Deciphering General Characteristics of Residues Constituting Allosteric Communication Paths

Girik Malik[1,2], Anirban Banerji[1], Maksim Kouza[1,3], Irina A. Buhimschi[2,4], and Andrzej Kloczkowski[1,4,5(✉)]

[1] Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA
Andrzej.Kloczkowski@nationwidechildrens.org
[2] Center for Perinatal Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA
[3] Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland
[4] Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA
[5] Future Value Creation Research Center, Nagoya University, Nagoya, Japan

Abstract. Allostery is one of most important processes in molecular biology by which proteins transmit the information from one functional site to another, frequently distant site. The information on ligand binding or on posttranslational modification at one site is transmitted along allosteric communication path to another functional site allowing for regulation of protein activity. The detailed analysis of the general character of allosteric communication paths is therefore extremely important. It enables to better understand the mechanism of allostery and can be used in for the design of new generations of drugs.

Considering all the PDB annotated allosteric proteins (from ASD - AlloSteric Database) belonging to four different classes (kinases, nuclear receptors, peptidases and transcription factors), this work has attempted to decipher certain consistent patterns present in the residues constituting the allosteric communication sub-system (ACSS). The thermal fluctuations of hydrophobic residues in ACSSs were found to be significantly higher than those present in the non-ACSS part of the same proteins, while polar residues showed the opposite trend.

The basic residues and hydroxyl residues were found to be slightly more predominant than the acidic residues and amide residues in ACSSs, hydrophobic residues were found extremely frequently in kinase ACSSs. 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 communication. ACSS structures recorded low RMSD and high Akaike Information Criterion (AIC) scores among themselves. While the ACSS networks for all the groups of allosteric proteins showed low degree centrality and closeness centrality, the betweenness centrality magnitudes revealed nonuniform behavior. 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. Barring one particular case, the entire ACSS for any class of allosteric proteins did not

I. Rojas et al. (Eds.): IWBBIO 2019, LNBI 11466, pp. 245–258, 2019.
https://doi.org/10.1007/978-3-030-17935-9_23

demonstrate “small world” behavior, though the sub-graphs of the ACSSs, in certain cases, were found to form small-world networks.

Keywords: Allosteric communication sub-system · B-factor of allosteric residues · Cliques and communities · Closeness centrality · Betweenness centrality · Maximum-common-subgraph · Small-world network

1 Introduction

Starting from Monod−Wyman−Changeux [1] and Koshland−Némethy−Filmer [2] models, investigations of allosteric regulation of protein function have over half-a-century long, rich and multifaceted history. There are so many excellent reviews that have attempted to capture the essence of various aspects of research [3–8]. To summarize these efforts, one can merely observe that while a lot has been unearthed about the physicochemical nature of allosteric signal transduction, the various modes through which the long-distant communication is achieved, the structural details of cooperativity revealed during this process, there are still significant aspects of allosteric regulation, especially in the context of generalized characterization of the process, that need to be better understood. The present work reports a few generalized findings about the allosteric communication.

Because the allosteric communication paths are constituted by a certain subset of residues, we attempted in the present work, to provide a quantifiable difference between the residues involved in allosteric communications and those which are not involved. Because of that, our study revolved principally around identifying the statistical and graph-theoretical differences between the two aforementioned set of residues. We tried to decipher some consistent patterns embedded latently in structural, biophysical and topological nature of allosteric communication sub-structures (ACSS).

We focused also on the analysis of mobilities of residues forming ACSS. We compared protein fluctuations derived from crystallographic Debye-Waller B-factors of experimentally solved crystal structures with those obtained from the root mean square fluctuations (RMSF) profile from computational modeling.

We were interested to know whether the sub-structures of the allosteric communication paths have structural similarities among themselves, so that the kinase’s allosteric communication paths will be characterized by a certain set of canonical parameters, while the nuclear receptor’s allosteric communication paths will be different by certain (structural) degrees, etc.

2 Materials

The curated database ASD (Allosteric Database) [9] was used to retrieve protein structures with information about the identified allosteric communication paths. Cases with differences in the description of protein structures provided by the ASD and PDB were not considered for the study. Retaining the typification scheme provided by the ASD, the finally selected set of 30 proteins were further divided in four groups:

kinases, nuclear receptors, peptidases and transcription factors. The PDB IDs of these 30 proteins are: 1CZA, 1DKU, 1E0T, 1PFK, 1S9I, 1SQ5, 2BTZ, 2JJX, 2OI2, 2VTT, 2XRW, 3BQC, 3EQC, 3F9M, 3MK6, 4AW0 (kinases); 1IE9, 1XNX, 2AX6 3S79 (nuclear receptors); 1SC3, 2QL9, 4AF8 (peptidases); and 1JYE, 1Q5Y, 1R1U, 1XXA, 2HH7, 2HSG, 3GZ5 (transcription factors).

