Received: 30 May 2023 | Revised: 31 August 2023 | Accepted: 14 September 2023

DOI: 10.1002/prot.26606

RESEARCH ARTICLE

PROTEINS WILEY

The alteration of structural network upon transient association between proteins studied using graph theory

Vasam Manjveekar Prabantu[1] | Himani Tandon[1,2] | Sankaran Sandhya[3] | Ramanathan Sowdhamini[1,4,5] | Narayanaswamy Srinivasan[1]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

[2]Structural Studies Division, MRC Laboratory of Molecular Biology, Cambridge, UK

[3]Faculty of Life and Health Sciences, Department of Biotechnology, Ramaiah University of Applied Sciences, Bangalore, India

[4]National Centre for Biological Sciences (TIFR), Bangalore, India

[5]Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

Correspondence
Ramanathan Sowdhamini, Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India.
Email: mini@ncbs.res.in

Funding information
Department of Biotechnology, Ministry of Science and Technology, India, Grant/Award Number: BT/PR40187/BTIS/137/9/2021; Institute of Bioinformatics and Applied Biotechnology, Grant/Award Number: IBAB/MSCB/182/2022; Science and Engineering Research Board, Grant/Award Number: JBR/2021/000006

Abstract

Proteins such as enzymes perform their function by predominant non-covalent bond interactions between transiently interacting units. There is an impact on the overall structural topology of the protein, albeit transient nature of such interactions, that enable proteins to deactivate or activate. This aspect of the alteration of the structural topology is studied by employing protein structural networks, which are node-edge representative models of protein structure, reported as a robust tool for capturing interactions between residues. Several methods have been optimized to collect meaningful, functionally relevant information by studying alteration of structural networks. In this article, different methods of comparing protein structural networks are employed, along with spectral decomposition of graphs to study the subtle impact of protein–protein interactions. A detailed analysis of the structural network of interacting partners is performed across a dataset of around 900 pairs of bound complexes and corresponding unbound protein structures. The variation in network parameters at, around, and far away from the interface are analyzed. Finally, we present interesting case studies, where an allosteric mechanism of structural impact is understood from communication-path detection methods. The results of this analysis are beneficial in understanding protein stability, for future engineering, and docking studies.

KEYWORDS
allostery, graph spectral analysis, network alterations, protein structural networks, protein–protein interaction, transient association

1 | INTRODUCTION

The importance of molecular flexibility for a protein to function is well documented.[1–3] The analysis of different structural states helps in analyzing its function.[4] Several proteins are very rigid and exist in fewer distinct conformations. However, most proteins exist predominantly in an inactive state and seldom in an active state. The difference between the structural states arises from subtle changes in local conformations at specific sites or from large structural excursions with altered topology. Some proteins also populate an intermediate state, which may be stabilized in a dynamic equilibrium due to strong interactions.[5,6] The stabilization of such a transient state is generally brought about by external perturbations like post translational modifications, mutations, or various binding partners.[7–9] One such scenario is the perturbation of the protein structure due to binding with another protein. The changes brought about at the site of their interface is generally allosterically transmitted across the structure of the protein to impact or present its active site.[10–12] Although this is intuitive to understand, the mechanism of this signaling has not been thoroughly analyzed across all proteins.

The analysis of allosteric signaling can often be approached using protein structural networks (PSN).[13–15] It is a simplistic

Ramanathan Sowdhamini and Narayanaswamy Srinivasan are joint corresponding authors.

mathematical model of the residue interactions that exist in a well-ordered globular protein structure, where the residues are considered as nodes and any interaction between them is considered by drawing an edge between them. The edge signifies a relationship between the interacting residues, generally pertaining to features such as non-covalent bonding, proximity between atoms in space and sometimes also based on energy parameters that can be defined for interactions.[16–19]

A further advancement on the use of these structure networks is the application of a graph-spectral method for the global comparison of proteins.[20,21] This method is robust and sensitive to even minute structural perturbations that can be observed between any pair of networks that are given as input.[22,23] The method described by Vasundhara and co-workers is an advanced comparison metric that has been applied in the validation of protein structure models and in understanding the amount of structural variability in an ensemble of structures of a given protein.[24–26] The method has been employed in the current work for the global comparison of PSNs.

We have compared the bound and the unbound forms of the same protein that has been crystallized when bound to interacting partners as well as independently. The ProPairs database is a large consortium of available bound and unbound proteins structures and provides this information. Any changes in network parameters, both local and global, are discussed along with interesting clinically relevant scenarios where the network may have changed considerably without affecting its overall structure topology. The rearrangement, gain and loss of network connectivity are discussed further, illustrated using case studies.

2 | MATERIALS AND METHODS

2.1 | Dataset

The Propairs database[27] of legitimate protein–protein docking complexes enlists crystal structures of bound complexes, defined as biological complexes in the Protein Data Bank (PDB),[28] and their corresponding unbound protein structure(s). A total of 2378 bound complexes that have corresponding unbound structures of the interacting partners were collected. It was ensured that the crystal structures of the bound complex are of better than 3 Å resolution. Structures in the bound form were required to have the same oligomeric state as in the unbound form to ensure that the perturbation is not influenced by the docking of any other protein molecule. The dataset was further curated to remove structures with missing residues. Any multiple occupancies were corrected to obtain a conformer with single occupancy, based on highest occupancy value, for all its atoms. A total of 895 chains from a bound complex with interfacial sites were paired with the corresponding chain in the unbound form to make the working dataset for this analysis. Topologically equivalent residues are identified using the TM Align tool. The list of all the protein chains along with the analysis of their network information can be found in Table S1.

