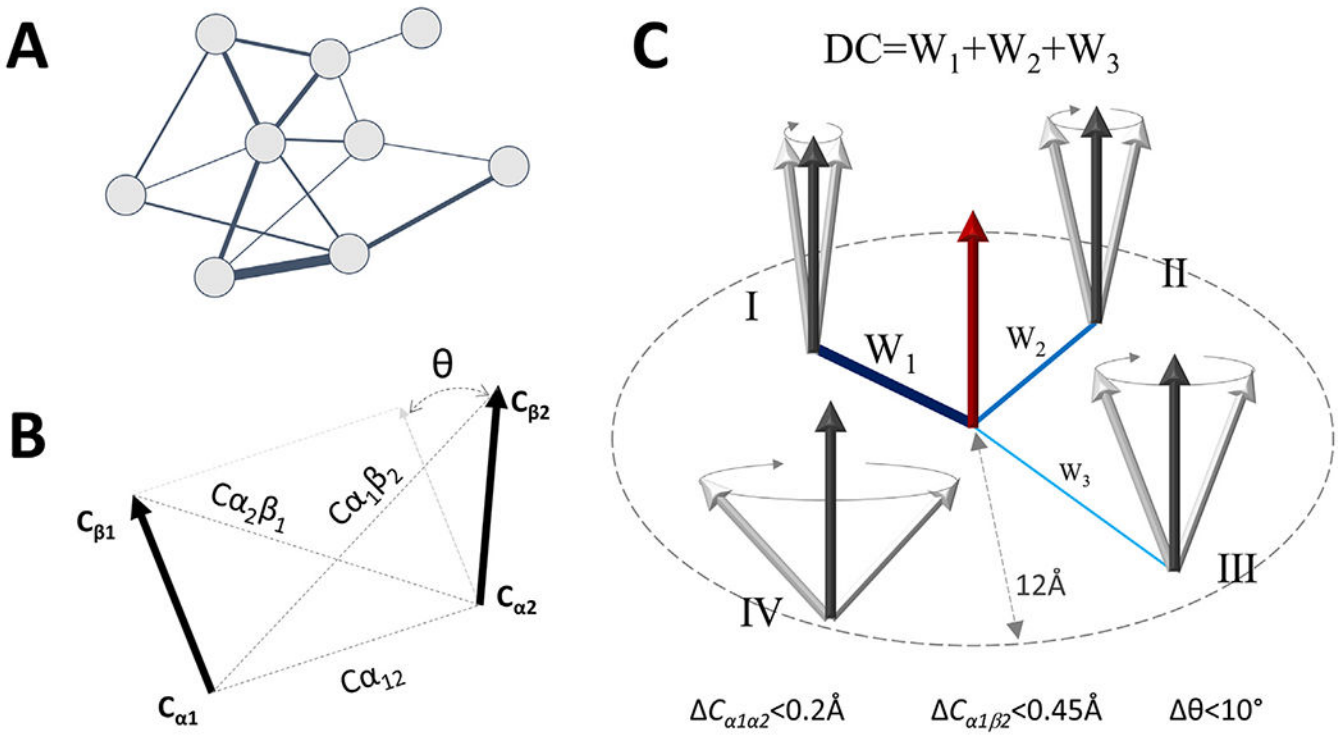

Figure 2. Interpretation of Degree Centrality in LSP-based Protein Residue Networks.

A) In unweighted graphs, the Degree Centrality (DC) value for a node is the total number of its connections. For weighted graphs, connections carry specific values or "weights" (represented by lines of varying width). In this case, DC is determined by summing the weights of all connections linked to the node. B) Mutual positions of Cα:Cβ vectors in LSP-based PRNs are described by four parameters for each residue pair: three distances ($C\alpha a_{12}$, $C\alpha_1\beta_2$, $C\alpha_2\beta_1$) and one dihedral angle Θ. C) Computation of DC for the central residue: Four residues within a predefined cutoff level for $C\alpha_{12}$ distance (12 Å) are shown, labeled I-IV. The weights of their links reflect the mutual stability of each residue with respect to the central one during the MD simulation of the protein. Shown stabilities decrease from residue I to IV. Spatial variations over time for each residue are compared against the shown cut-off levels and calculated by the formula shown in Materials and Methods. W_1 has the highest weight due to the least fluctuations. Residue IV is not assigned any weight as its variations exceed the thresholds. The DC is the cumulative sum of the three weights (W_1, W_2, and W_3).

