NAPS

Network Analysis of Protein Structures

Reference Manual

April 10, 2016

Broto Chakrabarty and Nita Parekh,
Centre for Computational Natural Sciences and Bioinformatics,
International Institute of Information Technology-Hyderabad, India

Contents

1. About NAPS

NAPS is an online portal for the construction and analysis of Protein Contact Network (PCN), also referred as Residue Interaction Network (RIN), or protein structure graph (PSG). A protein structure can be represented as a network by providing the PDB id (or uploading the PDB file) on the portal: http://bioinf.iiit.ac.in/NAPS/index.php. The snapshot of the home page is shown in Figure 1.

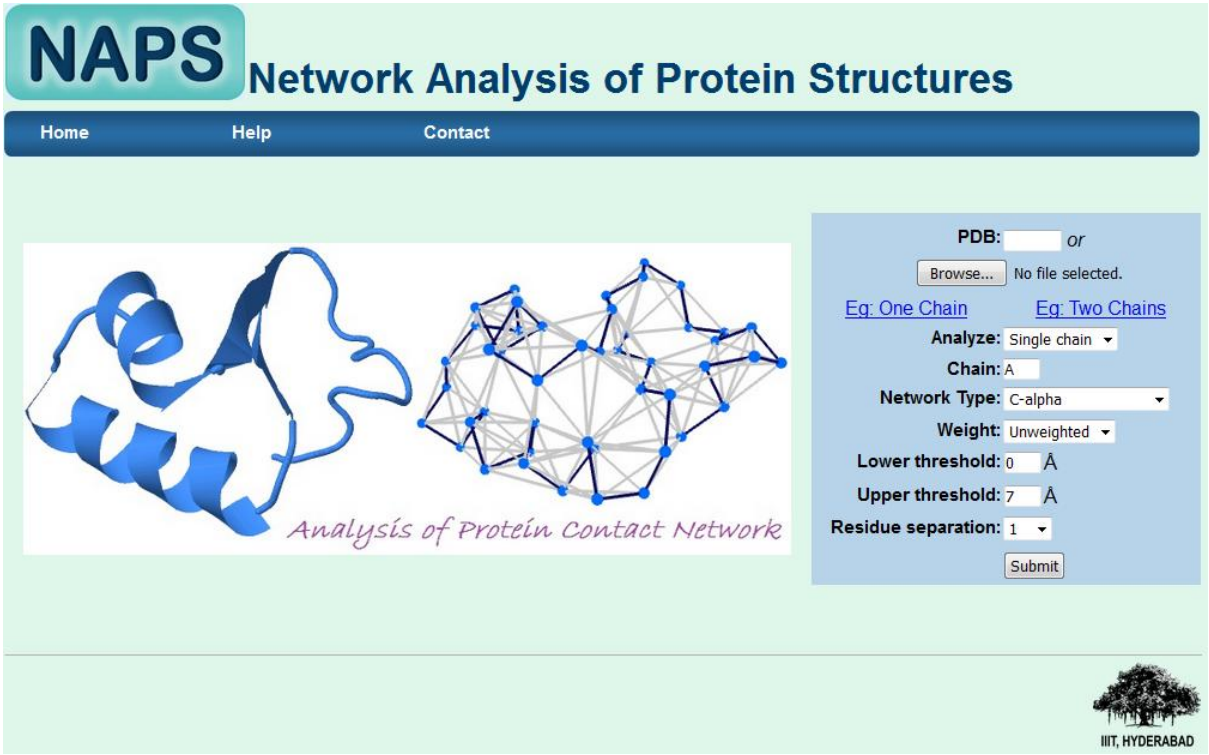

Figure 1: Home page of NAPS portal.

2. System requirement

The portal is browser independent and needs no added plugins to be installed in the browser. On providing the PDB id (or upload PDB file), the construction of the network and other computations are performed on the backend server and the data is sent back to the frontend web browsers. The 3D structure of the protein, network visualization and other features are displayed using javascript applets which are browser independent and utilize the client resources (user system resources). Therefore the upper limit of the structure size that can be handled by the portal is determined by the system configuration of the user. For each page, data transfer to/from the server is required only once while loading the page, after which all the activities on the page are performed at the client site with no data transfer required to/from the server. This enhances the performance speed of the portal.