2.2 | Interface and non-interface definition based on atom contacts

Non-covalent bonding between associating protein units form the interface in-between partnering molecules that form a complex. In this analysis, any pair of residues from the interacting partners, whose atoms fall within a distance of 4.5 Å are considered to form the interfacial sites. This concept is based on proximity and is called an atom contact, is also similar to the method used in the construction of edges to make a structural network of the protein. The distance-based method effectively captures the interfacial sites in the bound form of the complex (Table S5). Any non-adjacent residue within the structural network that comes directly in contact with the interfacial sites by making atom contacts are termed as primary contacts. These sites make up the rim of the interface around the interfacial core. All the remaining residues that are away from the interface makeup the non-interfacial sites.

2.3 | Network construction and analysis

The PSN is a weighted graph representation of residues interconnected to depict a network of interactions. Each residue in the PSN is a node and they are connected to other nodes using edges based on their 3D coordinates. Spatial distance, chemical non-covalent bonds, charge, energy, or many other features are several means of defining network edges. In the current analysis, inter-atomic distance between atoms of non-adjacent residues are used to define interactions. A distance proximity cut-off of 4.5 Å is used to define atom contacts.[25] The weight of an edge is defined as the ratio of the total number of atom contacts made between a pair of residues and the maximum number of atom contacts found in the entire dataset between the corresponding amino acids.[29]

The degree parameter and node strength obtained from the sum of all edge weights connected to a node describes the connectivity around the corresponding residues in the PSN. A higher degree would suggest that the well-connected node is a hub in the network. Upon binding, any significant change in the node strength or degree of sites, especially hubs, corresponding to loss or gain of crucial connectivity in the PSN are analyzed.

2.4 | Global network and structure comparison

Structures in their bound form are aligned with the corresponding structures in their unbound form and information of structural deviation (RMSD) between all pairs of chains are obtained. TM-align is used to compute the RMSD, which is a measure of Cα deviation between topologically equivalent residues. This is followed by global network comparison of the same equivalent residues using the network dissimilarity method discussed by Gadiyaram and coworkers.[25]

The method of computing network dissimilarity involves storing the PSN as an adjacency matrix which is normalized to a Laplacian of

the network graph followed by its spectral decomposition to obtain the eigen values and eigen vectors. A Forbenious norm of the difference between the adjacency matrices of the two structural networks provides the edge difference score (EDS). A difference in the local and global clustering of residues in the network is captured using the Edge-weighted cosine score (EWCS) and the correspondence scores, respectively. The component scores are then used to compile the Network Dissimilarity Score (NDS) between the compared networks. Identical networks would have a NDS of zero. No two networks can have a dissimilarity greater than $\sqrt{3}$.

2.5 | Control dataset

The global comparison of the structure network was performed earlier on a dataset of single-chain single-domain proteins to determine their structure variability.[26] There, all pairs of available multiple conformers of the same protein are compared to determine the network dissimilarity (NDS) and structure deviation (RMSD) scores. In this work, the data from the previous analysis have been used as a control to understand the significance of the variability obtained upon transient association. The two-sample Kolmogorov–Smirnov test[30] (KS test) is performed to test the significance of the data obtained in the working dataset as compared to the corresponding scores of the control dataset.

3 | RESULTS

The dataset used for the analysis consists of 895 cases, corresponding to a set of proteins that includes enzymes, few transport protein complexes, and other structural oligomers. Proteins from diverse organisms including eukaryotes, plants, and various animals are also present in this working dataset. A total of 366 cases in the dataset are human proteins of which 96 are known to have enzymatic activity. A detailed list of all interacting partners undergoing transient associations that are analyzed in this work have been presented in Table S2 along with additional information regarding each case. In each case, the protein chain from the bound complex is compared against its corresponding chain in the unbound form. The structures are superposed to identify topologically equivalent residues, modeled as nodes in the PSN construct. Those residues that are in contact between the binding partners are considered as interfacial sites and those in contact with the interface are primary contacts.

A threshold of a minimum of 50 residues is used so that globular domains are considered and not peptides. The smallest case with least number of nodes is the ovalbumin enzyme inhibitor bound to different binding partners (PDB ID 1HJA). Here, turkey ovalbumin protein of 51 residues is complexed with the trimeric alpha-chymotrypsin to inhibit its function (with 16 interfacial sites and 21 are primary contact sites). The largest case, from PDB ID 2J8S, is the Acriflavine resistance protein B trimer of 1032 residues in each chain bound to two molecules of designed inhibitor (DARPin). One of the chains from the trimer is unbound and the other two have differential contact. While one chain has 20 interfacial sites with 31 primary contacts, the other has only 7 interfacial sites with 20 primary contacts.

3.1 | Analyzing alteration in PSN upon transient association

Preliminary analysis of the constructed PSNs was performed to study the alteration of basic network parameters like average degree and strength, change in hub status, loss or gain of edges and centrality information. Figure S1 shows the degree of all nodes in the dataset. The maximum degree of any node in all PSNs analyzed in this dataset is 19. Nodes in the PSNs that have a very high degree are special hotspots with high connectivity that can alter the overall network when perturbed. These nodes in the PSNs are identified as hubs. From the degree distribution of the dataset, it is observed that nodes with degree equal to 11 crosses the 90th percentile. Henceforth, for this analysis, any node with degree greater than 10 is considered as a hub node.

The change of degree and strength at all sites between the PSNs of the bound and unbound forms of the proteins were next analyzed. The alteration at the interfacial sites, primary contacts, and non-interfacial sites were considered as the changes at, around, and away from the interface, respectively. Figure S2 summarizes the changes observed across all sites in the dataset. The variation upon transient association are recognized by treating the unbound form as the reference structure and measuring all changes in the network when complexed. Hence, any edges that are unique to the unbound form are considered to be lost and those unique to the bound form are found to be gained. Likewise, altering the hub status of nodes in the networks are also noted. If the size of the protein is large and has a compact structure, there is an increased chance of finding more hubs. Loss or gain of numerous hubs can point to change in local conformation of sidechains. The information on gain and loss of edges and hubs in all cases is reported in Table S1.