3 Methodology

Resorting to a coarse-grained representation of residues, and a reduced amino acid alphabet is more likely to lead to generalized ideas from the investigation of the ACSS of 30 proteins. A mere two-letter hydrophobic-polar classification of the residues would have been too broad to reveal the complexity of the problem. Thus we resorted to a scheme [10, 11] which has been found to be extremely successful in protein structure prediction studies [12, 13]. Here the 20 amino acids are expressed with a reduced 8-letter alphabet scheme; that is: GLU and ASP - as acidic, ARG, LYS and HIS - as basic, GLN and ASN - as amides, SER and THR - as hydroxyls, TRP, TYR, PHE, MET, LEU, ILE and VAL - as hydrophobic, and GLY and ALA - as small residues. PRO and CYS are special among the 20 amino acids because of their special status; each one of them are placed as singleton groups. This coarse-grained description was used to study both population characteristics of the ACSS constituents and to undertake the network-based investigations of ACSS.

Various tests were conducted throughout the study to compare residues constituting ACSS with non-ACSS ones, to measure the extent by which residues involved in allosteric communication differ from all remaining residues. Atoms of the residues not identified (viz., color-coded) by the ASD as part of allosteric communication paths, were considered to be non-ACSS residues and atoms.

We used THESEUS 2.0 software [14] that superposes multiple protein structures without throwing away gaps in them and without causing significant information loss.

For network-based and complex network studies, Python's NetworkX was used as the graphing library, while matplotlib was used for image generation. Apart from these, Python's igraph was used to investigate cliques and communities.

Because statistical tests are necessary to categorically establish the general traits in the allosteric proteins and yet, because the present study considers a limited set of allosteric proteins as belonging to different four classes two non-parametric tests (Wilcoxon signed-rank test and Friedman's non-parametric test) [15–17] were employed to ascertain the traits of the obtained results.

In investigating the "small world network" characteristics, methodologies elaborated in [18] were implemented by us; details about the theoretical basis of the methodology, thus, can be found there. To gather the answer to the question of whether or not the ACSSs are SWNs or not, at multiple resolutions, we studied the problem by generating the Erdös-Rényi (E–R) random graph at three probabilities: 0.3, 0.5 and 0.7.

In our work, the CABS-flex method [19] was used for predicting protein fluctuations. CABS-flex employs a coarse grained CABS model [20] - efficient and versatile tool for modeling protein structure, dynamics and interactions [21–24]. Conformations obtained by CABS-flex simulations further can be reconstructed to physically sound atomistic

systems using coarse-grained to atomistic mapping methods [21, 25]. The interactions between atoms are described by a realistic knowledge-based potential, while protein-solvent interactions are approximated using implicit solvent model [20, 26].

4 Results

4.1 The Thermal Fluctuation of Residues in Allosteric Communication Paths

Alongside the dynamics needed to ensure the propagation of the structural signal through protein, the ACSS residues possess their inherent thermal fluctuational dynamic. The construction of the residual-interaction networks depends on the value of the cutoff distance, that may be different than the commonly used value of 6.5 Å [27]. To quantify the extent of fluctuations of the ACSS residues, versus fluctuations of non-ACSS residues, we extracted B-factors from the coordinate files of the protein structures in protein data bank (PDB) [28] for all 30 proteins. Table 1 contains the details of this investigation.

To assess whether and by what extent the B-factors of different families of allosteric proteins differ from each other, we subjected the mean values to Friedman's non-parametric test (alternatively referred to as 'non-parametric randomized block analysis of variance') [15, 16]. We chose to employ Friedman's test because, ANOVA requires the assumptions of a normal distribution and equal variances (of the residuals) to hold, none of which is found to be existing in our case (viz., that in Table 1), while Friedman test is free from the aforementioned restrictions. The null hypothesis for the test was that the B-factors of the four types of ACSS are the same across repeated measures. Result obtained from the test categorically demonstrates that there indeed exists a substantial difference in the B-factors of these four classes of ACSSs. Results obtained from B-factors of four types of ACSS was Freidman $X^2 = 20.4 > 16.266$ (P value at 0.001, with 3 degrees of freedom), whereby the null hypothesis was rejected comprehensively.

To ascertain the degree to which the B-factors of ACSS residues in each of the four classes of allosteric proteins differ from the B-factors of the non-ACSS residues, each of the classes were subjected to Wilcoxon signed rank test (36), which is a non-parametric analogue of paired t-test for correlated samples, without assuming that the population is normally distributed. The null hypothesis for each of the comparisons was that the median difference between pairs of observations is zero. Result obtained from the tests revealed that the B-factors of ACSS residues in each of the classes differed significantly than the B-factors of the non-ACSS residues. For the kinase class of allosteric proteins we found, $W_{kinase} = 87 \gg 23$ ([$W(\alpha = 0.01,\ 17) = 23$); for the peptidase class $W_{peptidase} = 8 > 2$ ([$W(\alpha = 0.05, 7) = 2$] (we note that W is not defined in 0.01 at degrees of freedom 7 (though $W(0.01,\ 8) = 0$), whereby, the critical value comparison is being reported at the weaker 0.05 level); for the Nuclear Receptors, $W_{NR} = 15 > 5$ ([$W(\alpha = 0.01, 11) = 5$]); and for the transcription factors, $W_{TF} = 6 > 5$ ([$W(\alpha = 0.01,\ 11) = 5$]). Thus, the null hypothesis was rejected in each of the four cases with extremely high confidence.

