Deciphering general characteristics of residues constituting allosteric communication paths

Girik Malik[1] Anirban Banerji[1] Andrzej Kloczkowski[1,2]

[1]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital [2]Department of Pediatrics, College of Medicine, The Ohio State University

Introduction

- There has been great interest and excitement in studying proteins' dynamics and structural changes involved in function to gain insights into the machinery of proteins.
- Due to difficulty in retaining atomic details in mode decomposition of large system dynamics, there have been significant computational challenges, which make the study of large system dynamics very complex.
- Considering all the PDB annotated allosteric proteins (from ASD, the allosteric database) belonging to four different classes, this work has attempted to decipher certain consistent patterns present in the residues constituting the allosteric communication sub-system (ACSS).
- Despite having different sequences and different lengths of ACSS, they were found to be structurally quite similar to each other – suggesting a preferred structural template for allosteric communication.
- Though Cliques and Communities could be identified within the ACSS, maximal-common-subgraph, considering all the ACSS, could not be generated, primarily due to the diversity in the dataset.
- ACSS, in general, did not demonstrate "small-world" behavior, but the subgraphs in certain cases were found to be small world networks (SWNs).

Materials

- The curated allosteric database(ASD)(Huang et al,2014) was used to retrieve protein structures with information about the identified allosteric communication paths.
- Cases with differences in the description of protein structure provided by ASD and PDB, were not considered for this study.
- Retaining the typification scheme provided by ASD, the finally selected set of 30 proteins were typified in four groups – Kinases, Nuclear Receptors, Peptidases, Transcription Factor.
- The PDB ids of these 30 proteins are:
  - Kinases - 1CZA, 1DKU, 1E0T, 1PFK, 1S9I, 1SQ5, 2BTZ, 2JJX, 2OI2, 2VTT, 2XRW, 3BQC, 3EQC, 3F9M, 3MK6, 4AW0
  - Nuclear Receptors - 1IE9, 1XNX, 2AX6, 3S79
  - Peptidases - 1SC3, 2QL9, 4AF8
  - Transcription Factor - 1JYE, 1Q5Y, 1R1U, 1XXA, 2HH7, 2HSG, 3GZ5
- A coarse-grained (8 alphabet) representation of the residues was used, to reduce the complexity of the problem:
  - Acidic(GLU and ASP), Basic(ARG, LYS and HIS), Amides(GLN and ASN), Hydroxyls(SER and THR), Hydrophobic(TRP, TYR, PHE, MET, LEU, ILE and VAL), Small Residues(GLY and ALA). PRO and CYS are special among the 20 amino acids for obvious reasons.

Methods

- Given the central theme to measure the extent to which residues involved in allosteric communication differ from those which do not, various tests were conducted throughout the study, between residues constituting ACSS and those constituting non-ACSS.
- Python's NetworkX was used as the primary graphing library, Matplotlib was used for image generation and igraph was used to investigate cliques and communities.
- For investigation of "small world network" characteristics, methodologies of Humphries et al, 2008, were implemented.
- To characterize ACSSs as SWNs or not, at multiple resolutions, we generated Erdös-Rényi (E-R) random graphs at three probabilities: 0.3, 0.5, 0.7
- The network was constructed with a Euclidean distance threshold of 6.5Å.

Results

- Basic residues are more frequent in ACSS than acidic residues.
- ACSS structures, despite being from different proteins and despite having different sequences and lengths, demonstrated lower RMSD values and significantly higher AIC values.
- The magnitude of average degree centrality is consistently lower in ACSS, suggesting that they are shielded from many forms of perturbations (say, mutation of a residue somewhere else in the protein) which may destabilize the protein residual interaction network.
- The ACSS networks studied here suggest that the allosteric signal communication may not take place along the shortest path.
- ACSS modules can be partitioned into cliques and communities.
- ACSS may not display small world characteristics as a whole, but may display when portioned into sub-graphs.
- Barring PDB ID 1XNX (Nuclear Receptor), structural graphs of both ACSS and non-ACSS were fragmented into >1 sub-graph for displaying SWN characteristics.