Across the dataset, a net gain in hubs is observed more: In terms of hubs, the net gain, loss, and indifference are 437, 333, and 125 cases, respectively. Human serum albumin in complex with beta-2-microglobulin and IgG receptor (PDB ID 4K71) has the highest net gain in the number of hubs (23) across the dataset. However, taking percentage gain proportional to the length of the protein, the highest gain in hubs is observed in another IgG Fab fragment (PDB ID 1NSN) where there is a gain of 15 hubs in the 187-residue structure, which is more than 8% of the protein. Such a tremendous gain in hubs upon binding would suggest that the residues move closer to each other, and the structure would be more compact upon binding. On the other hand, the highest loss (33 hubs and 103 edges) is observed in the case of Glutathione S-transferase when bound to Jasmonic amino acid synthetase (PDB ID 5GZZ). Interestingly, 164 edges are lost, whereas only 61 edges are gained and the loss of hubs from unbound form is not accompanied by any gain of hubs when bound, which suggest that there is a severe loss of connectivity in the protein as a result of protein binding. The information necessary to analyze each case in the dataset is provided in Table S1.

3.2 | Global comparison in working dataset as compared to control

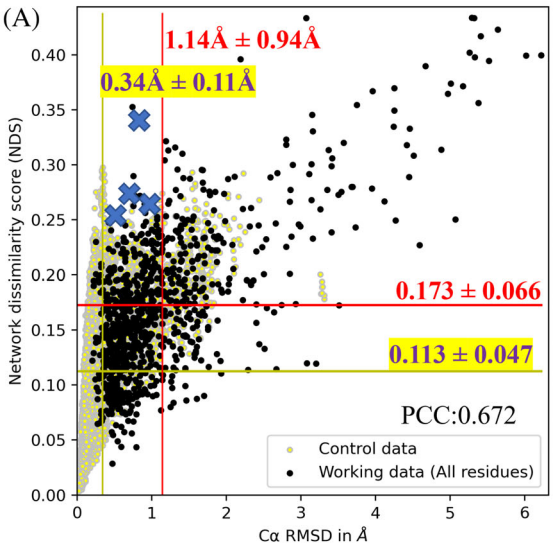
Next, the 895 cases from the bound versus unbound pairs of protein chains are subject to global comparison by computing the NDS and RMSD for structure deviation between backbone conformations. The information obtained from the working dataset is illustrated as a scatter plot as shown in Figure 1A and the comparison scores have been presented in Table S3. Similar information obtained in the control where only multiple structures of the same protein that are known to be single-chain single-domain is also plotted in the background of the scatter.[26] The statistics of the data points corresponding to each of the datasets is marked with horizontal and vertical lines to depict the mean of the data points. The structure deviation in the control was about 0.34 Å with a standard deviation of about 0.3 Å, whereas the mean deviation of the working dataset is about 1.14 Å with a standard deviation of about 0.94 Å. By performing a KS test, it is also understood that the RMSD scores obtained in the working dataset is significantly much higher than what was observed in the control (p-value < .0001). Hence, the spread of the topological variation of the backbone is much larger when proteins undergo transient associations than the structural variability obtained from multiple conformers of the same protein.

Likewise, the network dissimilarity observed in the control had a mean of 0.112 ± 0.047, whereas that observed from the working dataset is 0.173 ± 0.065. Furthermore, KS-test results (p-value < .0001) show that the dissimilarity in the networks is significantly higher in transiently binding proteins as compared to the dissimilarity in PSNs of multiple conformers of the same protein (control dataset). It is interesting to analyze those cases where the variability in the network arises without much change in the backbone conformation. The contribution to network dissimilarity specifically arises from variability in local side-chain conformations. The network comparison scores, and structure deviation of these cases are greater and lower than mean of the dataset, respectively and can be found clustered on the top left of the scatter shown in Figure 1A. A total of 194 cases from the working dataset show such a trend (with RMSD <1.14 Å and NDS >0.173). A list of top 12 cases with such network alterations are listed in Figure 1B, a few highlighted cases are studied in detail.

3.3 | Network alteration without change in the overall topology

The scenario where the protein connectivity is altered upon transient association without much deviation in the backbone conformation are interesting. This suggests that the signal from the interface to non-interface residues can be transmitted via small, concerted, side-chain conformational changes without undergoing any large backbone conformational changes. The data points corresponding to such cases can be found on the upper left quadrant of the scatter plot, where the NDS is high and RMSD is very low. A detailed list of all the cases that

(A)



(B)

Bound	Unbound	NDS	RMSD	Non/Enzyme	Name of protein complex
2J8S : B	2GIF : B	0.353	0.74	Non-Enzyme	Multidrug efflux pump subunit AcrB
2J8S : A	2GIF : A	0.341	0.84	Non-Enzyme	Multidrug efflux pump subunit AcrB
1S3S : F	1E32 : F	0.29	0.75	Enzyme	Transitional ER ATPase
1BVN : P	1PPI : A	0.274	0.67	Enzyme	Pancreatic alpha-amylase
3BUZ : A	1GIR : A	0.272	0.84	Non-Enzyme	Iota toxin component Ia
5ECM : B	5ECS : A	0.272	0.78	Enzyme	Glutathione S-transferase U20
3OUR : A	4I4C : B	0.269	0.76	Enzyme	Esterase FrsA
5T94 : B	1BK6 : B	0.264	0.9	Non-Enzyme	Karyopherin subunit alpha
2YVJ : A	2GR2 : A	0.262	0.75	Non-Enzyme	Ferredoxin reductase
5ECR : B	5ECS : B	0.256	0.91	Enzyme	Glutathione S-transferase U20
1ZY8 : A	1ZMC : G	0.255	0.54	Enzyme	Dihydrolipoyl dehydrogenase
1EWY : A	1GJR : A	0.253	1.07	Enzyme	Ferredoxin--NADP reductase