Table 1. B-factor of residues constituting ACSS and those constituting non-ACSS

Kin. ACSS		Kin. non-ACSS		Pept. ACSS		Pept. non-ACSS		N.R. ACSS		N.R. non-ACSS		T.F. ACSS		T.F. non-ACSS	
Res.	B-factor	Res.	B-factor	Res.	B-factor	Res.	B-factor	Res.	B-factor	Res.	B-factor	Res.	B-factor	Res.	B-fact.
ALA	36.53(24.21)	ALA	30.34(17.73)			ALA	20.89(14.36)			ALA	32.89(22.47)	ALA	25.94(4.15)	ALA	37.71(19.94)
ARG	35.21(21.93)	ARG	37.69(21.34)	ARG	25.69(13.14)	ARG	23.68(16.86)	ARG	27.79(17.02)	ARG	46.01(24.03)	ARG	46.38(17.76)	ARG	46.82(22.25)
ASN	59.04(30.24)	ASN	36.84(19.79)			ASN	24.04(16.16)	ASN	13.86(5.30)	ASN	44.61(23.59)	ASN	28.83(19.51)	ASN	52.78(24.03)
ASP	35.64(13.31)	ASP	38.09(21.57)	ASP	14.62(3.92)	ASP	26.64(19.43)			ASP	40.96(24.86)	ASP	38.26(12.27)	ASP	42.54(21.98)
		CYS	29.03(16.94)	CYS	35.69(1.38)	CYS	20.16(14.56)			CYS	38.71(27.67)	CYS	52.52(26.14)	CYS	28.70(13.13)
GLN	28.44(9.25)	GLN	39.92(22.96)			GLN	25.74(21.61)	GLN	54.70(5.09)	GLN	35.91(20.80)	GLN	36.90(4.65)	GLN	39.24(20.99)
GLU	39.59(17.93)	GLU	38.99(20.70)	GLU	15.95(4.99)	GLU	32.80(18.88)	GLU	51.66(18.04)	GLU	47.97(26.94)			GLU	53.41(24.19)
GLY	27.87(9.56)	GLY	32.86(17.22)			GLY	22.80(16.68)			GLY	47.41(27.43)			GLY	39.62(22.36)
HIS	44.97(20.35)	HIS	35.71(20.97)			HIS	27.35(20.58)	HIS	13.88(2.28)	HIS	37.59(23.26)	HIS	37.62(18.24)	HIS	34.54(21.18)
ILE	42.26(16.44)	ILE	29.78(15.62)			ILE	21.97(15.70)			ILE	44.0(28.05)			ILE	38.45(19.11)
LEU	40.69(19.53)	LEU	29.97(16.05)			LEU	19.34(13.89)	LEU	54.75(5.83)	LEU	34.14(23.81)			LEU	40.07(20.85)
LYS	38.72(11.67)	LYS	38.81(20.22)			LYS	31.09(18.67)			LYS	46.38(27.14)	LYS	47.32(3.57)	LYS	53.42(23.40)
MET	43.12(22.56)	MET	31.10(15.56)			MET	18.20(14.14)	MET	59.16(0.80)	MET	41.21(24.96)			MET	42.64(23.60)
PHE	28.48(6.08)	PHE	29.68(16.62)			PHE	22.49(15.94)	PHE	16.25(2.38)	PHE	37.18(23.42)			PHE	41.84(23.56)
		PRO	33.25(17.87)			PRO	21.53(16.09)			PRO	39.67(25.50)			PRO	41.43(19.17)
SER	34.86(13.03)	SER	33.41(19.38)	SER	11.33(0.60)	SER	23.36(17.02)	SER	8.37(0.82)	SER	38.91(26.37)	SER	17.47(1.97)	SER	41.46(21.70)
THR	45.03(27.45)	THR	32.29(18.18)	THR	12.35(0.40)	THR	20.91(15.51)			THR	42.18(25.24)	THR	51.90(20.00)	THR	40.01(19.26)
		TRP	31.22(23.75)	TYR	36.82(1.73)	TRP	27.45(22.24)			TRP	44.96(32.35)			TRP	32.68(13.28)
TYR	52.48(24.13)	TYR	31.12(16.42)			TYR	22.70(18.29)	TYR	11.47(3.65)	TYR	43.32(28.58)	TYR	67.23(2.49)	TYR	39.88(18.00)
VAL	38.10(23.12)	VAL	28.78(14.69)			VAL	19.41(15.28)	VAL	57.18(0.37)	VAL	38.22(24.76)			VAL	37.22(21.24)

B-factors were calculated at the residual level in ACSS and non-ACSS, they are presented as Mean (Std. Dev.)

4.2 Robustness of the Results Against CABS-Flex Simulations

The computational modelling is a key to solving many fundamental problems of molecular biology. Prediction of protein structures and interactions [29] as well as structural transformations taking places during unfolding, folding and aggregation processes have been studied by computer simulations at different levels of resolution and timescales [30–38]. For more efficient simulations one uses coarse-grained (CG) models which reduce the complexity of each amino acid by representing it by a single node or group of pseudo atoms [20, 29, 39, 40].