The network visualization is performed using WebGL which is present in all modern browsers. Some of the popular browsers with WebGL are:

Firefox Moxilla: Version 4.0 and above

Google Chrome: Version 9 and above

Safari: Version 6.0 and above installed on OS X Mountain Lion, Mac OS X Lion and Safari 5.1 on Mac OS X Snow Leopard.

The portal displays an error message if WebGL support is not available. If the browser version is above the minimum required version stated above, it implies that WebGL is turned off in the browser. The user needs to manually enable WebGL in such a case.

3. Network construction

The portal provides an option to construct following types of network representations of a protein structure, depending on the type of analysis required.

a) **C-alpha**: An amino acid residue represented by the C-alpha atom is considered as node in the network and an edge is constructed if the distance between a pair of C-alpha atoms is within the lower and upper thresholds defined by the user (default upper threshold = 7 Å; lower threshold = 0 Å).

   i. **Unweighted**: All edges are considered equally important.

   ii. **Weighted**: Edge weight for a C-alpha weighted network is given by:

$$w_{ij} = \frac{1}{d_{ij}}$$

   where d_{ij} is the euclidean distance between C-alpha atoms of i^{th} and j^{th} residues.

b) **C-beta**: An amino acid residue represented by the C-beta atom is considered as node in the network and an edge is constructed if the distance between the C-beta atoms (C-alpha for GLY) is within the lower and upper thresholds defined by the user (default upper threshold = 7 Å; lower threshold = 0 Å).

   i. **Unweighted**: All edges are considered equally important.

   ii. **Weighted**: Edge weight weighted network is given by:

$$w_{ij} = \frac{1}{d_{ij}}$$

   where d_{ij} is the euclidean distance between C-beta atoms of i^{th} and j^{th} residues.

c) **Atom pair contact**: An amino acid residue is considered as node in the network and an edge is constructed if the distance between any pair of atoms of the residue pair is within the lower and upper thresholds defined by the user (default upper threshold = 5 Å; lower threshold = 0 Å).

   i. **Unweighted**: All edges are considered equally important.

   ii. **Weighted**: Edge weight is given by the number of atom pairs within cutoff distance.

d) **Centroid (centre of mass)**: An amino acid residue is considered as node in the network and an edge is constructed if the distance between centre of mass of the residue pair is within the lower and upper thresholds defined by the user (default upper threshold = 8.5 Å; lower threshold = 0 Å).

   i. **Unweighted**: All edges are considered equally important.

   ii. **Weighted**: Edge weight weighted network is given by:

$$w_{ij} = \frac{1}{d_{ij}}$$

   where d_{ij} is the euclidean distance between centre of mass of i^{th} and j^{th} residues.

e) **Interaction strength**: An amino acid residue is considered as node in the network and an edge is constructed if the interaction strength between two residues is more than the threshold defined by the user (default = 4%) (1). The interaction strength as proposed by Brinda and Vishveshwara (2) is calculated as:

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i * N_j}} * 100$$

where, n_{ij} is the number of side chain atom pairs of the residues *i* and *j* within 4.5 Å. N_i and N_j are the normalization values of the residues i and j given by Kannan and Vishveshwara (1) as shown in Table 1.

i. **Unweighted**: All edges are considered equally important.

ii. **Weighted**: The interaction strength (I_{ij}) is considered as edge weight.

The choice of network type and threshold depend on the biological problem to be addressed. The different network types and some example problems for which they have been used in the past, are listed in Table 2. The analysis of one protein structure as an unweighted network is shown in sections 4 – 11. The advanced options for analysis of protein complex are provided in section 12.

Long range interaction network (LIN)

A LIN is constructed by considering edges between residues that are sequentially separated by 10-12 residues along the protein backbone. LIN can be constructed by selecting a suitable threshold residue separation at the home page. A range of 1 to 15 is available as threshold for residue separation. Minimum residue separation of 1 is used as default, where edge is drawn between any pair of residues satisfying the criteria of network construction, including the adjacent residues of the protein backbone. Threshold of 2 or 3 is usually used to remove noise in the network.