FIGURE 1 Diversity in structural network and backbone topology upon transient association. (A) The structure deviation information obtained by computing Cα-RMSD (Å) is compared against NDS information obtained by comparing structural networks based on atom contacts. The RMSD is plot on the x-axis and the NDS is plot on the y-axis. The variability observed in 46% of the cases is lower than the mean comparison scores. About 194 cases (21.65%) show strong dissimilarity in network even though there is no significant change in their topology, the alteration of the network in these cases is analyzed in detail. The control data shown in yellow dots correspond to similar information obtained from structure variability of single chain single domain proteins that do not associate transiently. The mean of the working dataset is significantly greater than the mean of the control. (B) Those cases with high network dissimilarity without much structural deviation found on the top left region of the scatter plot as listed. Only the top 12 cases out of which four interesting cases are picked for detailed case studies are highlighted. Their corresponding datapoints are marked on the scatter plot shown on the left. The list of 197 cases that fall in this region are tabulated in Figure S4.

fall into this category can be found in Table S4 along with their computed comparison metrics. Of the 194 cases that show such variability, few interesting cases have been chosen for detailed studies below.

3.3.1 | The acriflavine resistance protein B undergoes network rearrangement

The acriflavine resistance protein B (AcrB) is a multidrug efflux pump found in the inner membrane component of the AcrAB-ToiC drug efflux system found in *Escherichia coli*. The overexpression of this protein machinery in the organism renders it resistant to a wide range of drugs as it functions to efflux the molecules out of the cell. The mechanism of function of the protein have been well studied. There are three distinct domains in each subunit of the AcrB protein, the transmembrane domain (TM-domain), pore domain, and the ToiC docking domain. Drug molecules enter the protein from the central cavity in the TM-domain domain, which are passed into the pore through channels that can open and close based on conformation coupling in the helices of the protein. The molecules efflux into the docking domain via the central funnel.[31–33]

In the bound form, the AcrB trimer is bound to two molecules of designed ankyrin-repeat protein (DARPin) inhibitors (PDB ID 2J8S shown in Figure S3A). This binding mimics the different functional states of the proteins, where the three chains of AcrB are locked in three different states, which reveal the mechanism of drug efflux from the channels of each chain. Alteration of the network around the channel in the pore domain of AcrB (Chain A) is observed upon binding with the DARPin molecule (Figure 2A). The pore domain has an open channel that allows for the efflux of drug molecules, which is closed in the inhibitor bound form of the protein. The closing of this channel is mediated by rearrangement of sidechain interactions facing outwards of the channel to face inwards thus closing the gate. This leads to a significant rearrangement of the network in the protein and hence the high network dissimilarity is observed.

3.3.2 | Net loss of connectivity in pancreatic α-amylase protein when bound to inhibitor

The porcine pancreatic α-amylase protein (PPA) that is secreted by pancreatic acinar cells is responsible for the catalyzing the initial step in starch hydrolysis and is an essential enzyme for producing glucose. Function of the protein is the endohydrolysis of (1–4)-alpha-D-glucosidic linkages in polysaccharides containing three or more (1–4)-alpha-linked D-glucose units.[34] This is performed when the carboxylic oxygens of the catalytically competent residues Glu233 and Asp300

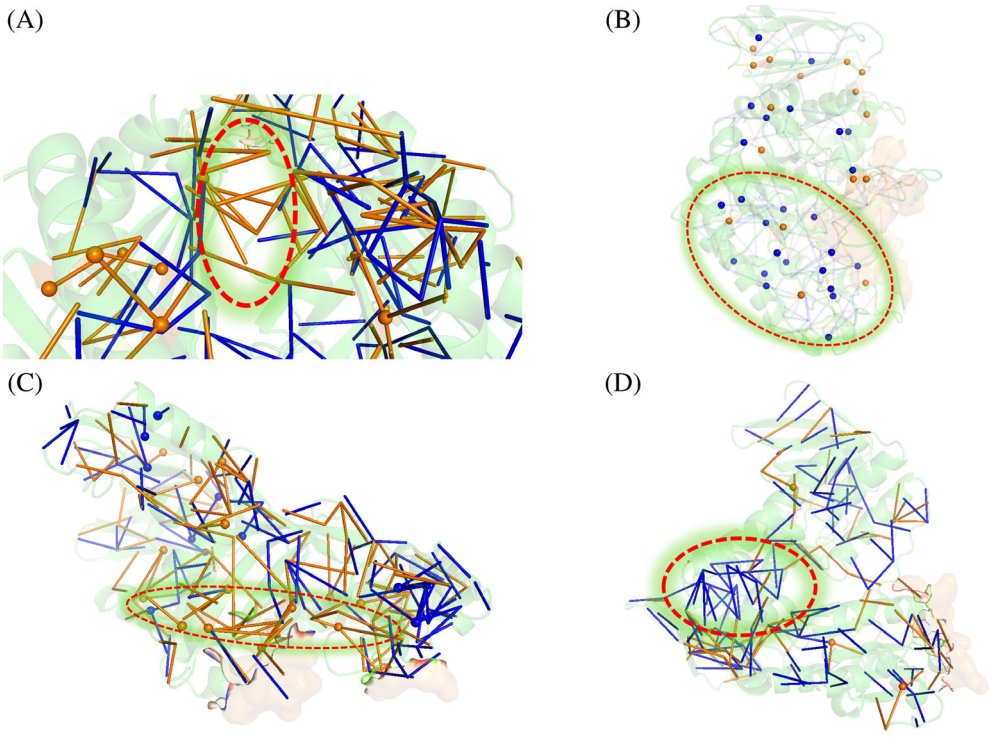


FIGURE 2 Alteration of structural network in case studies analyzed from change in basic network parameters. The observed changes are illustrated on cartoon diagram of the structure networks. Edges that are gained and lost are shown as orange and blue lines, respectively. Those hubs that are gained and lost are shown as orange and blue spheres, respectively. (A) Rearrangement of edges in the pore domain of the chain A in AcrB protein is shown, where most of the edges around the pore are lost and those within the pore are gained. (B) Loss of hubs in the C-terminal domain of the amylase protein is observed. (C) Several edges and hubs are gained around the binding site of the Iota toxin component upon interaction with actin protein. (D) Loss of edges far from the binding site is observed in human DLD protein.