In order to address the question whether B-factors extracted from PDB file are consistent with root mean square fluctuations (RMSF) of atoms from simulations, limited simulations have been carried out with the help of CABS-flex simulations [19]. We computed the values of RMSF for three conceptually different proteins, 1Q5Y from transcription factors group, 1SC3 from the peptidases and 2JXX from the kinase group of allosteric proteins. RMSF profiles are shown as red curves (on right Y-axis) on middle plots in Figs. 1a, b and c. The values of B-factors are shown as black curves (on left Y-axis).

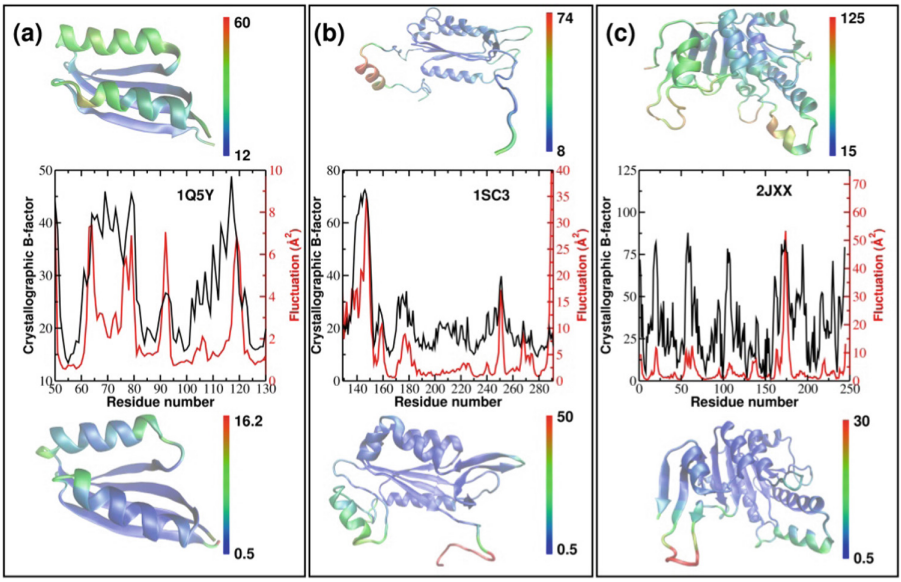

Fig. 1. Comparison of RMSF values with B factors for 3 different proteins: 1Q5Y (a), 1SC3 (b) and 2JXX (c). (Color figure online)

Although quantitative comparison between B-factors and RMSFs is not possible due to different temperatures and environmental factors used in simulations and experiments, qualitatively the data agree. The most fluctuating protein residues during near-native state simulations result in a series of peaks in RMSF profile (red curve, right X-axis) which correlate with experimentally measured B-factor values (black curve, left Y-axis).

Upper and bottom snapshots in Fig. 1 correspond to protein representation colored by crystallographic B-factor values from PDB and by RMSF values from CABS-flex simulation, respectively. The overall trend is that the mobility of atoms obtained from simulations are in reasonable agreement with the crystallographic B-factors.

4.3 Composition of the ACSS Population

Allosteric signalling achieved at the structural level show certain differences for various proteins [41–43]. Thus, we expect to observe differences in composition of ACSS residues for four different classes of proteins. We found that basic residues are more frequent in ACSS than the acidic ones. To demonstrate this prevalence let us take a closer look at the composition of ACSSs for kinases: the acidic residues were found in 12/133 cases, while the basic residues occurred in 32/133 cases. For the class of transcription factors the basic residues in ACSSs occurred in 13/35 cases, whereas the acidic residues occurred in 5/35 cases. The hydrophobic residues were found to occur in ACSSs of kinases with significant frequency (59/133 cases), but were be notably small in ACSS of transcription factors (2/35) and in peptidases (1/11).

Hydroxyl residues were found to be more common in ACSSs than the amide residues, for kinase ACSS: amide residues 6/133, and hydroxyl residues 14/133 cases. PRO and CYS populations although are extremely small in ACSS, show that CYS occurs slightly more frequently than PRO. The small amino acids (GLY and ALA) were found in very small frequency in ACSSs, while TRP was not found as part of any of the ACSSs.

4.4 Structural Superimposition of Multiple Allosteric Communication Paths

Results obtained from the structural superimposition of multiple ACSSs demonstrated clearly that the allosteric communication paths, for any type of allosteric protein, match closely each other in their structures. We superimposed the PDB-coordinates of ACSSs of all proteins for each of the four classes using ‘Theseus’ software. Here we report the two most prominent results, a: RMSD for the superposition, and b: the Akaike Information Criterion (AIC). AIC proposed by Akaike [44] has become commonly used tool for statistical comparison of multiple theoretical models characterized by different numbers of parameters. Because the RMSD of two superposed structures indicates their divergence from one another a small value is interpreted as a good superposition. In contrast, the higher magnitude of AIC indicates better superposition. We found that the ACSS paths, despite belonging to different proteins and corresponding to sequences of varying lengths, consistently demonstrated lower RMSD values and significantly higher AIC values in comparison to non-ACSSs parts of the structures.