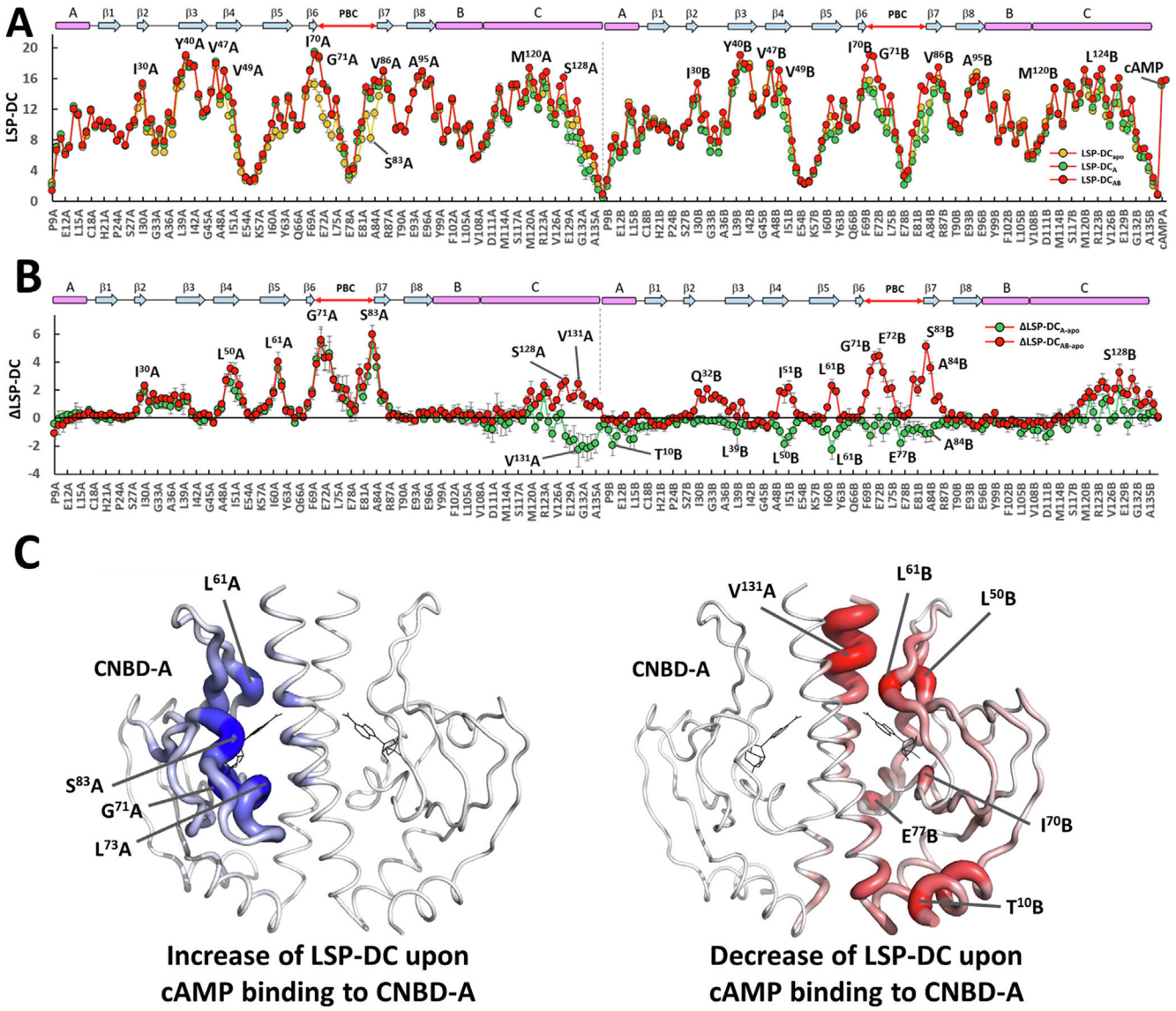

Figure 3. Changes of Degree centrality in CAP LSP-based PRNs during cAMP binding.

