



Residue interaction network analysis of Dronpa and a DNA clamp

Guang Hu*, Wenying Yan, Jianhong Zhou, Bairong Shen*

Center for Systems Biology, Soochow University, Suzhou 215006, China



HIGHLIGHTS

- Network analysis reveals the topological principles of proteins.
- Centrality analysis recalls the relationship with protein dynamics.
- Communication pathways in a DNA clamp are predicted.

ARTICLE INFO

Article history:

Received 19 October 2013

Received in revised form

19 December 2013

Accepted 18 January 2014

Available online 28 January 2014

Keywords:

Protein topology

Network theory

Small-world

Communication pathway

ABSTRACT

Topology is an essential aspect of protein structure. The network paradigm is increasingly used to describe the topology and dynamics of proteins. In this paper, the effect of topology on residue interaction network was investigated for two different proteins: Dronpa and a DNA clamp, which have cylindrical and toroidal topologies, respectively. Network metrics including characteristic path lengths, clustering coefficients, and diameters were calculated to investigate their global topology parameters such as small-world properties and packing density. Measures of centrality including betweenness, closeness, and residue centrality were computed to predict residues critical to function. Additionally, the detailed topology of the hydrophobic pocket in Dronpa, and communication pathways across the interface in the DNA clamp, were investigated using the network. The results are presented and discussed with regard to existing residue interaction network properties of globular proteins and elastic network models on Dronpa and the DNA clamp. The topological principle underlying residue interaction networks provided insight into the architectural organization of proteins.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Proteins consist of one or more linear polymer chains that typically folded into one or more specific spatial conformations crucial to biological function. Topology and symmetry are important aspects in the inherent structural organizations. Although complex systems in their own right, protein structures adopt discrete topologies and they can be globular, cylindrical or toroidal. Residue interaction networks based on network theory can be used to analyze and characterize protein structures (Bode et al., 2007; Krishnan et al., 2008; Sun and He, 2011; Greene, 2012; Taylor, 2013). Using this approach, residues correspond to network nodes, and each interaction between residues is represented as an edge linking the corresponding nodes. Residue interaction networks have enriched chemistry with a novel paradigm (Paola et al., 2013), by extending topological descriptors from chemical graph theory to delineate ontology of proteins. These analyses can be thought of as

mathematical chemistry, and are rapidly moving towards full biological network analysis of systems biology.

Residue interaction network analysis has revealed that proteins display small-world behavior (Watts and Strogatz, 1998). Combining these small-world network properties with network parameters provided a novel approach to studying protein folding (Vendruscolo et al., 2002; Bagler and Sinha, 2007; Fang et al., 2010) and stability (Brinda and Vishveshwara, 2005), identifying functional residues (Amitai et al., 2004) and hot-spot residues (del Sol and O'Meara, 2004), mutation (del Sol et al., 2006) and allosteric communication analysis (Vishveshwara et al., 2009), detecting protein decoys (Zhou et al., in press; Zhou et al., accepted for publication), as well as understanding interactions within protein complexes (Brinda and Vishveshwara, 2005; Hu et al., 2013) and protein–protein docking (Jiao and Chang, 2011). It has also been shown that graph theory provides a basis for topological analysis of proteins. Kannan and Vishveshwara (1999) used graph spectral approach to detect a variety of side-chain clusters, and in other study, universal topological characteristics were used to describe topological cavities in proteins (Estrada, 2010). Other types of residue interaction networks based on long- and short-range interaction were constructed to

* Corresponding authors.

E-mail addresses: huguang@suda.edu.cn (G. Hu), bairong.shen@suda.edu.cn (B. Shen).

investigate more properties (Greene and Higman, 2003; Paci et al., 2012). Other network-based approaches, such as the elastic network model (ENM) (Tirion, 1996) and the Gaussian network model (GNM) (Bahar et al., 1997) treat all atomic distances weighted with harmonic potentials, forming spring networks which exploit the protein dynamics encoded in the molecular topology. In particular, Atilgan et al. (2004) demonstrated the average shortest path lengths are highly correlated with residue fluctuations, revealing an interesting link between protein topology and dynamics.

The protein backbone plays a fundamental role in determining the small-world properties (Bartoli et al., 2007). Hence, it is of particular interest to study the network properties of proteins with different connectivities. To date, network approaches have been most effectively applied to the study of globular proteins. Bagler and Sinha (2005) constructed residue interaction networks and calculated their network parameters for four types of globular proteins. Their results showed that globular proteins have small-world properties irrespective of their structure classes. More recently, this methodology has been applied to transmembrane proteins (Emerson and Gothardam, 2012a, 2012b), which show small-world properties similar to globular proteins. In this study, we restricted ourselves to two representative proteins with particular topologies. Dronpa (Wilmann et al., 2006) has a typical cylindrical topology, containing of 11 β -strands that form a β -barrel (Fig. 1a). Two central helices, α_1 and α_2 , connected by the chromophore, run through the β -barrel, and several interconnecting loops sit at the top and bottom to protect the chromophore. The other representative protein studied was the β -clamp DNA polymerase from *Escherichia coli*, which also has cylindrical symmetry but a toroidal topology (Kong et al., 1992). The β clamp is a homodimer with two almost identical monomers (Fig. 1c). The torus-shaped structure is lined by 12 β -helices forming the inner surface, and six β -sheets forming the outer surface, that together form a β -wheel. Since both Dronpa and the β clamp have special topological backbones, they are

ideal model structures with which to investigate potentially features of residues interaction networks.

In this study, the network theory was used to investigate the topological and geometrical properties of Dronpa and the β clamp. Both proteins were modeled using undirected graphs in which the amino acids are represented as nodes and their interactions as edges. The effects of protein connectivity in a cylinder and torus were investigated using various network parameters. Residue centrality was calculated by removing the node and its corresponding edges from the residue interaction network to predict central residues, and to describe the robustness of proteins. Network theory was also used to probe the hydrophobic pocket in Dronpa and the communication pathways across the interface of the β clamp. These results were compared with existing globular protein network studies and with the results of previous ENM studies on Dronpa and the β clamp.