4.5 Network Analyses of Allosteric Communication Paths

4.5.1 Centrality of ACSS

The process of allosteric signal communication is directional, but the richness of the constructs available to study networks becomes apparent by using non-directional graph-theoretical framework. Thus, instead of asking 'what is the route of allosteric signal propagation for a specific protein?', which is already provided by ADB, we asked questions like: 'how robust the ACSSs are, compared to non-ACSS parts of the proteins?', or, 'how does the fluctuation of one arbitrarily-chosen residue influence the spread of allosteric signal through ACSS?', or, 'how probable is it that allosteric communication occurs through a randomly chosen shortest path between two residues belonging to ACSS?', etc.

To answer these and similar questions of general nature, we started our investigation by studying the centrality aspects of the ACSS network. The centrality metrics quantify the relative importance of a protein residue (viz. the vertex) or a residue-to-residue communication path (viz., an edge) in the network description of ACSS. There are many centrality measures, we chose to concentrate upon three fundamental measures outlined in Freeman's classic works [45, 46], namely: degree centrality, betweenness centrality and closeness centrality.

4.5.2 Degree Centrality

Degree centrality for any protein residue in an ACSS network is calculated in a straightforward way, by counting the number of residue-residue communication links connecting that residue (implementing the classical definition [45] to the context of ACSS). Degree centrality of any ACSS residue provides an idea about the local structure around that residue, by measuring the number of other residues connected to it. We note that degree centrality is a local measure that does not provide any information about the network's global structure. We have found that the average degree centrality of ACSS residues, irrespective of the type of allosteric proteins, is lower than the average degree centrality of the non-ACSS residues. This result, alongside that obtained from the other centrality measures are presented in Table 2.

Table 2. The centrality indices for ACSS and non-ACSS for four groups of allosteric proteins.

	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	0.596	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

The three types of major centrality measures calculated on the ACSS and non-ACSS graphs of the same size.

We note that the average degree centrality of ACSS fragments consistently show lower values than similar non-ACSS fragments. We note also that the typical differences between average degree centrality of ACSS and non-ACSS fragments are: $\sim$0.10 for kinases, $\sim$0.8 for nuclear receptors, $\sim$0.9 for peptidases, and $\sim$0.13 for transcription factors. Thus, the consistently lower values of average degree centrality observed in ACSS fragments suggests that nature attempts to shield them from perturbations which may destabilize allosteric communication.

4.5.3 The Global Centrality Measures

While the degree centrality provides a measure to assess the possibility of immediate involvement of a residue in influencing the signal communication in residue interaction network of a protein, the concepts of closeness centrality and betweenness centrality provide ideas of how the global topology of the network influences the signal propagation. Closeness centrality of any connected graph measures how "close" a vertex is to other vertices in a network; this is computed by summing up the lengths of the shortest paths between that vertex and other vertices in the network. Closeness of a vertex, thus, can be interpreted as a predictor of how long it may take for that vertex to communicate with all other vertices. In the framework of protein residue connectivity network, the residues with low closeness score can be identified as ones that are separated by short distances from other residues. It can be expected that they receive the structural signal (i.e. instantaneous fluctuation or perturbation) faster, being well-positioned to receive this information early. We indeed found that the average closeness centrality of the ACSS network is lower in comparison to the non-ACSS fragments, for all the types of allosteric proteins. However, the difference between the extent of average closeness centrality between ACSS and non-ACSS fragments was found to vary over a larger scale than what was observed for average degree centrality (see Table 2).

The betweenness centrality provides more idea about the global network structure; for every vertex of the network the betweenness centrality specifies the fraction of the shortest paths (geodesics) that pass through that vertex. In this sense, such measure assesses the influence that a given vertex (residue) has over the transmission of a structural signal. A residue with large betweenness centrality score can be expected to have a large influence on the allosteric signal propagating through the ACSS network. Results obtained by us shown in Table 2.

4.6 Cliques and Communities in ACSS

Cliques are the complete subgraphs, where every vertex is connected to every other vertex. A clique is considered maximal only if it is not found to be a subgraph of some other clique. Communities are identified through partitioning the set of vertices, whereby each vertex is made a member of one and only one community. Because of their higher order connectivity, the cliques detected in protein structures are considered to indicate regions of higher cohesion (in some cases, rigid modules). Do the ACSSs embody certain common characteristics in their connectivities which can be revealed through the cliques and communities? To answer this question, we subjected the ACSSs of each of the four classes of proteins to investigation, which implemented [47]

and [48] algorithms through the Python-igraph package. We found that indeed the ACSS modules can be partitioned into cliques and communities.

4.7 Maximum Common Subgraphs to Describe the ACSS

A maximal common subgraph of a set of graphs is the common subgraph having the maximum number of edges. Many attempts have been made for the last two decades to apply this methodology in protein science [49–51]. Finding the maximal common subgraph is a NP-complete problem [52]. To solve this difficult problem a backtrack search algorithm proposed by McGregor [53] and a clique detection algorithm of Koch [52], are traditionally used. However, for our ACSSs, some of which are quite large in size, neither McGregor's nor Koch's algorithm was found to be applicable; primarily because of the huge computational costs incurred by the exponential growth of intermediary graphs of varying sizes. Thus, upon generating the subgraphs for each of ACSSs (using Python's NetworkX), we had to resort to the brute-force method to identify the maximum common subgraph for each of the ACSS classes. In some cases, the number of cliques was found to be large; e.g. for 2BTZ while in some other cases only one clique was found (e.g. for 2VTT or for 2XRW).