A) DC values for the CAP homodimer, including cAMP molecules. The secondary structure of CAP is shown as a cartoon at the top. **B)** Changes in DC after binding of the first cAMP to CNBD-A ($\Delta DC_{A\text{-}apo}$) and two molecules of cAMP ($\Delta DC_{AB\text{-}apo}$). Standard errors of the mean for six measurements are shown. **C)** Positive (left) and negative (right) values of $\Delta DC_{A\text{-}apo}$ shown on the B panel (green dots) mapped on the CAP structure. Positive $\Delta DC_{A\text{-}apo}$ correspond to more stable regions (blue color is chosen to depict stability), negative $\Delta DC_{A\text{-}apo}$ show increase in the mobility of Cα:Cβ vectors (red color represents higher mobility). Two cAMP binding sites are indicated by black lines.

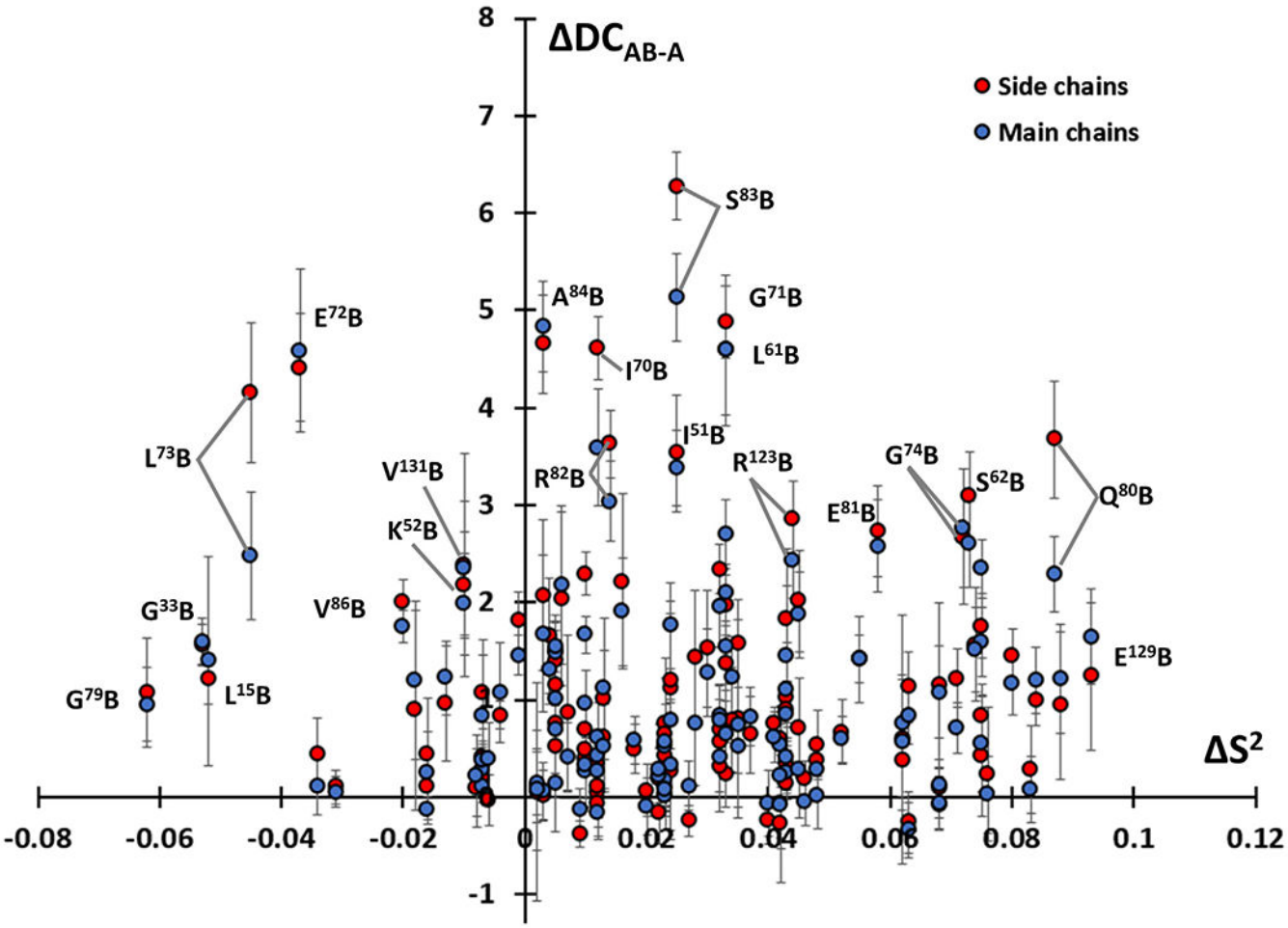

Figure 4.