2. Materials and methods

2.1. Residue interaction networks

Residue interaction networks are constructed by considering non-adjacent amino-acid residues or C_α as nodes, which are connected by non-covalent interactions. These graphs can be either un-weighted or weighted in which edges are defined based on predefined cutoff such as the strength of interaction (I_{min}) and/or distances. Alternatively, protein contact maps can be used to generate similar abstracted representations of protein structures. In this study, the definitions of residue interaction networks and protein contact maps were specified by two parameters. Each amino-acid residue is represented by a vertex, and edges represented contacts between atoms that had at least one Van der Waals interaction (a higher cut-off representing longer interaction

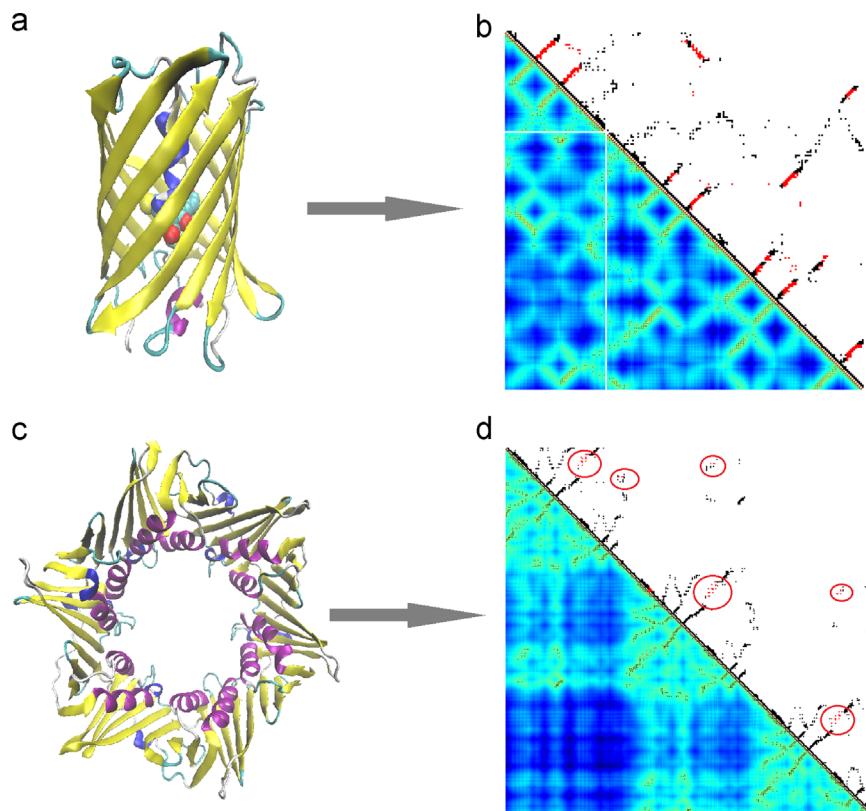


Fig. 1. The three-dimensional structures of (a) cylindrical Dronpa and its (b) corresponding protein contact maps, and (c) toroidal β clamp and (d) the corresponding protein contact maps of the monomer. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

distances utilized a London-Van der Waals forces of approximately 5 Å). Therefore, our residue interaction networks had weighted links instead of distance cut-offs (Aftabuddin and Kundu, 2006). The weight for the network is defined as the Van der Waals contact score, which corresponds to the approximate average free energy of each interaction and arose from average values from both simulations and experimental data.

2.2. Network parameters

After construction of the whole grid of links, the network can be characterized by evaluation of quantitative descriptors as previously described (Bonchev and Buck, 2007; Doncheva et al., 2012).

The *clustering coefficient* (C_i , range 0–1) is a fundamental network parameter for describing the hierarchical structure of proteins, and normalizes the number of edges between the first neighbors of the vertex i by dividing it by the maximal number of such edges. The average clustering coefficient $\langle C \rangle$ is calculated by averaging the clustering coefficients of all vertices.

The shortest path is identified as the path through which the two concerned nodes are connected by the smallest number of intermediate nodes. The average shortest path length, also known as the *characteristic path length*, L , indicates the expected distance between two connected nodes.

The *network diameter*, D , is the maximum length of shortest paths between two nodes. If a network is disconnected, its diameter is the maximum of all diameters of its connected components. Therefore, D is a measure of compactness for a network. The Wiener index in chemical graph theory is half of the graph distance, and the other major distance descriptor is the average graph distance, or *graph radius*.

The *network density* is the normalized version of the average number of neighbors, which indicates the average connectivity of a node in the network. The value of network density is between 0 and 1, and it reflects how densely the network is populated with edges. A network that contains no edges, only solely isolated nodes, has a density of 0.

Based on the variance of the connectivity, the *network heterogeneity* is used to measure the connection tendency of a network that includes hub nodes. Biological networks tend to be very heterogeneous: while some hub nodes are highly connected, the majority of nodes tend to have very few connections.

In addition to these parameters related to connectedness and distance, betweenness and closeness are two important centrality measures. The *betweenness centrality* B_k of a node k is the number of times that a node is included in the shortest path between each pair of nodes, normalized by the total number of pairs. It is defined as

$$B_k = \sum_{s \neq n \neq t} (\sigma_{st}(k)/\sigma_{st}), \quad (1)$$

where s and t are nodes in the network other than k , σ_{st} denotes the number of shortest paths from s to t , and $\sigma_{st}(n)$ is the number of shortest paths from s to t on n lies. The betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes in the network.

The *closeness centrality* C_k of a node k is the reciprocal of the average shortest path length, which can be calculated as follows:

$$C_k = \frac{(x-1)}{\sum_{m \in U, m \neq k} L(m, k)}, \quad (2)$$

where U is the set of all nodes and x is the number of nodes in the network. Closeness centrality is a measure of how quickly information spreads from a given node to other reachable nodes in the network.

The *residue centrality* is the change of the characteristic path length L by removing node k and its corresponding edges (del Sol

et al., 2006)

$$\Delta L_k = |L - L_{rem,k}|, \quad (3)$$

where $L_{rem,k}$ denotes the characteristic path length after the removal of node K with its links from the network.

Key residues for communication and robustness are determined using statistical z-score values of the betweenness centrality and the residue centrality, respectively, and are calculated as follows:

$$z_k = (B_k - \bar{B})/\sigma, \quad (4)$$

$$z_k = (\Delta L_k - \bar{\Delta L})/\sigma, \quad (5)$$

where \bar{B} and $\bar{\Delta L}$ denote the betweenness centrality and the residue centrality averaged over all the residues, and σ is the standard deviation.

2.3. Computational details

The proteins studied in this work are Dronpa and the β clamp (Fig. 1), corresponding to PDB codes 2IE2 and 2POL, respectively. Calculation and visualization of protein contact maps were performed using CMView (Vehlow et al., 2011). The RING (Martin et al., 2011) program was used to generate their residue interaction networks. REDUCE was applied for adding H-atoms to the original X-ray structure, and PROBE was used to identify noncovalent residue interactions. The residue interaction network was visualized using Cytoscape (Shannon et al., 2003) and the plugin RINalyzer (Doncheva et al., 2011). The simple network parameters were calculated by NetworkAnalyzer plugin (Assenov et al., 2008) of Cytoscape, as well as closeness and betweenness were also calculated by RINalyzer. The Shortest Path Cytoscape plugin was subsequently used to identify the shortest path or distance between two selected nodes. The network analysis considered different edge types for all possible interactions, Van der Waals contacts, H-bonds, overlaps, and main-chain and side-chain interactions. All measures incorporated weighted edges where more than one interaction occurred between two nodes. The script for calculating residue centrality of the network was written and executed using Fortran 6.6. The calculation of Root mean square fluctuation (RMSF) was based on the statistical mechanics foundations of ENM as previously described (Hu et al., 2012).