4.8 How Frequently Do the Allosteric Communication Paths Form Small World Network?

Investigating whether in general the ACSS residues belonging to the four different classes of allosteric proteins constitute 'small world' networks (SWN) or not is important; because SWNs are more robust to perturbations, and may reflect an evolutionary advantage of such an architecture [54, 55]. The SWN [56], constitute a compromise between the regular and the random networks, because on one hand they are characterized by large extent of local clustering of nodes, like in regular networks, and on the other hand they embody smaller path lengths between nodes, something that is distinctive for random networks. Because of the ability to combine these two disparate properties, not surprisingly, it has been shown that networks demonstrating the 'small-world' characteristics tend to describe systems that are characterized by dynamic properties different from those demonstrated by equivalent random or regular networks [56–61]. We have found that whether ACSSs exhibit SWN nature or not - is a complex problem; while the complete ACSS of a protein may not always demonstrate SWN characteristics, many sub-graphs of non-trivial lengths of the same ACSS reveal SWN character.

5 Conclusions

The aim of the present work was to decipher some general patterns of residues forming the ACSS of 30 allosteric proteins, and compare them with non-ACSS residues in the same proteins. Our aim was to report the general quantifiable differences between these two (aforementioned) sets of residues and not to study the general mechanism of allosteric communication. By performing the CABS-based simulations of proteins around their native conformations we demonstrated that protein fluctuations depicted

by RMSF profiles can be mapped to B-factors and show satisfactory degree of agreement with experimental data.

Our results may benefit the protein engineering community and those studying the general mechanism of allosteric communication or in general, long-distance communication in proteins. The knowledge of the topological invariants of communication paths and the biophysical, biochemical and structural patterns may help in a better understanding of allostery. As many recent papers [62–66] have pointed out, the long-distance communication features within proteins involve several types of non-linear characteristics that may often be dependent on transient fluctuations, making it difficult to arrive at a generalized dynamic picture. However a generalized static picture of the long-distance communication route can be obtained, which may help to better understand such communication schemes, especially those related to allostery. The present work attempted to report such generalized findings. While certain yet-unknown (to the best of our knowledge) patterns regarding the thermal fluctuation profile of ACSS atoms, the structural and topological nature of the ACSS have come to light, incongruities of our findings regarding the extent of betweenness centrality in ACSS network and their small-world nature indicates the need for more focused studies directed at these issues, which in turn, may shed new light on allosteric signal communication. For example, proteins, in general, are fractal objects with known characteristics of trapping energy [67–70]. Do the findings on betweenness and on small-world network nature reported in this work indicate the possibility of energy traps in ACSSs? - We plan to probe into many such questions in future.

Acknowledgements. The second author, Dr. Anirban Banerji, passed away in Columbus, OH on Aug. 12, 2015 at the age of 39. A.K. acknowledges support from The Research Institute at Nationwide Children's Hospital, from the National Science Foundation (DBI 1661391) and from National Institutes of Health (R01GM127701 and R01GM127701-01S1). M.K. acknowledges the Polish Ministry of Science and Higher Education for financial support through "Mobility Plus" Program No. 1287/MOB/IV/2015/0. I.A.B. acknowledges support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) R01HD084628 and The Research Institute at Nationwide Children's Hospital's John E. Fisher Endowed Chair for Neonatal and Perinatal Research.

References

1. Monod, J., Wyman, J., Changeux, J.P.: On the nature of allosteric transitions: a plausible model. J. Mol. Biol. **12**, 88–118 (1965)
2. Koshland Jr., D.E., Némethy, G., Filmer, D.: Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry **5**, 365–385 (1966)
3. Nussinov, R.: Introduction to protein ensembles and allostery. Chem. Rev. **116**, 6263–6266 (2016)
4. Ribeiro, A.A., Ortiz, V.: A chemical perspective on allostery. Chem. Rev. **116**, 6488–6502 (2016)
5. Dokholyan, N.V.: Controlling allosteric networks in proteins. Chem. Rev. **116**, 6463–6487 (2016)
6. Guo, J., Zhou, H.X.: Protein allostery and conformational dynamics. Chem. Rev. **116**, 6503–6515 (2016)