Table 1: Normalization value for amino acid residues used to construct Interaction Strength Network.

Residue Type	**Norm (*N*)**
Alanine	55.7551
Arginine	93.7891
Asparagine	73.4097
Aspartic acid	75.1507
Cystine	54.9528
Glutamine	78.1301
Glutamic acid	18.8288
Glycine	47.3129
Histidine	83.7357
Isoleucine	67.9452
Leucine	72.2517
Lysine	69.6096
Methionine	69.2569
Phenyl alanine	93.3082

Proline	51.3310
Serine	61.3946
Threonine	63.7075
Trptophan	106.703
Tyrosine	100.719
Valine	62.3673

Table 2: Different network types and the purpose they are used.

Network Type	**Edge weight**	**Purpose**
C_α	Unweighted	Analysis of global network properties (3, 4), Protein folding kinetics (5), Analysis of inter- and intra molecular, communications (6), Identification of proteins with similar folds (7), Structural repeat identification (8–10)
C_β	Unweighted	Protein dynamics (11)
Atom pair contact	Unweighted	Protein fold (12)
	Weighted	Network analysis based on physicochemical properties (13)
Centroid	Unweighted	Protein core and exposed residue analysis (14)
Interaction strength	Weighted	Structural stability (2), Identifying side chain clusters (1)

4. Global properties

The global properties of the network are displayed in the network property page along with details of network construction. This is the first page shown after construction of the network. The page can be browsed by selecting 'Properties' option from the pull-down menu of 'Analysis' tab at the top menu bar. The snapshot of the page for an example protein 1CRN (chain A) is shown in Figure 2.

The following global parameters are displayed on the network property page:

a) **N_r:** Number of nodes in the network. This represents the number of residues in the protein (or complex).

b) **N_e:** Number of edges in the network.

c) **D:** Shortest path distance between the pair of farthest nodes in the network.

d) **R**: Shortest path distance of the centre(s) of the network to the farthest node.

e) **k:** Average degree of all the nodes in the network.

f) **l**: Average of shortest path between all node pairs in the network.

g) **c**: Clustering coefficient of the network

The following network analysis links are available on the network property page:

a) View Network: Network visualization as described in section 5.

b) Node centrality: Centrality analysis as described in section 6.

c) Edge centrality: Edge centrality as described in section 7.

d) Shortest path: Shortest path analysis as described in section 8.

e) *k*-clique: *k*-clique analysis as described in section 9.

f) Graph spectra: Graph spectral analysis of adjacency and laplacian matrices as described in section 10.

g) Domain view: Visual analysis of multi-domain proteins as described in section 11.

The user can also select the type of analysis to be performed from the pull-down menu of the tab 'Analysis' from menu bar.

Note: The global parameters, node centrality and edge centrality are not available if the network has more than one connected components. A network is called one connected component, if at least one path exists between each node pair of the network.