make hydrogen bonds with the "glycosidic" NH group of the acarviosine group. This protein is a major component of pancreatic fluid making it the primary target for the treatment of Type 2 diabetes. Furthermore, its interaction with several inhibitors is well documented as it is studied from perspective of several different diseases.[35]

In our case study, the protein is bound to the microbial inhibitor tendamistat (PDB ID 1BVN shown in Figure S3B).[36] This bound form has the same topological fold as the unbound PPA; however, their network is found to be altered.[37] A net loss of hubs and edges is observed along with loss of essential connectivity around the active site residue, 233E, which functions as a proton donor. Most of the lost edges are found in the C-terminal domain of the proteins, which binds to the tedamistat inhibitor where, 17 hubs are lost, of which one is an active site and only four hubs are gained (Figure 2B). Hence, the connectivity in the functional C-terminal domain is lowered when bound to the inhibitor which also blocks the active site of protein from performing hydrolysis.

3.3.3 | Iota toxin-component has increased connectivity when bound to actin protein

The Iota toxin component Ia is an ADP ribosylating toxin (ADPRT). The C-terminal domain that houses the active site of this protein complexed to NAD is essential for its function. It is known that the active site loop along with residues Tyr60–Tyr62 of the toxin binds to actin protein and inhibits its activity. Actin-specific ADPRTs perform ADP ribosylation of G-actin at Arg117, leading to disorganization of the cytoskeleton and cell death.[38,39]

In our analysis, significant dissimilarity is observed between the actin bound form of the protein (PDB ID 3BUZ shown in Figure S3C) and the unbound toxin. The actin-bound protein has few more new interactions at the binding site and a net overall gain in connectivity. New edges are found to be gained at the functional sites Arg352, Glu378, Glu380 of the protein. A net gain in number of edges and hubs is observed, which shows a more compact structure especially around the binding site (Figure 2C).

3.3.4 | Dihydrolipoyl dehydrogenase loses connectivity away from the binding site

The human mitochondrial dihydrolipoyl dehydrogenase (DLD, hE3 or E3) along with pyruvate dehydrogenase (E1) and dihydrolipoyl transacetylase (E2) form the pyruvate dehydrogenase complex (PDC), which is known to link the glycolysis metabolic pathway with the citric acid cycle. The primary function of the complex is to convert pyruvate to acetyl-CoA, which is necessary for cellular respiration. The DLD protein exists as a homodimer, where each subunit consists of a FAD- and NAD-binding domain along with a central and an interface domain that interacts with the E3-binding protein (E3BP) to form the PDC. At the FAD-binding domain active site, flavin-mediated oxidation takes place to oxidize the substrate, which is reversed to FAD resting state using NAD^+ resulting in NADH and H. Deficiency of this protein is associated with autosomal recessive metabolic disorders.[40–42]

In this case, the E3 homodimer is bound to E3-binding domain of E3BP forming a subcomplex with a strong hydrophobic interface (PDB ID 1ZY8 shown in Figure S3D). It is reported that the central hydrophobic patch along with numerous ionic and hydrogen bonds between residues of the three chains add to the stability of the subcomplex. The network dissimilarity between the bound and the unbound form of this protein is found to be very high. A significant loss in the number of edges, mostly far from the site of the interface is observed (Figure 2D). While 19 hubs are lost all around the protein, only three hubs are gained leading to a net severe loss of connectivity.

4 | DISCUSSION

The ability of a protein to make and break interactions with another protein in order to bind to it temporarily and perform a certain function allows for transient association between the proteins. The binding event is mostly brought about by non-covalent bonds at the interface between the interacting proteins, which also determine the nature and strength of interaction. The implication of such an event can range from no structural change (due to weak interaction) to long range (allosteric communication) of the perturbation signal. The focus of our work was to study how direct associations play a role in perturbing the connectivity of residues in globular proteins. The effect of the perturbation at, around, and away from the interface is studied by constructing a PSN of the connectivity within the protein and analyzing the alteration of the network.

A collection of structures from the ProPairs database was used for the analysis. After filtering structures based on their crystallographic properties, 895 protein chains have been identified that transiently associate with other proteins and have the same oligomeric state in the bound complex as well as the unbound form. The structure of the protein chain obtained as a part of a protein complex, which is the bound form of the protein, is compared with the structure of the same protein chain when it is not bound to the interacting partner but yet in the same oligomeric state. At the local level, the change in network parameters such as number of edges, hubs, and centrality measures are studied. The global comparison is made in terms of the topological change, as in, their backbone deviation (RMSD) and in terms of their structural network dissimilarity (NDS).

Graph spectral comparison methods are used in computing the dissimilarity between the networks that involve spectral decomposition to obtain the eigen vectors and eigen values of the PSNs. Few case studies with high network variation were identified using the global and basic network comparison. A major contribution to the NDS in these cases arises from its component, EWCS, which is mostly responsible to compute the change in local clustering of residues. The Fiedler vectors (Fv) between a pair of PSNs can be examined to identify the sites with high variations in the clustering of nodes. The eigen vector corresponding to the second smallest eigen

value is called the Fiedler vector. This vector can provide meaningful information on the algebraic connectivity of the network and can be used in partitioning the network into clusters.[43] This is illustrated with a case study.

The first step is to identify the Fv from the spectra of the pair of graphs. Figure 3A shows the aligned Fv between the bound and the unbound form of the DLD protein. The absolute difference between the aligned vectors is computed to find regions of the protein that are not in agreement. Figure 3B shows the difference between the Fv and highlights the region with variation between the vectors. The sites with the most variation, having the highest absolute difference are shown as sticks in Figure 3C. The cartoon diagram of the chain A of DLD is colored based on the absolute difference between the vectors and chain B is shown as gray surface. The interacting partner E3BP protein is shown using yellow surface representation. Side chains of the top five residues with highest absolute difference, yellow in the bound form, and red in the unbound form, are shown using spheres.