7. Papaleo, E., Saladino, G., Lambrughi, M., Lindorff-Larsen, K., Gervasio, F.L., Nussinov, R.: The role of protein loops and linkers in conformational dynamics and allostery. Chem. Rev. **116**, 6391–6423 (2016)
8. Wei, G.H., Xi, W.H., Nussinov, R., Ma, B.Y.: Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. Chem. Rev. **116**, 6516–6551 (2016)
9. Huang, Z.M., Mou, L.K., Shen, Q.C., Lu, S.Y., Li, C.G., Liu, X.Y., et al.: ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res. **42**, D510–D516 (2014)
10. Feng, Y.P., Kloczkowski, A., Jernigan, R.L.: Four-body contact potentials derived from two protein datasets to discriminate native structures from decoys. Proteins **68**, 57–66 (2007)
11. Feng, Y., Jernigan, R.L., Kloczkowski, A.: Orientational distributions of contact clusters in proteins closely resemble those of an icosahedron. Proteins **73**, 730–741 (2008)
12. Faraggi, E., Kloczkowski, A.: A global machine learning based scoring function for protein structure prediction. Proteins **82**, 752–759 (2014)
13. Gniewek, P., Kolinski, A., Kloczkowski, A., Gront, D.: BioShell-threading: versatile Monte Carlo package for protein 3D threading. BMC Bioinform. **15**, 22 (2014)
14. Theobald, D.L., Steindel, P.A.: Optimal simultaneous superpositioning of multiple structures with missing data. Bioinformatics **28**, 1972–1979 (2012)
15. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. **32**, 675–701 (1937). J. Am. Stat. Assoc. **34**, 109 (1939)
16. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bull. **1**, 80–83 (1945)
17. Friedman, M.: A comparison of alternative tests of significance for the problem of m rankings. Ann. Math. Stat. **11**, 86–92 (1940)
18. Humphries, M.D., Gurney, K.: Network 'small-world-ness': a quantitative method for determining canonical network equivalence. PLOS One **3**, e0002051 (2008)
19. Jamroz, M., Kolinski, A., Kmiecik, S.: CABS-flex: server for fast simulation of protein structure fluctuations. Nucleic Acids Res. **41**, W427–W431 (2013)
20. Kolinski, A.: Protein modeling and structure prediction with a reduced representation. Acta Biochim. Pol. **51**, 349–371 (2004)
21. Kmiecik, S., Gront, D., Kouza, M., Kolinski, A.: From coarse-grained to atomic-level characterization of protein dynamics: transition state for the folding of B domain of protein A. J. Phys. Chem. B **116**, 7026–7032 (2012)
22. Wabik, J., Kmiecik, S., Gront, D., Kouza, M., Kolinski, A.: Combining coarse-grained protein models with replica-exchange all-atom molecular dynamics. Int. J. Mol. Sci. **14**, 9893–9905 (2013)
23. Blaszczyk, M., Kurcinski, M., Kouza, M., Wieteska, L., Debinski, A., Kolinski, A., et al.: Modeling of protein-peptide interactions using the CABS-dock web server for binding site search and flexible docking. Methods **93**, 72–83 (2016)
24. Jamroz, M., Orozco, M., Kolinski, A., Kmiecik, S.: Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. J. Chem. Theory Comput. **9**, 119–125 (2013)
25. Gront, D., Kmiecik, S., Kolinski, A.: Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. J. Comput. Chem. **28**, 1593–1597 (2007)
26. Jamroz, M., Kolinski, A., Kmiecik, S.: Protocols for efficient simulations of long-time protein dynamics using coarse-grained CABS model. Methods Mol. Biol. **1137**, 235–250 (2014)
27. Sun, W.T., He, J.: From isotropic to anisotropic side chain representations: comparison of three models for residue contact estimation. PLOS One **6**, e19238 (2011)

28. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al.: The protein data bank. Nucleic Acids Res. **28**, 235–242 (2000)
29. Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E., Kolinski, A.: Coarse-grained protein models and their applications. Chem. Rev. **116**, 7898–7936 (2016)
30. Sulkowska, J.I., Kloczkowski, A., Sen, T.Z., Cieplak, M., Jernigan, R.L.: Predicting the order in which contacts are broken during single molecule protein stretching experiments. Proteins-Struct. Funct. Bioinform. **71**, 45–60 (2008)
31. Scheraga, H.A., Khalili, M., Liwo, A.: Protein-folding dynamics: overview of molecular simulation techniques. Annu. Rev. Phys. Chem. **58**, 57–83 (2007)
32. Nasica-Labouze, J., Nguyen, P.H., Sterpone, F., Berthoumieu, O., Buchete, N.V., Cote, S., et al.: Amyloid beta protein and Alzheimer's disease: when computer simulations complement experimental studies. Chem. Rev. **115**, 3518–3563 (2015)
33. Kouza, M., Co, N.T., Nguyen, P.H., Kolinski, A., Li, M.S.: Preformed template fluctuations promote fibril formation: insights from lattice and all-atom models. J. Chem. Phys. **142**, 145104 (2015)
34. Kouza, M., Banerji, A., Kolinski, A., Buhimschi, I.A., Kloczkowski, A.: Oligomerization of FVFLM peptides and their ability to inhibit beta amyloid peptides aggregation: consideration as a possible model. Phys. Chem. Chem. Phys. **19**, 2990–2999 (2017)
35. Kmiecik, S., Kouza, M., Badaczewska-Dawid, A.E., Kloczkowski, A., Kolinski, A.: Modeling of protein structural flexibility and large-scale dynamics: coarse-grained simulations and elastic network models. Int. J. Mol. Sci. **19**, 3496 (2018)
36. Kouza, M., Banerji, A., Kolinski, A., Buhimschi, I., Kloczkowski, A.: Role of resultant dipole moment in mechanical dissociation of biological complexes. Molecules **23**, 1995 (2018)
37. Kouza, M., Co, N.T., Li, M.S., Kmiecik, S., Kolinski, A., Kloczkowski, A., et al.: Kinetics and mechanical stability of the fibril state control fibril formation time of polypeptide chains: a computational study. J. Chem. Phys. **148**, 215106 (2018)
38. Lan, P.D., Kouza, M., Kloczkowski, A., Li, M.S.: A topological order parameter for describing folding free energy landscapes of proteins. J. Chem. Phys. **149**, 175101 (2018)
39. Shakhnovich, E.: Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chem. Rev. **106**, 1559–1588 (2006)
40. Liwo, A., He, Y., Scheraga, H.A.: Coarse-grained force field: general folding theory. Phys. Chem. Chem. Phys. **13**, 16890–16901 (2011)
41. Banerji, A.: An attempt to construct a (general) mathematical framework to model biological "context-dependence". Syst. Synth. Biol. **7**, 221–227 (2013)
42. Tuncbag, N., Gursoy, A., Nussinov, R., Keskin, O.: Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. Nat. Protoc. **6**, 1341–1354 (2011)
43. Ozbabacan, S.E.A., Gursoy, A., Keskin, O., Nussinov, R.: Conformational ensembles, signal transduction and residue hot spots: application to drug discovery. Curr. Opin. Drug Disc. **13**, 527–537 (2010)
44. Akaike, H.: A new look at the statistical-model identification. IEEE Trans. Autom. Control **19**, 716–723 (1974)
45. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**, 215–239 (1979)
46. Freeman, L.C., Borgatti, S.P., White, D.R.: Centrality in valued graphs - a measure of betweenness based on network flow. Soc. Netw. **13**, 141–154 (1991)
47. Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. Phys. Rev. E **74**, 016110 (2006)