3. Results

3.1. Protein contact maps

Protein contact maps represent residue interaction networks in two dimensions, which facilitate the identification of structural features such as interactions within and between secondary structural elements and domains (Barah and Sinha, 2008; Bhavani et al., 2011). In the protein contact and distance maps for Dronpa (Fig. 1b), the black dots in contact map (upper right) and red color in distance map (lower right) indicate interactions between the corresponding residues. The protein contact map clearly revealed the topology of the cylinder, while 10 red bars represent interactions between 11 neighboring β -strands. Similarly, the β clamp contact map clearly represented the contact topology of a monomer (Fig. 1d), with the three domains of the monomer falling along the diagonal. In addition, six circled clusters indicated interactions between α helices that form a circle.

3.2. General network topologies

Additional network parameters such as network diameter D , network radius R , average number of neighbors K , network density, and network heterogeneity are computed (Table 1). The results of chain A of the β clamp are listed to show the topological difference between the β clamp the monomer and dimer.

Network diameters for Dronpa, and for the monomer and dimer of β clamp were 10, 15, and 19, respectively, and the network radii were 5, 8, and 14, respectively. The radius of Dronpa network is equal to the half of the diameter, whereas the radius of the β clamp is greater than half of the diameter. This indicates that the structure of Dronpa is more regular, as expected for the cylindrical symmetry. The density of Dronpa (0.037) was also found to be larger, and this protein is therefore more compact. The density is smaller in the dimer of the β clamp, indicating that toroidal topology is not favorable for closed packing. Furthermore, the average number of neighbors and the network heterogeneity are similar for all three systems, suggesting appreciable network robustness.

The characteristic path length, L , and the mean coefficient $\langle C \rangle$ are used to quantify the structural characteristics of the small-world networks. For a random network with N nodes, $L_{\text{random}} \sim \ln N / \ln K$; $C_{\text{random}} \sim K/N$, whereas for regular network, $L_{\text{regular}} = N(N+K-2)/[2K(N-1)]$; $C_{\text{regular}} = 3(k-2)/[4(k-1)]$. Therefore, random networks are characterized by small values of $\langle C \rangle$ and L , while regular networks are characterized by high values of $\langle C \rangle$ and L . Small-world topology is, however, an alternative lying between random and regular network topologies, with small values of L and high values of $\langle C \rangle$.

Intuitively, we would expect two systems have high $\langle C \rangle$ but small L and exhibit small world behavior, and this was confirmed to be the case (Table 2). For Dronpa, $\langle C \rangle = 0.443$, $L = 4.292$, whereas for the corresponding random network and regular network, these values were 0.037 and 0.64, and 2.812 and 13.973, respectively. For the β clamp, $\langle C \rangle = 0.434$, $L = 8.681$, whereas for the corresponding random network and regular network, these values were 0.01 and 0.637, and 3.473 and 48.325, respectively. The relatively high $\langle C \rangle$ and the relatively small L suggest that networks for Dronap and the β clamp fall into the class of small-world networks. This is further supported by the residue degree distributions for Dronpa and the β clamp (Fig. 2), which were both found to be bell-shaped and Poisson-like, as is typical of small-world networks. The degrees of residues are most distributed between 15 and 20, and the number of nodes with higher degree will fall off immediately.

Table 1
Network parameters for three network systems.

Networks	Size (N/E)	Diameter	Radius	Density	Heterogeneity	K
Dronpa	213/1577	10	5	0.037	0.342	7.869
β clamp	732/5249	19	14	0.01	0.356	7.653
β clamp_A	366/2601	15	8	0.021	0.359	7.519

Table 2
Characteristic path length L and clustering coefficient $\langle C \rangle$ for three protein networks, compared to their random and regular graphs.

Networks	$\langle C \rangle$	C_{random}	C_{regular}	L	L_{random}	L_{regular}
Dronpa	0.443	0.037	0.64	4.292	2.812	13.973
β clamp	0.434	0.01	0.637	8.681	3.473	48.325
β clamp_A	0.436	0.018	0.635	6.475	3.142	24.773

3.3. Network centrality

Network centrality measures span the entire network topology from local to global. The centralization of any network reflects the centrality of the most central node compared with all other nodes. It is of important to analyze the centrality of cylindrically symmetric proteins, and betweenness (B_k) and closeness (C_k) are particularly appropriate for this. B_k and C_k for each node were calculated and their distributions were investigated (Figs. 3 and 4, respectively). Nodes with bigger circles have larger B_k values, as well as nodes with brighter (red means brightest, and green means darkest) colors have larger C_k . The centrality analysis indicated that residues with larger betweenness and closeness were distributed around the central portion of the Dronpa cylinder, and within the circumference of the torus of the β clamp. In order to quantify the distributions, B_k and C_k were calculated as a function of residue number for Dronpa and the β clamp (Figs. 5 and 6). B_k and C_k were found to be positively correlated, with correlation coefficient of 0.7146 for Dronpa and 0.6271 for the β clamp.

Peaks in the profile of B_k correspond to the most connected residues, which are considered to play a key role in protein folding. For Dronpa, Fig. 5 predicted that residues Asn65, Val76, Ser88, and Cys113 have high B_k . For the β clamp, Fig. 6 predicted that residues Lys12, Gln143, Phe144, Met204, Ala268, Arg269, Leu325 and Asp326 have high B_k . However, the distributions of C_k for both Dronpa and the β clamp are more even, which indicates that most of the residues in these two proteins are conservative. In order to describe the global topologies, we also computed the average betweenness and average closeness, which were 0.0216 and 0.0674 for Dronpa, and 0.0097 and 0.0323 for the β clamp.

3.4. Prediction of important residues by residue centrality

Error tolerance and attack vulnerability are generic properties of network robustness (Albert et al., 2000). To address the error tolerance of two residue interaction networks, we investigated the change in characteristic path length when the node and its corresponding edges are removed (residue centrality). The larger of residue centrality of a node, the greater the vulnerability to attacks, and removal may alter the network topology. Residue centrality has also been used to identify centrally conserved residues important for protein folding (del Sol et al., 2006; Emerson and Gothandam, 2012).

In this study, the residue centrality for both proteins was analyzed, and residues with a z-score of 2.0 or larger are predicted to be key residues for robustness. Network analysis identified seven key residues in Dronpa (Phe61, Arg66, Phe68, Trp89, Arg91, Met93, Tyr116), which are distributed around the chromophore. These residues may play important role in maintaining communication between the chromophore and other residues, as well as being key residues for protein fold or maintaining the fold. Ten key residues were detected in the β clamp analysis (Leu14, Gln16, Cys79, Phe144, Gln 143 in chain A; Gln16, Gln143, Phe144, Met204, Met206 in chain B; Fig. 7). All of these residues are distributed around the central cavity of the β clamp, which is the region most important for allosteric communication. Specifically, Leu14 and Gln16 belong to α_1 helix, whereas Cys79, Met204, and Met206 belong to α_2 helix. It is perhaps somewhat surprising that most important residues for robustness are distributed at the most interconnected parts within a single chain, but not at the dimerization interface. However, additional key residues at the interface are involved in allosteric communication in the β clamp.