Any alteration of network parameters close to the site of interface is expected as the interfacial sites make new interactions with the binding partner. However, more often, alterations are also observed far from the site of binding due to allostery, which is the transmission of the perturbation. The path of this allosteric signal can be analyzed by drawing the shortest path between the site of perturbation to the site of significant network alteration. The change in shortest path between the site of binding to the site of perturbation is analyzed in the case study and discussed in the Figure S4. A new edge between spatially proximal nodes GLU 437 and ASP 350 in the bound form reduces the shortest path when compared to the several possible short paths in the unbound form.

Most of the variability observed in the dataset occurs at the non-interfacial sites. This is also evident from the variation of degree and strength at, around, and away from the interface observed in Figure S2. The network variation obtained only by considering the structure network of non-interface sites says that in almost 89% of the cases, the network dissimilarity is greater than 50% of the NDS scored from all residues. This result also suggest that most of the network away from the binding site is affected by the perturbation, but all sites are not perturbed proportionally. As viewed from the absolute difference between Fv of the DLD protein case study, only five nodes of the entire length of the long protein were strongly affected to cluster differently. The effect on all other sites are feeble and is a result of subtle changes in the local conformation of sidechains, which shows that most residues predominantly still remain in the same topology, any interactions that are broken are counteracted by other interactions being made. Hence there is predominantly a rearrangement of interactions that is being observed. However, in about 60% of the 285 cases that are identified as enzymes, in the working dataset, a net loss in connectivity is observed. Hence when the interacting protein is an enzyme, the structure of the bound form may be less compact than the unbound form, which may serve the process of functioning to catalyze several different reactions.

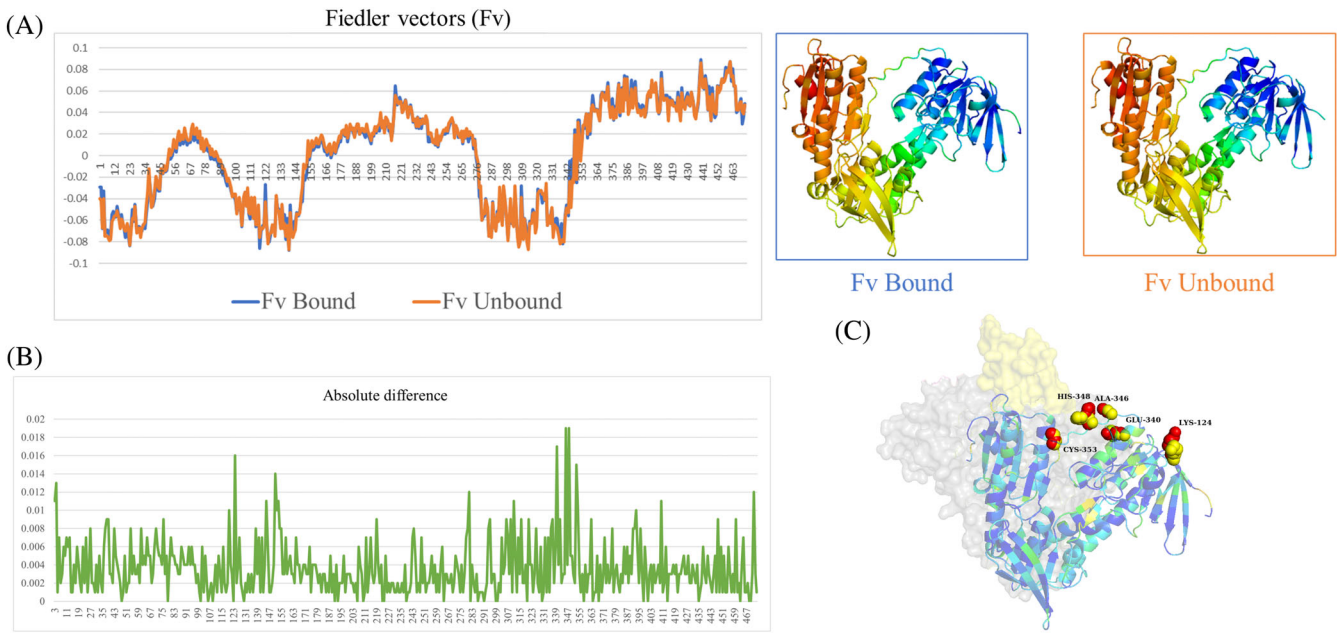


FIGURE 3 Absolute difference between Fiedler vectors (Fv) of DLD protein. Fv components cluster the nodes of the PSN into groups. (A) shows the aligned Fv components of the bound and unbound PSN of DLD protein. The clustering of most sites look almost similar since the topological change in the backbone is not much. (B) The absolute difference between the vectors point to those nodes whose local clustering have changed due to variation of edge weights. When the absolute difference between these vectors is taken, it is found that there are spikes at specific sites that have been perturbed by being clustered differently. (C) These specific sites are mapped onto the cartoon diagram of the bound protein. The binding partners are shown as surface. The sidechains of identified sites in the bound and unbound structures are shown as yellow and red spheres, respectively.

The case studies have been chosen such that their analysis has certain clinical relevance and an impact on understanding their mechanism. The alteration of the network in all the case studies can be related to several human disease conditions like antibacterial resistance, diabetes, toxin-induced cell death, and latic acidosis. We also related to known information about the mechanisms of their function. Hence, the analysis of the structure network is a necessary and beneficial tool in the analysis of structural excursions. The development of such tools that can analyze the impact of protein–protein interactions will help in understanding allostery mechanism and the network analysis of protein structures for stability engineering and docking studies.