48. Traag, V.A., Bruggeman, J.: Community detection in networks with positive and negative links. Phys. Rev. E **80**, 036115 (2009)
49. Grindley, H.M., Artymiuk, P.J., Rice, D.W., Willett, P.: Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. J. Mol. Biol. **229**, 707–721 (1993)
50. Koch, I., Lengauer, T., Wanke, E.: An algorithm for finding maximal common subtopologies in a set of protein structures. J. Comput. Biol. **3**, 289–306 (1996)
51. Raymond, J.W., Willett, P.: Maximum common subgraph isomorphism algorithms for the matching of chemical structures. J. Comput. Aid. Mol. Des. **16**, 521–533 (2002)
52. Koch, I.: Enumerating all connected maximal common subgraphs in two graphs. Theor. Comput. Sci. **250**, 1–30 (2001)
53. McGregor, J.J.: Backtrack search algorithms and the maximal common subgraph problem. Softw. Pract. Exp. **12**, 23–34 (1982)
54. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
55. Barabasi, A.L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**, 101–113 (2004)
56. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
57. Barahona, M., Pecora, L.M.: Synchronization in small-world systems. Phys. Rev. Lett. **89**, 054101 (2002)
58. Nishikawa, T., Motter, A.E., Lai, Y.C., Hoppensteadt, F.C.: Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? Phys. Rev. Lett. **91**, 014101 (2003)
59. Roxin, A., Riecke, H., Solla, S.A.: Self-sustained activity in a small-world network of excitable neurons. Phys. Rev. Lett. **92**, 198101 (2004)
60. Lago-Fernandez, L.F., Huerta, R., Corbacho, F., Siguenza, J.A.: Fast response and temporal coherent oscillations in small-world networks. Phys. Rev. Lett. **84**, 2758–2761 (2000)
61. del Sol, A., O'Meara, P.: Small-world network approach to identify key residues in protein-protein interaction. Proteins **58**, 672–682 (2005)
62. Kim, H., Zou, T.S., Modi, C., Dorner, K., Grunkemeyer, T.J., Chen, L.Q., et al.: A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. Structure **23**, 34–43 (2015)
63. Na, H., Lin, T.L., Song, G.: Generalized spring tensor models for protein fluctuation dynamics and conformation changes. Adv. Exp. Med. Biol. **805**, 107–135 (2014)
64. Song, G., Jernigan, R.L.: An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins **63**, 197–209 (2006)
65. Jamroz, M., Kolinski, A., Kihara, D.: Structural features that predict real-value fluctuations of globular proteins. Proteins **80**, 1425–1435 (2012)
66. Yang, Y.D., Park, C., Kihara, D.: Threading without optimizing weighting factors for scoring function. Proteins **73**, 581–596 (2008)
67. Enright, M.B., Leitner, D.M.: Mass fractal dimension and the compactness of proteins. Phys. Rev. E **71**, 011912 (2005)
68. Banerji, A., Ghosh, I.: Revisiting the myths of protein interior: studying proteins with mass-fractal hydrophobicity-fractal and polarizability-fractal dimensions. PLOS One **4**, e7361 (2009)
69. Leitner, D.M.: Energy flow in proteins. Annu. Rev. Phys. Chem. **59**, 233–259 (2008)
70. Reuveni, S., Granek, R., Klafter, J.: Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure. Proc. Natl. Acad. Sci. U.S.A. **107**, 13696–13700 (2010)