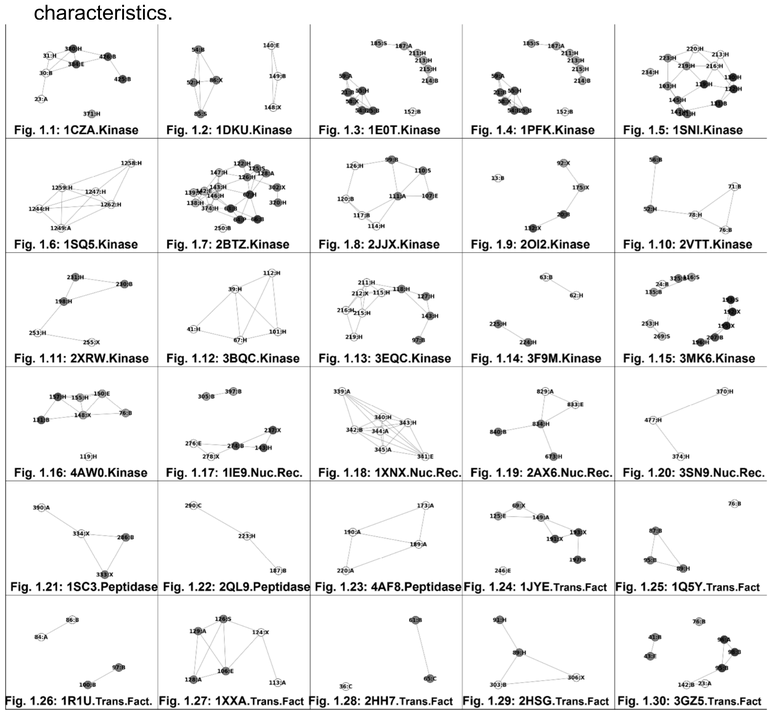

***Figure 1**: Cliques and Communities in ACSS networks. Cliques and communities found in 30 ACSS under study. Residues are identified in the format [Residue-Number in a protein:The coarse-grained character of the residue]. The coarse-grained character labeling scheme employed is: Acidic-A, Basic-B, Cysteine-C, Proline-P, Hydrophobic-H, Hydroxyl-X, Amide-E, Small-S.*

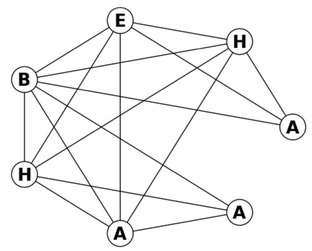

***Figure 2**: The SWN of PDB ID 1XNX's ACSS. The entire scope of ACSS of 1XNX was found to generate a SWN. Details of the SWN can be found in Table 3. The coarse-grained character labeling scheme employed is: Acidic-A, Basic-B, Cysteine-C, Proline-P, Hydrophobic-H, Hydroxyl-X, Amide-E, Small-S.*

	Kinase ACSS Residues	Kinase non- ACSS Residues	Peptidase ACSS Residues	Peptidase non-ACSS Residues	Nuclear receptor ACSS Residues	Nuclear receptor non- ACSS Residues	Transcription Factor ACSS Residues	Transcription Factor non- ACSS Residues
Average Degree Centrality	0.411	0.518	0.722	0.814	0.613	0.690	0.426	0.659
Average Closeness Centrality	0.488	0.598	0.809	0.843	0.707	0.754	0.474	0.719
Average Betweenness Centrality	0.075	0.122	0.194	0.038	0.137	0.141	0.062	0.119

***Table 1**: The three types of Centrality Indices for ACSS and non-ACSS graphs of same size for four groups of allosteric proteins.*

	Kinase		Nuclear Receptors		Peptidases		Transcription Factor	
	ACSS	Non-ACSS	ACSS	Non-ACSS	ACSS	Non-ACSS	ACSS	Non-ACSS
Global Clustering Coefficient Average	0.144	0.558	0.209	0.593	0.0	0.521	0.0	0.584
Global Average Shortest Path Length	1.228	7.396	0.976	6.586	1.0	6.136	1.119	5.042

***Table 2**: Comparison of Global Clustering Coefficient Average and Global Average Shortest Path Length for ACSS and non-ACSS.*

Group	Protein	Is it a Small world for Erdös-Rényi random graph			Nodes in Graph	Number of Nodes	Number of Edges	Average Clustering Coefficient	Average Shortest Path Length
		P=0.3	P=0.5	P=0.7					
Group-2_NuclearReceptors/ Fig-3_NuclearReceptors_ONLY_AlloRes	1XNX	N	Y	Y	['A', 'D', 'E', 'B', 'D', 'A', 'A']	7	16	0.838	1.238

***Table 3**: Summary of SWN characteristics for ACSS residues of 1XNX.*

Group	Protein	Is it a Small world for Erdös-Rényi random graph			Nodes in Graph	Number of Nodes	Number of Edges	Average Clustering Coefficient	Average Shortest Path Length
		P=0. 3	P=0. 5	P=0. 7					
Kinases	2BTZ _subset	Y	Y	N	['H', 'X', 'E', 'H', 'H', 'H', 'H']	7	10	0.59	1.619
	2JJX(A)_subset_1	Y	Y	N	['E', 'S', 'A']	3	3	1	1
	2JJX(A)_subset_2	Y	Y	N	['B', 'H', 'B']	3	2	0	1.333
	3EQC_subset	Y	Y	N	['H', 'H', 'X', 'H', 'H', 'H', 'H']	7	8	0.571	2.048
	3MK6(A)_subset	Y	N	N	['X', 'H', 'B']	3	2	0	1.333
Nuclear Receptors	2AX6_subset	N	Y	N	['E', 'H', 'A']	3	2	0	1.333
Peptidases	No SWN was observed even among the sub-graphs of residues constituting the peptidase ACSS.								
Transcription Factor	No SWN was observed among the sub-graphs of residues constituting the Transcription Factor ACSS								

***Table 4**: Summary of SWN characteristics in the sub-graphs of ACSSs considered in the present dataset.*

Discussion & Conclusion

- The main aim was to decipher some general patterns of distinct set of residues which formed the ACSS of 30 allosteric proteins.
- The intent was not on explaining the mechanism of allosteric communication in any of these cases in singularity.
- The work can be beneficial to those attempting to decipher general mechanisms of allosteric long-distance communication within protein structures.
- The structural and topological nature of ACSS have come to light.
- SWN and betweenness centrality indicate the need for more focused studies.

References

Huang Z, Mou L, et al. ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res. (2014); 42(Database issue):D510-D516

Humphries MD, Gurney K. Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence, PloS One (2008).

Monod J, Wyman J, Changeux J-P. On the nature of allosteric transitions: a plausible model. J. Mol. Biol. (1965); 12:88−118.