5 | CONCLUSION

The alteration of the structure network upon transient interactions between the structure of a protein complex of interacting partners is analyzed by comparing it against an available unbound structure as a reference. Studying the effect of perturbations using a network graph of residue–residue interactions within the protein is a beneficial and robust tool in the analysis of structural excursions. Comparing several basic network parameters and using advanced graph spectral approaches, a local and global difference between the unbound and the bound form of protein chains is identified. It is understood that even when there is no significant change in the overall fold of a given protein, the network of interactions may rearrange themselves to yield a preferred function or phenotype. Change of protein function can be analyzed by studying the change in network parameters at the active site. The protein binding event is found to increase the connectivity within a protein in the case of a non-enzymatic toxin inhibitor. A significant loss of connectivity in the case of DLD protein is associated with the formation of a strong hydrophobic patch in the bound form of this homodimer enzyme. Probing the spectral properties of the PSN yielded Fiedler vectors, which are compared to find nodes that jump from one cluster to another. The path of communication from the site of binding to the highest perturbed site is obtained by building the shortest path between the nodes. These methods of studying the effect of perturbations throw light on understanding the allostery mechanisms.

AUTHOR CONTRIBUTIONS

Vasam Manjveekar Prabantu: Methodology; data curation; investigation; formal analysis; writing – original draft. **Himani Tandon:** Supervision; validation; investigation; project administration; writing – review and editing. **Sankaran Sandhya:** Investigation; validation; project administration; supervision; writing – review and editing. **Ramanathan Sowdhamini:** Project administration; writing – review and editing. **Narayanaswamy Srinivasan:** Conceptualization; methodology; project administration; resources; funding acquisition.

ACKNOWLEDGMENTS

V. M. P. would like to thank the continuous support and mentoring from Professor Saraswathi Vishveshwara and Dr Vasundhara Gadiyaram. This work would not have been completed without the helpful suggestions provided by all our colleagues from N. S. and R. S. lab. This article is dedicated to one of the authors, late Prof N. Srinivasan.

FUNDING INFORMATION

Research from NS group is supported by the Department of Science and Technology (DST), University grants commission (UGC), Department of Biotechnology (DBT), Government of India. R. S. acknowledges funding and support provided by JC Bose Fellowship (JBR/2021/000006) from Science and Engineering Research Board, India and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021). R. S. would also like to thank Institute of Bioinformatics and Applied Biotechnology for the funding through her Mazumdar-Shaw Chair in Computational Biology (IBAB/MSCB/182/2022).

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26606.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Vasam Manjveekar Prabantu https://orcid.org/0000-0001-8024-8708

Ramanathan Sowdhamini https://orcid.org/0000-0002-6642-2367

REFERENCES

1. Huang YJ, Montelione GT. Proteins flex to function. *Nature*. 2005;438(7064):36-37. doi:10.1038/438036a
2. Eisenmesser EZ, Millet O, Labeikovsky W, et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*. 2005;438(7064):117-121. doi:10.1038/nature04105
3. Teilum K, Olsen JG, Kragelund BB. Functional aspects of protein flexibility. *Cell Mol Life Sci*. 2009;66(14):2231-2247. doi:10.1007/s00018-009-0014-6
4. Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002;109(3):275-282. doi:10.1016/s0092-8674(02)00741-9
5. Acuner Ozbabacan SE, Engin HB, Gursoy A, Keskin O. Transient protein–protein interactions. *Protein Eng Des Sel*. 2011;24(9):635-648. doi:10.1093/protein/gzr025
6. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. *Curr Opin Struct Biol*. 2011;21(1):50-61. doi:10.1016/j.sbi.2010.10.002
7. Lee JM, Hammarén HM, Savitski MM, Baek SH. Control of protein stability by post-translational modifications. *Nat Commun*. 2023;14(1):1-16. doi:10.1038/s41467-023-35795-8
8. Sikosek T, Chan HS. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface*. 2014;11(100):20140419. doi:10.1098/rsif.2014.0419
9. Söding J, Lupas AN. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays*. 2003;25(9):837-846. doi:10.1002/bies.10321
10. Tsai CJ, Del Sol A, Nussinov R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol Biosyst*. 2009;5(3):207-216. doi:10.1039/b819720b