3.5. The Dronpa hydrophobic pocket

Dronpa is a novel type of green fluorescent protein, and the photoswitching properties are determined by the chromophore.

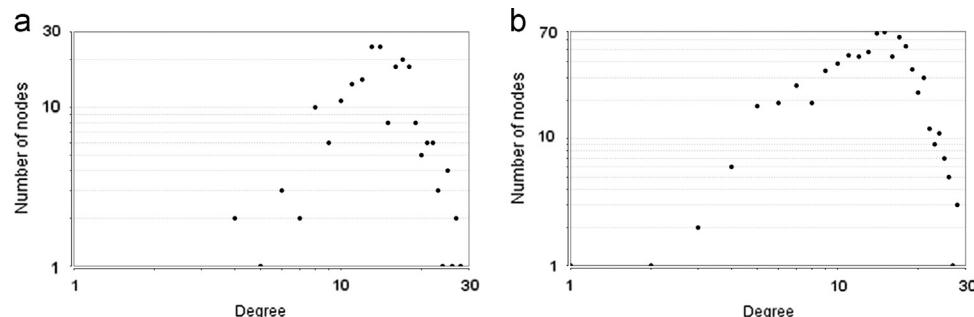


Fig. 2. Distribution of degree for (a) Dronpa and (b) the β clamp.

The hydrophobic pocket is comprised of Gln38, Met40, Thr58, Ile195, Leu209 and Glu211 (Fig. 8a), and the cysteine moiety of the chromophore lies in this pocket. It has been reported that this pocket markedly affects the spectral properties of the chromophore, and a detail network analysis could shed light on this phenomenon.

The sub-network of the hydrophobic pocket (Fig. 8b) gave a mean clustering coefficient $<C>$ and characteristic path length, L , of 0.01 and 3.473, respectively, and these values are also consistent with small-world property. In the sub-network, 6 nodes exhibited a degree between 4 and 7(Gln38, Leu209 and Glu211=7, Thr58 and Ile195=4, Met 40=5). All nodes had degrees 4 or larger, which may explain the stability of hydrophobic pocket. Betweenness for each residue was also calculated, and Glu211 had the largest value, suggesting this is a particularly important residue for the hydrophobic pocket. A total of 17 edges were present in the network, of which 14 corresponded to Van der Waals contacts, and 3 to hydrogen bonds. The hydrophobic pocket is therefore formed primarily from noncovalent interactions, especially Van der Waals contacts. It also indicates that a complex hydrogen bonding network exist in the vicinity of the chromophore.

For Dronpa, investigating the difference in network properties between β -strands and α -helices proved informative. As expected, β -strands require more interactions than α -helices since they need to connect each other along their entire lengths to form the barrel architecture. Comparison of degrees and betweenness for these two secondary structures found that they have similar degrees, but the α_1 helices have higher betweenness and closeness (Fig. 5). This suggests that α_1 helices have more connections than β -strands, and this may facilitate the photoswitching activity of the chromophore.

It is well known that medium- and long-range interactions play a key role in formation and stability of hydrophobic clusters in β -barrel proteins (Paci et al., 2012). In this study, network topology analysis of the hydrophobic pocket revealed structural principles that may impact on the conformational changes of the chromophore, providing new insights into the protons transfer network. To conclude, the residue interaction network established that (1) the hydrophobic pocket is formed from noncovalent interactions, with Van der Waals contacts particular important; (2) a complex hydrogen bonding network is present in the neighborhood of the chromophore; and (3) the α_1 helices have more connections than β -strands, which may facilitate the photoswitching activity of chromophore.

3.6. Communication across the dimer interface in the β clamp

Communication within and across proteins is crucial for the biological function. Experimental approaches such as mutational studies provide important information on the amino acids that are crucial for their function. The prediction of hot spots at protein interfaces using residue interaction networks is potentially simple way of identifying important residues. In order to study the communication across the dimer interface in the β clamp, interface hot

spots on both chains were first identified. Network analysis showed that the residues with a large betweenness contribute the most to the binding energy at the interfaces (del Sol and O'Meara, 2004). The Computational Hot Spots of Protein Interfaces (HotSprint) database (Guney et al., 2008) can also be used to distinguish functionally and structurally important residues at interface. A total of 19 residues for each protein chain were identified as interface hot spots in the β clamp using HotSprint (Table 3). In this study, betweenness centrality was used to determine key residues, and those with a z-score ≥ 2.0 are most important. Lys74, Gly81, Leu82, Arg269 and Gln299 in chain A, as well as Gly81, Pro83, Arg269, Ile272 and Glu300 in chain B were identified as important in this way. Residues identified by both HotSprint database and the network analysis are likely to play a major role in the stability of the interface, and mutation of these residues would likely severely destabilize the interface.

The shortest non-covalently connected path between two interface residues can be identified from the residue interaction network using the Floyd-Warshall algorithm. Accordingly, we identified a possible communication path using this approach (A: Pro71 (α_1) \rightarrow A: Lys74 \rightarrow B: Ile272 \rightarrow B: Arg269 (α_2); Fig. 9). This path is likely real for two reasons: (1) the starting residue Pro71 is located at helix α_1 in chain A, and the end residue Arg269 locates at helix α_2 in chain B, both in the required places. Furthermore, this agrees with previous harmonic analysis results that showed a strong correlation between α_1 and α_2 spatial adjacency (Hu et al., 2012). Two residues (Lys74 and Ile272) along this path exhibited among the largest B_k values of residues in chain A and B, and correspond to interface hot spots that contribute significantly to communication across the dimer interface. Table 4 shows more the shortest possible communication paths across the dimer interface, but these require experiment verification in future studies.

It should be noted that the several other methods could give more rigorous information about communication, such as the integration of network analysis and MD simulation (Ghosh and Vishveshwara, 2007; Vanwart et al., 2012), as well as ENM combined with structural perturbation methods (Antal et al., 2009; Erman, 2013; Raimondi et al., 2013).

4. Discussion

4.1. Comparison with globular proteins

Both symmetry and topology play important roles in the structure of proteins, which may take on globular, cylindrical and toroidal shapes. Bagler and Sinha (2005) and Emerson and Gothandam (2012a, 2012b) investigated the network properties for both globular and transmembrane proteins. It is worth noting that transmembrane proteins also adapt cylinder-like topology, and comparison with our results may give a deeper understanding of protein structure principles.