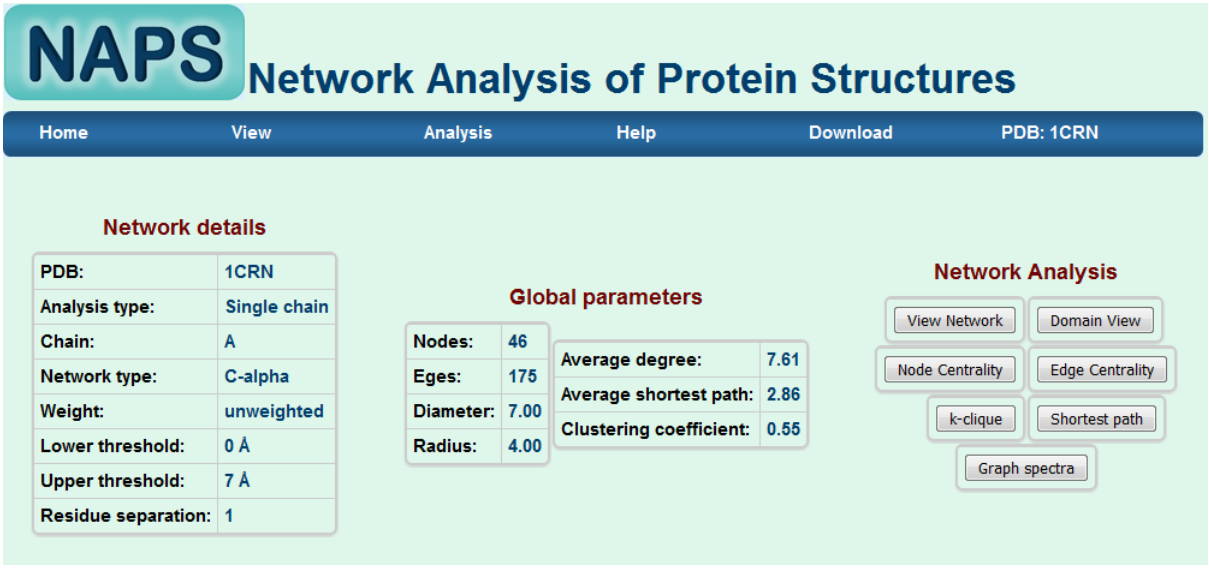

Figure 2: Properties of the Network.

The download option on the menu bar of this page provides an option to download:

a) Global parameters in a text file.

b) Edgelist file.

5. Network visualization

The network view page is dedicated to different options for visualizing the network along with the three dimensional structure of the protein. The network view page can be accessed from other analysis pages by selecting 'View' option from the top menu bar. The 3D structure of the protein is shown using open source javascript based applet, JSmol (15). There are four types of visualization options available in NAPS:

a) 3D structural view

The 3D structure of the protein is shown using open source javascript based applet, JSmol (15).

b) Network View

Depending on the type of network construction chosen, a 3D graphical view is shown in the left panel. To correlate the network to the protein structure, the nodes are plotted using the actual coordinates of the representative atoms (C-alpha/C-beta/centroid) of the corresponding residues in the PDB file. The view page with network view and 3D structure view of the protein 1CRN (chain A) is shown in Figure 3.