11. Swapna LS, Rekha N, Srinivasan N. Accommodation of profound sequence differences at the interfaces of eubacterial RNA polymerase multi-protein assembly. *Bioinformation*. 2012;8(1):6-12. doi:10.6026/97320630008006
12. Tsai CJ, Nussinov R. A unified view of "How Allostery Works". *PLoS Comput Biol*. 2014;10(2):e1003394. doi:10.1371/journal.pcbi.1003394
13. Amitai G, Shemesh A, Sitbon E, et al. Network analysis of protein structures identifies functional residues. *J Mol Biol*. 2004;344(4):1135-1146. doi:10.1016/j.jmb.2004.10.055
14. Taylor NR. Small world network strategies for studying protein structures and binding. *Comput Struct Biotechnol J*. 2013;5(6):e201302006. doi:10.5936/csbj.201302006
15. Faisal FE, Newaz K, Chaney JL, et al. GRAFENE: graphlet-based alignment-free network approach integrates 3D structural and sequence (residue order) data to improve protein structural comparison. *Sci Rep*. 2017;7(1):1-15. doi:10.1038/s41598-017-14411-y
16. Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. *Biophys J*. 2005;89(6):4159-4170. doi:10.1529/biophysj.105.064485
17. Brinda KV, Vishveshwara S, Vishveshwara S. Random network behaviour of protein structures. *Mol Biosyst*. 2010;6(2):391-398. doi:10.1039/b903019k
18. Vijayabaskar MS, Vishveshwara S. Interaction energy based protein structure networks. *Biophys J*. 2010;99(11):3704-3715. doi:10.1016/j.bpj.2010.08.079
19. Bhattacharyya M, Ghosh S, Vishveshwara S. Protein structure and function: looking through the network of side-chain interactions. *Curr Protein Pept Sci*. 2015;17(1):4-25. doi:10.2174/1389203716666150923105727
20. Krishnadev O, Brinda KV, Vishveshwara S. A graph spectral analysis of the structural similarity network of protein chains. *Proteins*. 2005;61(1):152-163. doi:10.1002/prot.20532
21. Vishveshwara S, Dighe A, Gadiyaram V. Graph spectral properties of the sidechain networks of protein structures: implications to Allostery and structure comparison. *Biophys J*. 2019;116(3):463a-464a. doi:10.1016/j.bpj.2018.11.2504
22. Chakrabarty B, Parekh N. PRIGSA: protein repeat identification by graph spectral analysis. *J Bioinform Comput Biol*. 2014;12(6):1442009. doi:10.1142/S0219720014420098
23. Gadiyaram V, Dighe A, Vishveshwara S. Identification of crucial elements for network integrity: a perturbation approach through graph spectral method. *Int J Adv Eng Sci Appl Math*. 2019;11(2):91-104. doi:10.1007/s12572-018-0236-7
24. Ghosh S, Gadiyaram V, Vishveshwara S. Validation of protein structure models using network similarity score. *Proteins*. 2017;85(9):1759-1776. doi:10.1002/prot.25332
25. Gadiyaram V, Ghosh S, Vishveshwara S. A graph spectral-based scoring scheme for network comparison. *J Complex Netw*. 2017;5(2):219-244. doi:10.1093/comnet/cnw016
26. Prabantu VM, Gadiyaram V, Vishveshwara S, Srinivasan N. Understanding structural variability in proteins using protein structural networks. *Curr Res Struct Biol*. 2022;4:134-145. doi:10.1016/j.crstbi.2022.04.002
27. Krull F, Korff G, Elghobashi-Meinhardt N, Knapp EW. ProPairs: a data set for protein-protein docking. *J Chem Inf Model*. 2015;55(7):1495-1507. doi:10.1021/acs.jcim.5b00082
28. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
29. Deb D, Vishveshwara S, Vishveshwara S. Understanding protein structure from a percolation perspective. *Biophys J*. 2009;97(6):1787-1794. doi:10.1016/j.bpj.2009.07.016
30. Hodges JL. The significance probability of the smirnov two-sample test. *Arkiv för Matematik*. 1958;3(5):469-486. doi:10.1007/bf02589501/metrics
31. Sennhauser G, Amstutz P, Briand C, Storchenegger O, Grütter MG. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol*. 2006;5(1):e7. doi:10.1371/journal.pbio.0050007
32. Seeger MA, Schiefner A, Eicher T, Verrey F, Diederichs K, Pos KM. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*. 2006;313(5791):1295-1298. doi:10.1126/science.1131542
33. Wang Z, Fan G, Hryc CF, et al. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *Elife*. 2017;6:e24905. doi:10.7554/elife.24905
34. Date K, Satoh A, Iida K, Ogawa H. Pancreatic α-amylase controls glucose assimilation by duodenal retrieval through N-glycan-specific binding, endocytosis, and degradation. *J Biol Chem*. 2015;290(28):17439-17450. doi:10.1074/jbc.m114.594937
35. Yadav R, Bhartiya JP, Verma SK, Nandkeoliar MK. The evaluation of serum amylase in the patients of type 2 diabetes mellitus, with a possible correlation with the pancreatic functions. *J Clin Diagn Res*. 2013;7(7):1291-1294. doi:10.7860/jcdr/2013/6016.3120
36. Wiegand G, Epp O, Huber R. The crystal structure of porcine pancreatic α-amylase in complex with the microbial inhibitor tendamistat. *J Mol Biol*. 1995;247(1):99-110. doi:10.1006/jmbi.1994.0125
37. Qian M, Haser R, Payant F, Buisson G, Duée E. The active center of a mammalian α-amylase. Structure of the complex of a pancreatic α-amylase with a carbohydrate inhibitor refined to 2.2-Å resolution. *Biochemistry*. 1994;33(20):6284-6294. doi:10.1021/bi00186a031
38. Tsuge H, Nagahama M, Nishimura H, et al. Crystal structure and site-directed mutagenesis of enzymatic components from *Clostridium perfringens* iota-toxin. *J Mol Biol*. 2003;325(3):471-483. doi:10.1016/S0022-2836(02)01247-0
39. Tsuge H, Nagahama M, Oda M, et al. Structural basis of Actin recognition and arginine ADP-ribosylation by *Clostridium perfringens* ι-toxin. *Proc Natl Acad Sci U S A*. 2008;105(21):7399-7404. doi:10.1073/pnas.0801215105
40. Brautigam CA, Chuang JL, Tomchick DR, Machius M, Chuang DT. Crystal structure of human dihydrolipoamide dehydrogenase: NAD+/NADH binding and the structural basis of disease-causing mutations. *J Mol Biol*. 2005;350(3):543-552. doi:10.1016/j.jmb.2005.05.014
41. Ciszak EM, Makal A, Hong YS, Vettaikkorumakankauv AK, Korotchkina LG, Patel MS. How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex. *J Biol Chem*. 2006;281(1):648-655. doi:10.1074/jbc.m507850200
42. Brautigam CA, Wynn RM, Chuang JL, Machius M, Tomchick DR, Chuang DT. Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex. *Structure*. 2006;14(3):611-621. doi:10.1016/j.str.2006.01.001
43. Gadiyaram V, Vishveshwara S, Vishveshwara S. From quantum chemistry to networks in biology: a graph spectral approach to protein structure analyses. *J Chem Inf Model*. 2019;59(5):1715-1727. doi:10.1021/acs.jcim.9b00002

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Prabantu VM, Tandon H, Sandhya S, Sowdhamini R, Srinivasan N. The alteration of structural network upon transient association between proteins studied using graph theory. *Proteins*. 2023;1-9. doi:10.1002/prot.26606