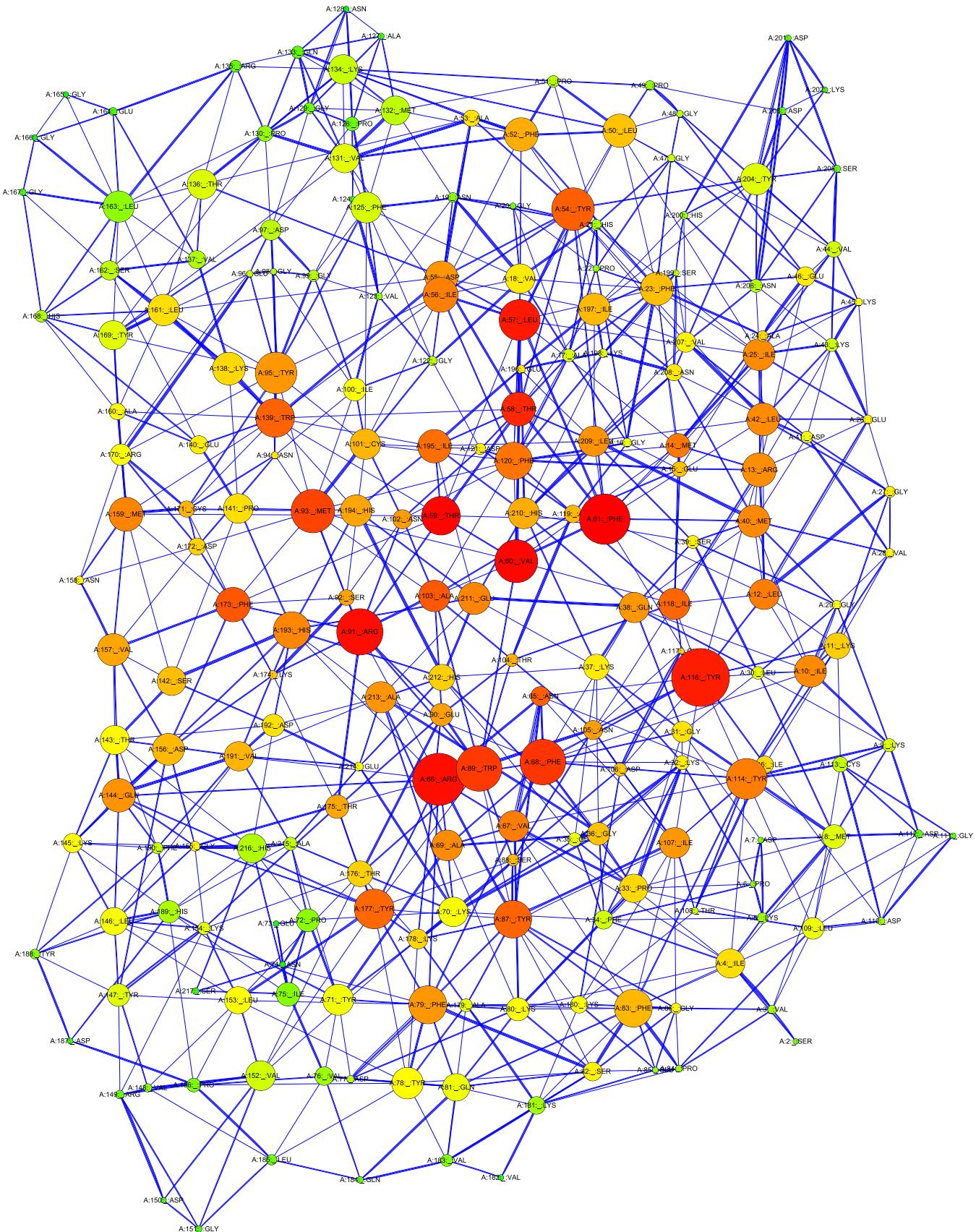


Fig. 3. The residue interaction network for Dronpa. Betweenness is denoted by node size (low values to small sizes), as well as closeness is denoted by node color (low values to bright colors).

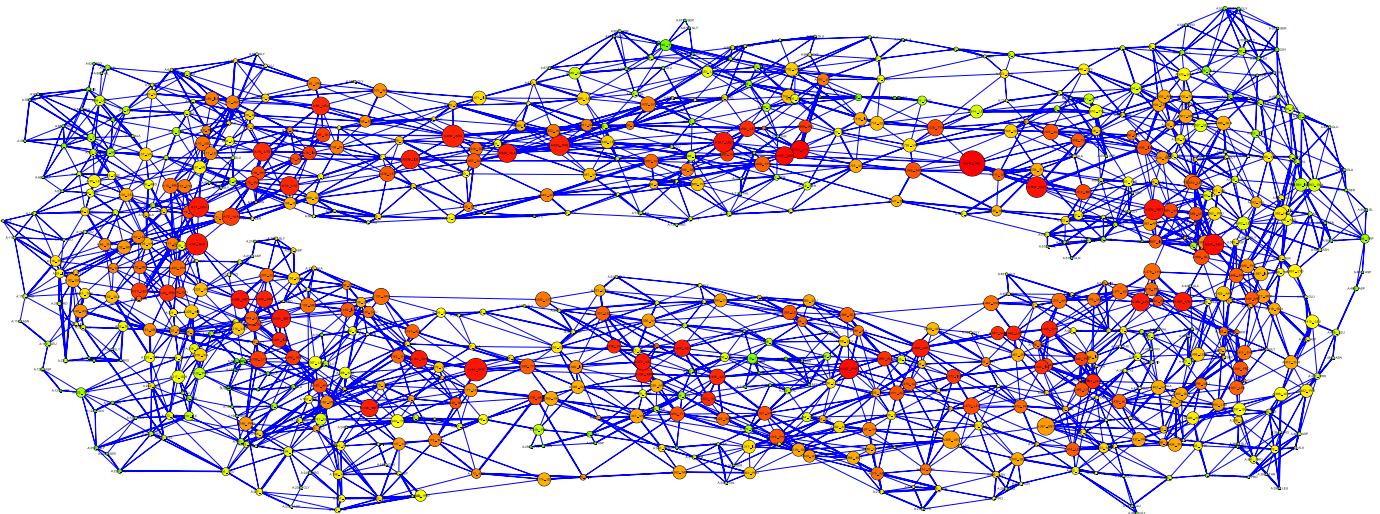


Fig. 4. The residue interaction network for the β clamp. Betweenness is denoted by node size (low values to small sizes), as well as closeness is denoted by node color (low values to bright colors).

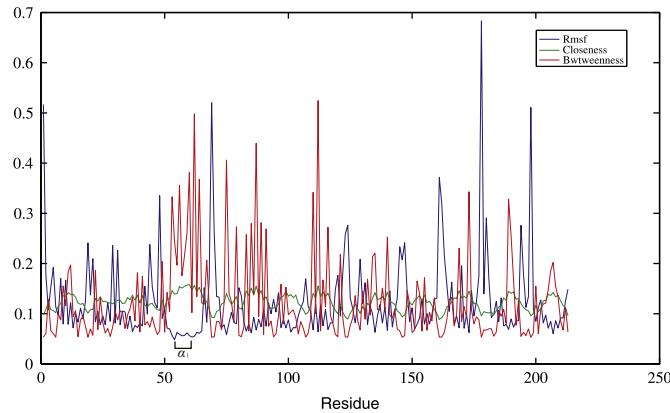


Fig. 5. Betweenness (the red line), closeness (the green line) and root mean square fluctuation (RMSF, the blue line) of Dronpa against residue numbers. The betweenness and closeness are rescaled to match the average of RMSF. It shows a positive correlation between betweenness and closeness, while they have anti-correlations with RMSF. Helix α_1 is indicated in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