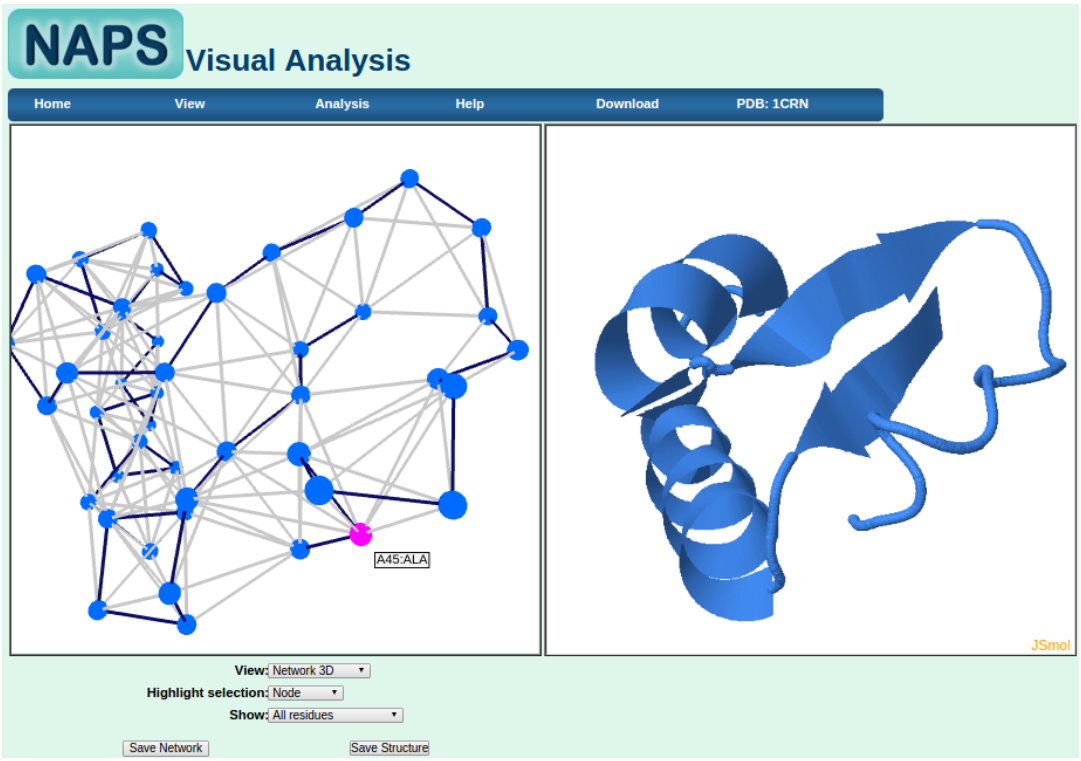

Figure 3: Network view page with 3D network representation.

The nodes in the network view are represented in blue color, while two types of edges are shown, dark blue and grey. The dark blue edges trace the backbone connectivity of the protein structure, while the grey edges represent all other nodes within a user defined cutoff threshold of a node. On hovering the mouse over a node in the network view, the node is highlighted with pink color and the chain number, residue number and residue type corresponding to the node are displayed, as shown in Figure 3. Two highlight selection options are provided on the network view page:

i. **Highlight node:** On selecting a node(s) by clicking, it gets highlighted with red color and the corresponding residue on the 3D structure representation also gets highlighted with red color as shown in Figure 4.

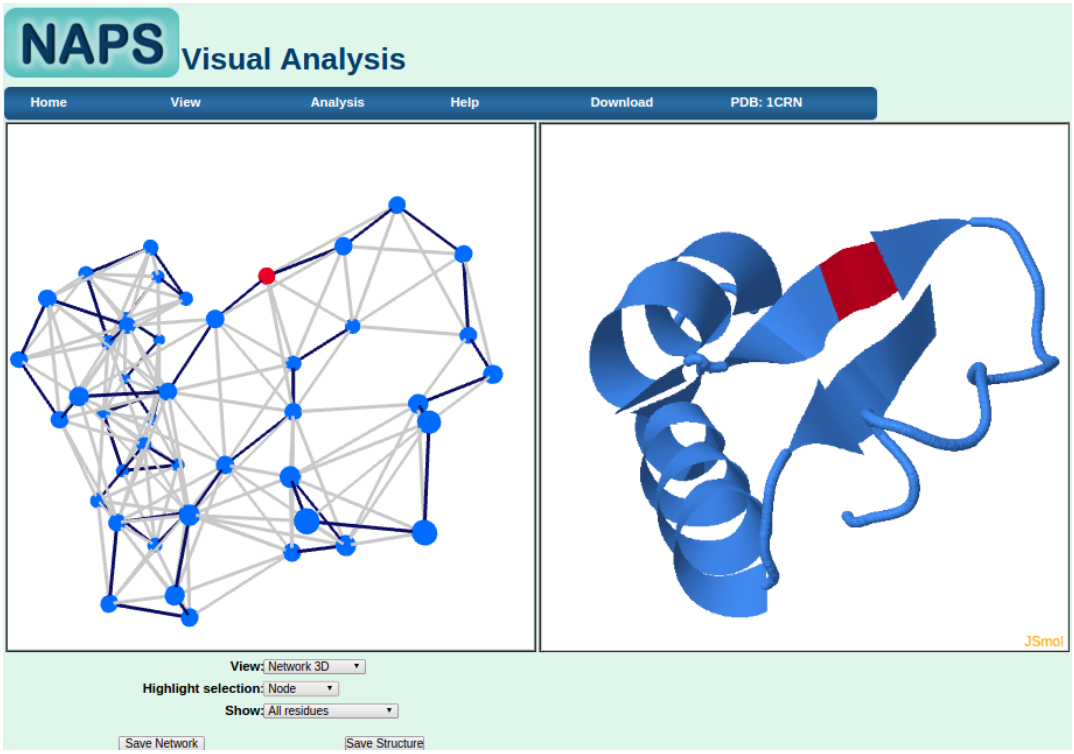

Figure 4: 3D representation with node selection.

ii. **Highlight neighbor:** In a dense network like PCN with many grey edges originating from a node, it is desirable to have easy identification of the immediate neighbors of a node. On selecting a node with highlight selection as neighbor, the node and the corresponding residue in JSmol applet are highlighted in red color, and its immediate neighboring nodes in the network view and the corresponding residues in JSmol applet are highlighted in yellow as shown in Figure 5. The residue ids of the selected node and its neighbors are displayed below the JSmol applet.