Difference between Degree Centralities in CNBD-B in two cAMP bound (DC_{AB}) and one cAMP bound (DC_A) CAP compared to the order parameter ΔS^2 from Ref.[9] DC was calculated for two different representations of residues: using Cα:Cβ vectors, representing stability of side chains (red dots) and NO vectors, representing main chains (blue dots). The values of ΔS^2 were approximated from the Figure S5 of Ref.[9]

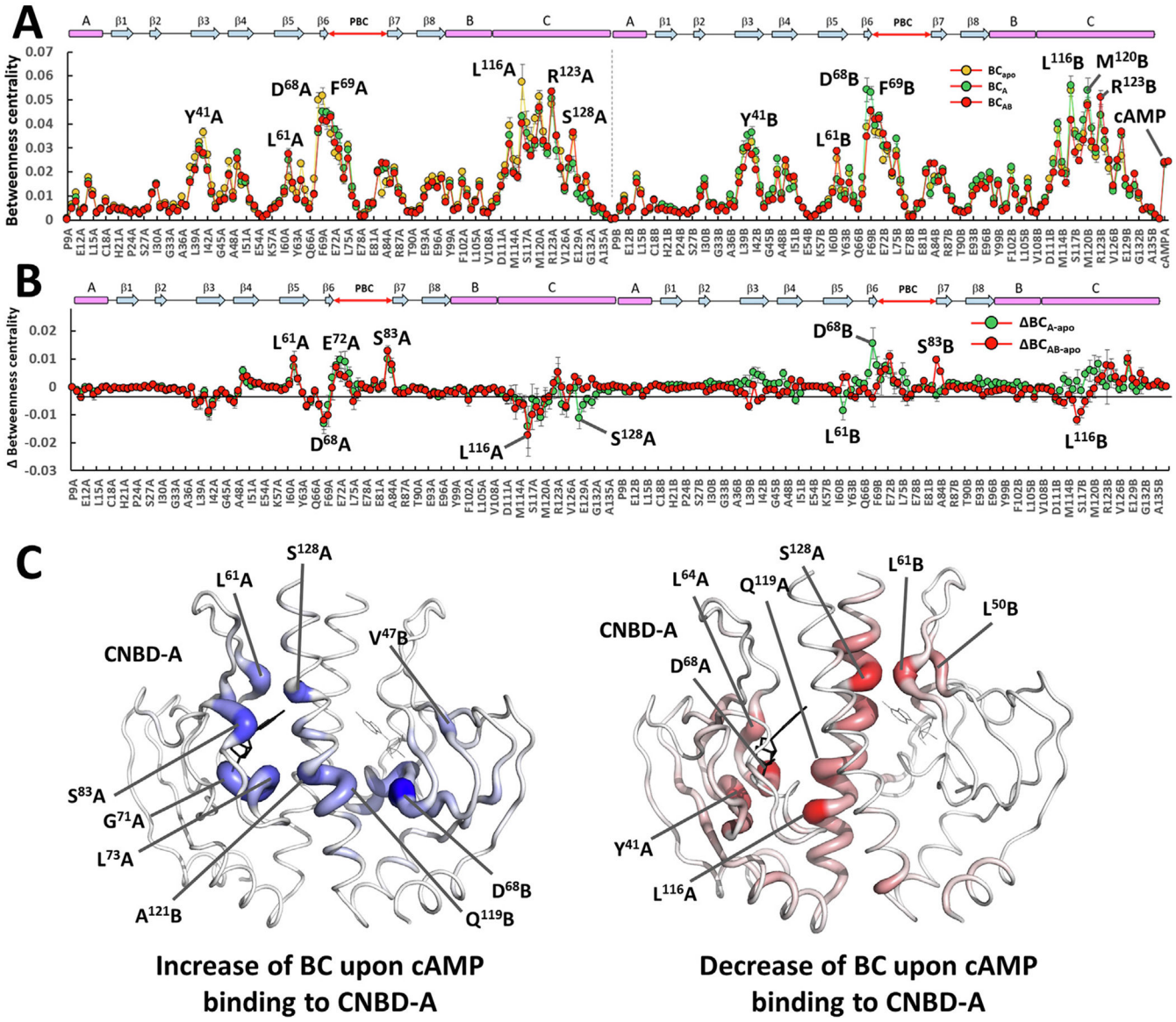

Figure 5. Changes in Betweenness Centrality in CAP LSP-Based PRNs during cAMP binding.