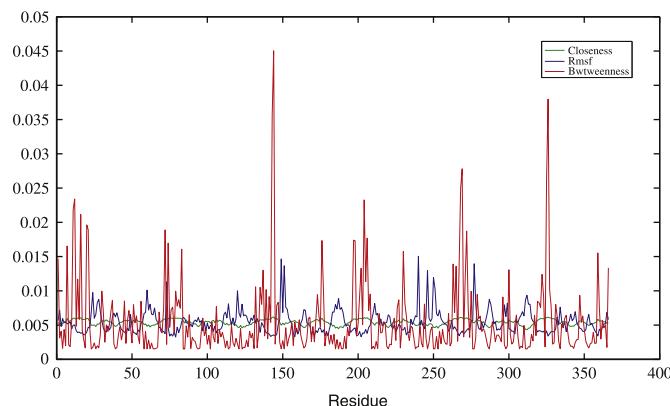


Fig. 6. Betweenness (the red line), closeness (the green line) and RMSF (the blue line) of the β clamp against residue numbers. The betweenness and closeness are rescaled to match the average of RMSF. It shows a positive correlation between betweenness and closeness, while they have anti-correlations with RMSF. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

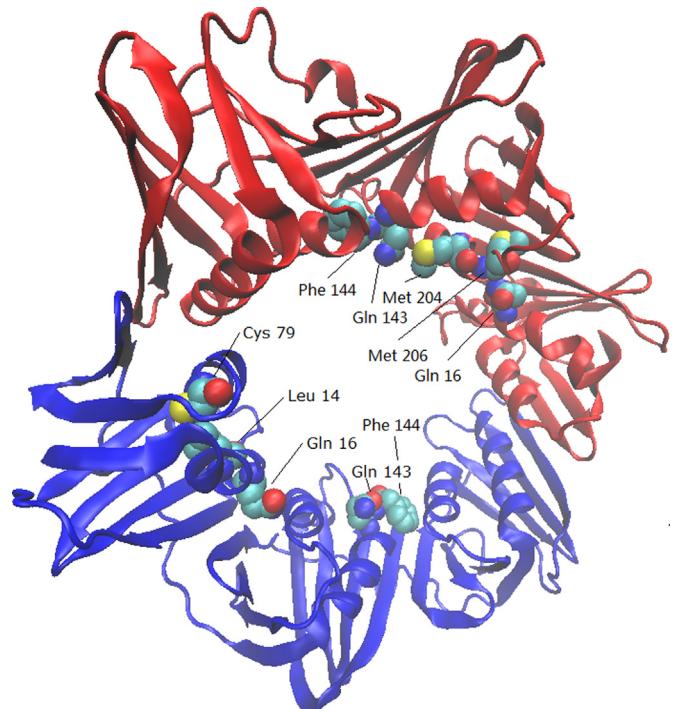


Fig. 7. Representation of ten key residues for robustness in the structure of the β clamp.

The network diameter value D for globular proteins is approximately 8, and the average $\langle C \rangle$ value is around 0.5418. This indicates that there is a significant difference in network diameter between globular, cylindrical and toroidal proteins. The larger D observed for the Dronpa is due to the elongated structure of the cylinder, whereas the larger D for the β clamp is due to the large cavity within the torus. Proteins with smaller value of D are spatially distributed, and globular proteins therefore have larger packing densities. This is further supported by the average closeness value, which was 0.0674 for Dronpa, 0.0323 for the β clamp, and 0.3226 for globular proteins. Larger average closeness values have larger packing densities. Although the $\langle C \rangle$ values for the three types of proteins is similar, we can use the

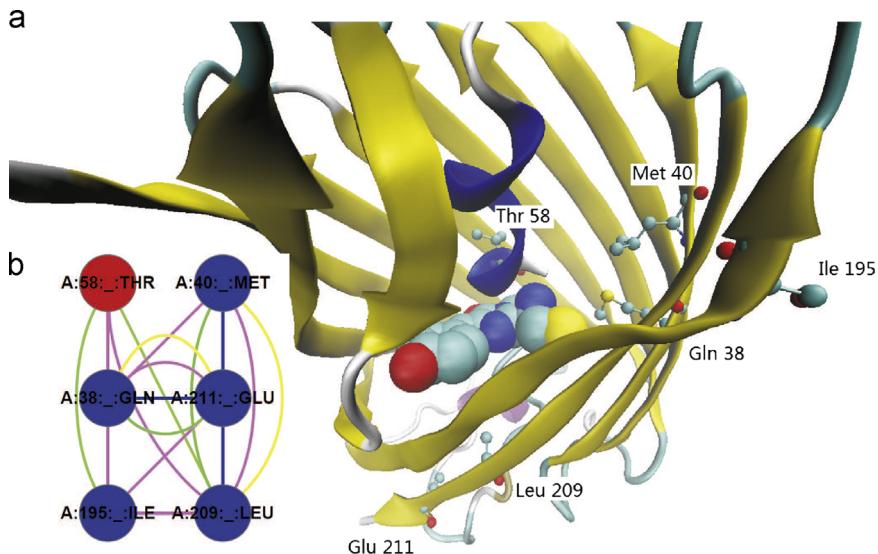


Fig. 8. The spatial structure and local contact network for the hydrophobic pocket in Dronpa. (a) The new cartoon structure of the hydrophobic pocket. Residues in the pocket are represented by CPK models, and the chromophore is represented by VDW model. (b) The sub-network for the hydrophobic pocket, where blue lines indicate hydrogen bonds.

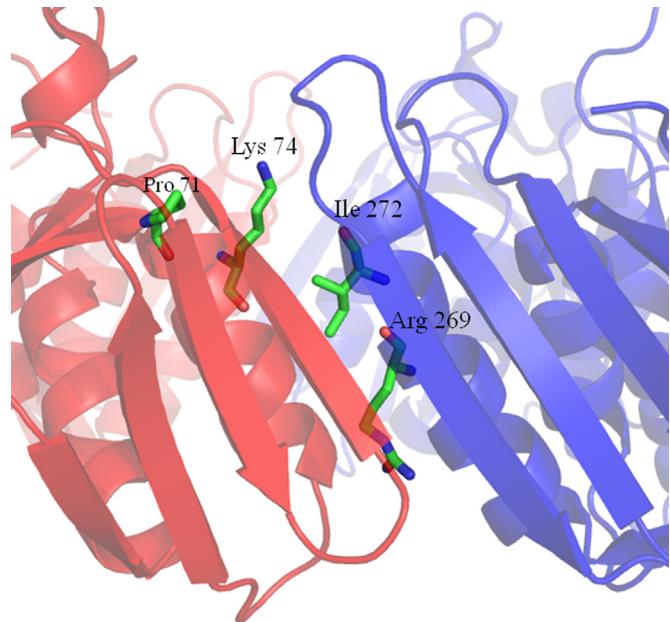


Fig. 9. A communication pathway at the interface of the β clamp. Two chains are represented by different colors.

ratio $\langle C \rangle / C_{\text{random}}$ to compare their small-world properties (Tao et al., 2009). Globular protein exhibited a ratio of 15.29, while the values for Dronpa and the β clamp are 11.97 and 43.4, respectively. According to this ratio, the toroidal topology is the most correlated with small-world properties. However, this ratio also depends on the network size (Kundu, 2005), and larger networks have bigger value of $\langle C \rangle / C_{\text{random}}$. Therefore, a deeper inspection of the relationship between small-world properties and protein topology is needed.