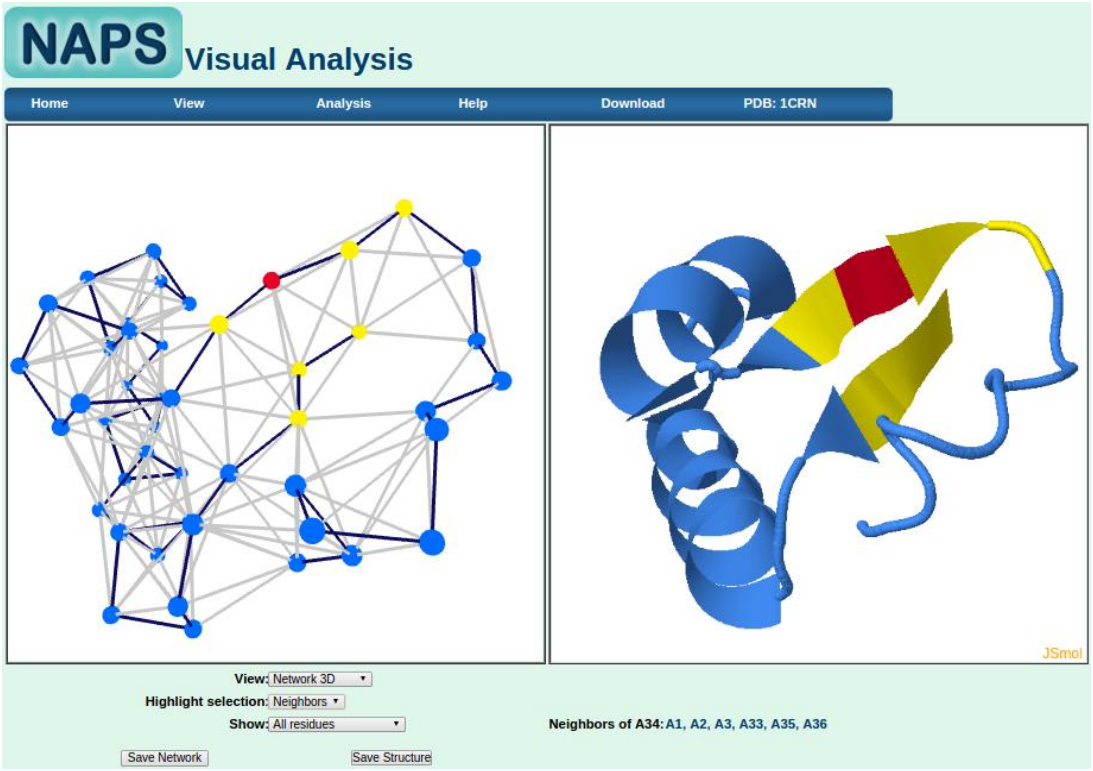

Figure 5: 3D Network view with highlight selection as neighbors.

c) Contact map view

Contact map view is a two-dimensional dot matrix representation of the network where a dot (*i*, *j*) represents an edge between the i^{th} and j^{th} nodes and is shown in Figure 6. The contact map view helps in visual inspection of interaction patterns and the long range interactions within the protein structure. The pattern of connectivity within a secondary structure remains conserved which can be observed in Figure 6 where the edges within helix are highlighted in red color. The residue id, chain and residue type of the two amino acids forming the edge are displayed on taking the mouse on an edge in the contact map view. On selecting the edge, the corresponding residues forming the edge are highlighted by red as shown in Figure 6. An option to add grid lines at intervals of 10 residues is provided as checkbox.