A) Absolute values of BC for CAP homodimer, including cAMP molecules. The secondary structure of CAP is shown as a cartoon on the top. **B)** Changes in BC upon binding of the first cAMP to CNBD-A ($\Delta BC_{A\text{-}apo}$) and two molecules of cAMP ($\Delta BC_{AB\text{-}apo}$). Standard errors of the mean for six measurements are shown. **C)** Positive (left) and negative (right) values of $\Delta BC_{A\text{-}apo}$ shown on the panel B (green dots) mapped on the CAP structure. Positive $\Delta BC_{A\text{-}apo}$ show global connectors newly formed upon the cAMP binding, negative $\Delta BC_{A\text{-}apo}$ show disruption of global connections upon this event. Two cAMP binding sites are indicated by black lines.

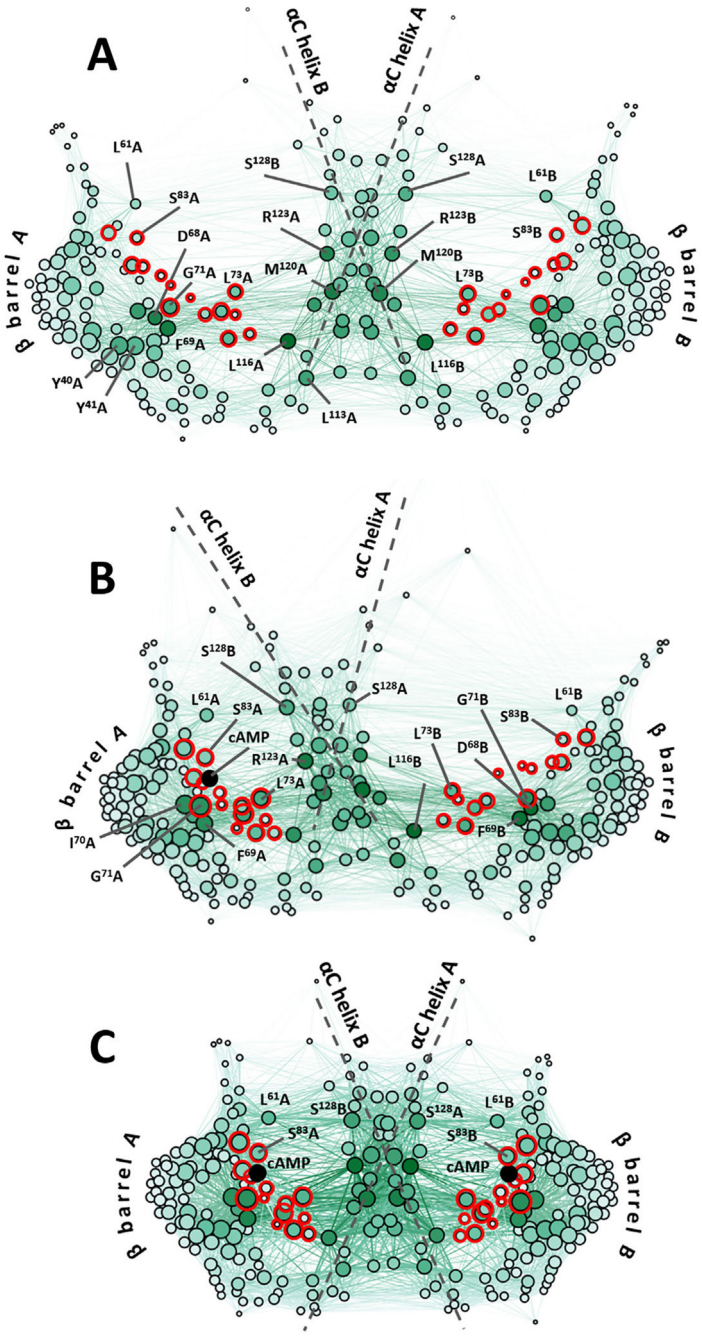

Figure 6. LSP-based PRNs visualized using ForceAtlas2 layout algorithm.

A) Apo state. **B)** cAMP bound to CNBD-A. **C)** Two cAMP molecules bound. Nodes' diameter is proportional to their DC values, darker color corresponds to higher BC values. PBC residues are shown as red circles, cAMP molecules are represented by black circles.