4.2. Comparison with the dynamical analysis

In previous work, the dynamical analysis of Dronpa and the β clamp was investigated to compare their dynamics information encoded in different topological structures (Hu et al., 2012). The elastic network model was used to detect collective motions and

RMSF of Dronpa and the β clamp. Compared with ENM, residue interaction network theory is an efficient computational tool that can quantify the structural organization of proteins using various network parameters. Since ENM is also a method based on the weighted network model, comparison between ENM and residue interaction network analysis is logical and of considerable importance. The correlation results for Dronpa showed that α -helices and β -strands are linked by hydrophobic interaction. The effect of these correlations is more pronounced in the network analysis of the hydrophobic pocket. The ENM results also showed inter-domain correlations between α_1 and α_2 , and between β_4 and β_8 of the β clamp. The residue interaction network analysis gives detailed information on which residues along this interface are important, as well as which may have higher correlations between them.

In ENM, the flexibility and rigidity of proteins is described in terms of the RMSF. Atilgan et al. (2004) previously investigated the relationship between ENM and residue interaction network analysis. We found an anti-correlation between B_k and RMSF (Figs. 5 and 6). B_k and C_k must be rescaled to match the average of RMSF so that these profiles are comparable. For dronpa, the rescaled factor of B_k was 5.68 and the correlation coefficient of $1/B_k$ and RMSF was 0.5634. For the β clamp, the rescaled factor of B_k was 0.5567 and the correlation coefficient of $1/B_k$ and RMSF was 0.5864. This is in good agreement with other results that showed that large B_k corresponded to highly connected nodes, and ENM shows that these hub residues have lower flexibility than sparsely connected residues.

4.3. Robustness of proteins

Protein structures must be robust to withstand mutational events and maintain function. Robustness of network systems can be analyzed by degree of tolerance against errors. Therefore, modeling proteins as networks opens up an alternative way to analyze robustness. The robustness of the residue interaction network is rooted in the homogeneity of the connectivity distribution. In this study, we addressed system robustness by identification of residues responsible for maintaining characteristic path length. The removal of these residues may enhance the vulnerability to attack, making the protein systems more easily damaged. It would be interesting to study the relationship between network attack and residue mutations, which may lead to predictions of disease-associated single amino acid polymorphisms (SAPs).

Table 3

Interface hot spots of the β clamp predicted by HotSprint. Residues highlighted by red have betweenness z-score values larger than or equal to 2.0.

Chain A	Pro71	Lys74	Asp77	Ile78	Gly81	Leu82	Pro83	Ser104	Phe106
Leu108	Arg269	Ile272	Leu273	Asn296	Glu298	Gln299	Glu300	Ala302	Glu304
Chain B	Pro71	Lys74	Asp77	Ile78	Gly81	Leu82	Pro83	Ser104	Phe106
Leu108	Arg269	Ile272	Leu273	Asn296	Glu298	Gln299	Glu300	Ala302	Glu304

Table 4

Six predicted communication pathways across the dimer interface in the β clamp.

- A: Lys74 → B: Ile272 → B: Arg269
- A: Gly81 → B: Arg269 → B: Ile272
- A: Gly81 → A: Ile78 → B: 273 → B: Glu300
- A: Leu82 → B: Arg269 → B: Ile272
- A: Leu82 → A: Phe106 → B: Leu273 → B: Glu300
- B: Gly81 → B: Asp77 → B: Lys74 → A: Glu300 → A: Gln299

Another structural characteristic of a protein is its backbone connectivity. In Dronpa, 11 anti-parallel β -strands form a classical β -barrel with cylindrical topology, which is stabilized by a network of backbone hydrogen bonds. This rigid β -barrel structure makes Dronpa a very stable protein. In comparison with Dronpa, the sliding β clamp forms a β -wheel with toroidal topology through polymerization of six very similar domains. Although such an open β -wheel structure does not form a single interconnected structure, it is stabilized by four strong hydrogen bonds and several other linkages that tether the β -strands across the dimer interface. This is another reminder that high robustness is reflected in proteins symmetry, form, and shape.

5. Conclusions

Structural analysis of Dronpa and the β clamp using topological indexes based on network theory, and the effects of protein connectivity on the cylindrical and toroidal topologies were investigated using various network metrics. The characteristic path lengths and clustering coefficients of both networks showed that both structures posses small-world properties. The larger network diameters and average closeness indicate that Dronpa and the β clamp are less densely packed then typical globular proteins, presumably due to their cylindrical and toroidal topologies. Residue centrality values identified key residues involved in protein folding and communication, and robustness in relation to these residues was discussed. Network theory also proved a simple and efficient way to analyze the topological structure of the hydrophobic pocket in Dronpa. This approach also proved useful for detecting communication pathways across the interface in the β clamp by combining betweenness and HotSprint predictions. A comparative analysis of network properties and ENM results may reveal information on protein dynamics and the relation of this to topology. Protein function can be investigated by exploring the relationships between topology and structural fluctuations. Dronpa, a green fluorescent protein, and the β clamp, a nuclear antigen found in proliferating eukaryotic and archaeal cells, are representative members of two large distant protein families. The topological properties revealed through network analysis in this work may be extended to other members of these large protein families.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (21203131 and 91230117), the Natural Science Foundation of the Jiangsu Higher Education Institutions (12KJB180014).

References

- Aftabuddin, M., Kundu, S., 2006. Weighted and unweighted network of amino acids within protein. *Physica A* 396, 895–904.
- Albert, R., Jeong, Hawoong., Barabási, A.L., 2000. Error and attack tolerance of complex networks. *Nature* 406, 378–382.
- Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanely, D., Venger, I., Pietrovoski, S., 2004. Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* 344, 1135–1146.
- Antal, M.A., Csaba, B., Csermely, P., 2009. Perturbation Waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr. Protein Pept. Sci.* 10, 161–172.
- Assenov, Y., Ramírez, F., Schelhorn, S.E., Lengauer, T., Albrecht, M., 2008. Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284.
- Atilgan, A.R., Akan, P., Baysal, C., 2004. Small-world communication of residues and significance for protein dynamics. *Biophys. J.* 86, 85–91.
- Bagler, G., Sinha, S., 2005. Network properties of protein structures. *Phys. A: Stat. Mech. Appl.* 346, 27–33.
- Bagler, G., Sinha, S., 2007. Assortative mixing in protein contact networks and protein folding kinetics. *Bioinformatics* 23, 1760–1767.
- Bahar, I., Attilgan, A.R., Erman, B., 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.* 2, 173–181.
- Barah, P., Sinha, S., 2008. Analysis of protein folds using protein contact networks. *Pramana – J. Phys.* 71, 369–378.
- Bartoli, L., Fariselli, P., Casadio, R., 2007. The effect of backbone on the small-world properties of protein contact maps. *Phys. Biol.* 4, L1–L5.
- Bhavani, S.D., Suvarnavani, K., Sinha, S., 2011. Mining of protein contact maps for protein fold prediction. *WIREs Data Min. Knowl. Discov.* 1, 362–368.
- Bode, C., Kovacs, I.A., Szalay, M.S., Palotai, R., Korcsmaros, T., Csermely, P., 2007. Network analysis of protein dynamics. *FEBS Lett.* 581, 2776–2782.
- Bonchev, D., Buck, G.A., 2007. From molecular to biological structure and back. *J. Chem. Inf. Model.* 47, 909–917.
- Brinda, K.V., Vishveshwara, S., 2005. Oligomeric protein structure networks: insights into protein–protein interactions. *BMC Bioinformatics* 6, 296.
- Brinda, K.V., Vishveshwara, S., 2005. A network representation of protein structures: implications for protein stability. *Biophys. J.* 89, 4159–4170.
- Doncheva, N.T., Klein, K., Domingues, F.S., Albrecht, M., 2011. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 36, 179–182.
- Doncheva, N.T., Assenov, Y., Domingues, F.S., Albrecht, M., 2012. Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.* 7, 670–685.
- Emerson, I.A., Gothandam, K.M., 2012a. Residue centrality in alpha helical polytopic transmembrane protein structures. *J. Theor. Biol.* 309, 78–87.
- Emerson, I.A., Gothandam, K.M., 2012b. Network analysis of transmembrane protein structures. *Phys. A: Stat. Mech. Appl.* 391, 905–916.
- Erman, B.A., 2013. Fast approximate method of identifying paths of allosteric communication in proteins. *Proteins* 81, 1097–1101.
- Estrada, E., 2010. Universality in protein residue networks. *Biophys. J.* 98, 890–900.
- Fang, Y.P., Ma, D.C., Li, M.L., Wen, Z.N., Diao, Y.B., 2010. Investigation of proteins folding rates and their properties of amino acid networks. *Chemometrics Intell. Lab. Syst.* 101, 123–129.
- Ghosh, A., Vishveshwara, S., 2007. A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15711–15716.
- Greene, L.H., 2012. Protein structure networks. *Brief. Funct. Genomics* 11, 469–478.
- Greene, L.H., Higman, V.A., 2003. Uncovering network systems within protein structures. *J. Mol. Biol.* 334, 781–791.
- Guney, E., Tuncbag, N., Keskin, O., Gursoy, A., 2008. HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res.* 36, D662–666.
- Hu, G., Michielssens, S., Moors, S.L.C., Ceulemans, A., 2012. The harmonic analysis of cylindrically symmetric proteins: a comparison of Dronpa and a DNA sliding clamp. *J. Mol. Gr. Model.* 34, 28–37.
- Hu, G., Zhou, J., Yan, W., Chen, J., Shen, B., 2013. The Topology and dynamics of protein complexes: insights from intra-molecular network theory. *Curr. Protein Pept. Sci.* 14, 121–132.
- Jiao, X., Chang, S., 2011. Scoring function based on weighted residue network. *Int. J. Mol. Sci.* 12, 8773–8786.
- Kannan, N., Vishveshwara, S., 1999. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* 292, 441–464.
- Kong, X.P., Onrust, R., O'Donnell, M., Kurian, J., 1992. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69, 425–437.
- Krishnan, A., Zbilut, J.P., Tomita, M., Giuliani, A., 2008. Proteins as networks: Usefulness of graph theory in protein science. *Curr. Protein Pept. Sci.* 9, 28–38.

- Kundu, S., 2005. Amino acids network within protein. *Physica A* 346, 104–109.
- Martin, A.J., Vidotto, M., Boscariol, F., Di Domenico, T., Walsh, I., Tosatto, S.C.E., 2011. RING: networking interacting residues, evolutionary information and energetics in protein structures. *Bioinformatics* 27, 2003–2005.
- Paci, P., Di Paola, L., Santoni, D., De Ruvo, M., Giuliani, A., 2012. Structural and functional analysis of hemoglobin and serum albumin through protein long-range interaction networks. *Curr. Proteomics* 9, 160–166.
- Paola, L.D., Ruvo, M.D., Paci, P., Santoni, D., Giuliani, A., 2013. Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.* 113, 1598–1613.
- Raimondi, F., Felline, A., Seeber, M., Mariani, S., Fanelli, F., 2013. A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems. *J. Chem. Theory Computation* 9, 2504–2518.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- del Sol, A., O'Meara, P., 2004. Small-world network approach to identify key residues in protein–protein interaction. *Proteins* 58, 672–682.
- del Sol, A., Fujihashi, H., Amoros, D., Nussinov, R., 2006. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* 2 (2006), 0019.
- Sun, W., He, J., 2011. From isotropic to anisotropic side chain representations: comparison of three models for residue contact estimation. *PLoS One* 6, e19238.
- Tao, Y., Zou, X.Q., Huang, S.Y., Zou, X.W., 2009. Cutoff variation induces different topological properties: A new discovery of amino acid network within protein. *J. Theor. Biol.* 256, 408–413.
- Taylor, N.R., 2013. Small world network strategies for studying protein structures and binding. *Computational Struct. Biotechnol. J.* 5, e201302006.
- Tirion, M.M., 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908.
- Vanwart, A.T., Eargle, J., Luthey-Schulten, Z., Amaro, R.E., 2012. Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Computation* 8, 2949–2961.
- Vehlow, C., Stehr, H., Winkelmann, M., Duarte, J.M., Petzold, L., Dinse, J., Lappe, M., 2011. CMView: interactive contact map visualization and analysis. *Bioinformatics* 27, 1573–1577.
- Vendruscolo, M., Dokholyan, N.V., Paci, E., Karplus, M., 2002. Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E* 65, 061910:1–061910:4.
- Vishveshwara, S., Ghosh, A., Hansia, P., 2009. Intra and inter-molecular communications through protein structure network. *Curr. Protein Pept. Sci.* 10, 146–160.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. *Nature* 393, 440–442.
- Wilmann, P.G., Turcic, K., Battad, J.M., Wilce, M.C.J., Devenish, R.J., Prescott, M., Rossjohn, J., 2006. The 1.7 Å crystal structure of Dronpa: a photoswitchable green fluorescent protein. *J. Mol. Biol.* 364, 213–224.
- Zhou, J., Yan, W., Hu, G., Shen, B., SVR_CAF: an integrated score function for detecting native protein structures among decoys. *Proteins*, <http://dx.doi.org/10.1002/prot.24421>, (in press).
- Zhou, J., Yan, W., Hu, G., Shen, B., 2014. Amino acid network for the discrimination of native protein structures from decoys. *Curr. Protein Pept. Sci.* (accepted for publication).