The download option on the menu bar of the visualization page provides allows download of:

a) Network view image in PNG format.

b) Structure view image in PNG format –Network view or Contact map, whichever is selected.

c) Edgelist - a text file listing the node pairs sharing an edge between them.

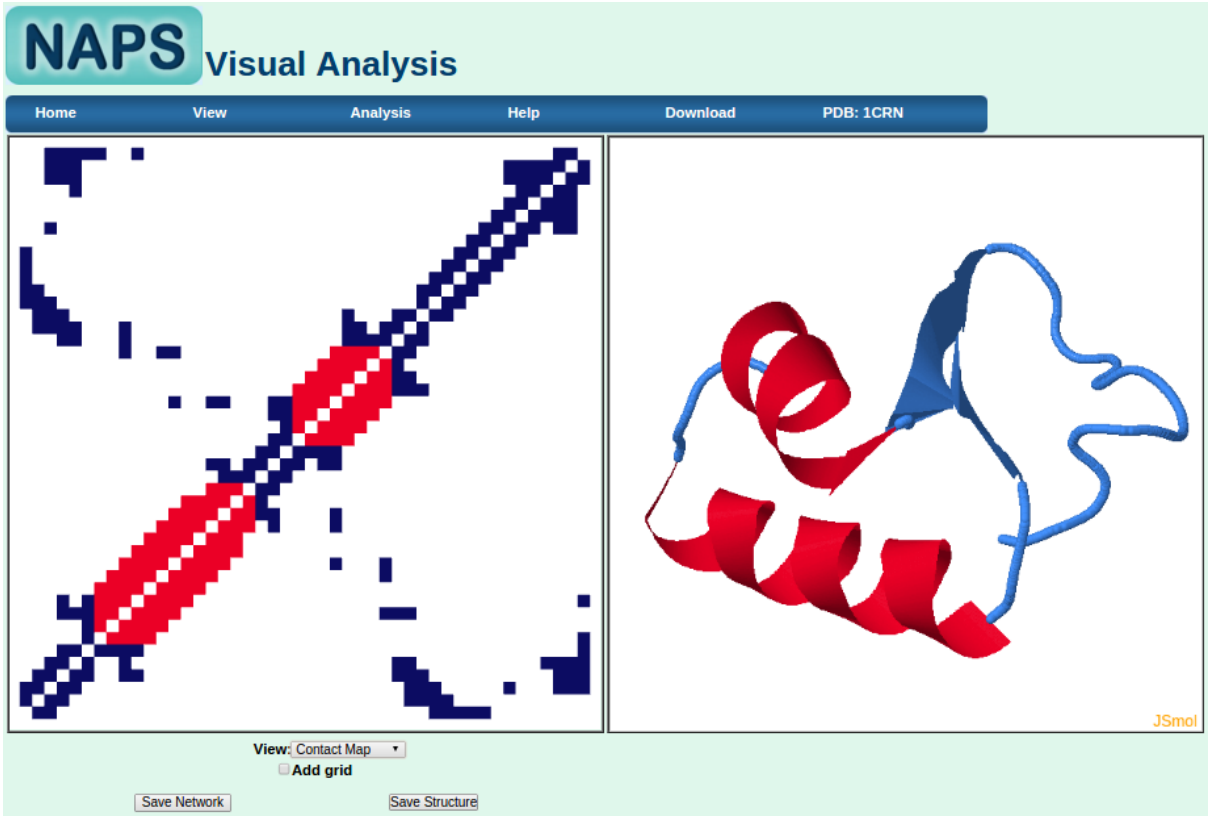

Figure 6: Contact Map view.

d) Distance matrix view

The distance matrix captures the distance between the amino acid residues. The element (*i*,*j*) of the matrix represents the distance (in Å) between i^{th} and j^{th} residues. The distance matrix view provides a 2D representation of the distance matrix with color representing the distance between the residues as shown in Figure 7. The residue id, chain and residue type of the two amino acids represented by a cell in the distance matrix, are displayed on taking the mouse on the distance matrix cell. An option to add grid lines at intervals of 10 residues is provided as checkbox.

Note: Contact map and distance matrix views are available for systems with up to 750 residues.

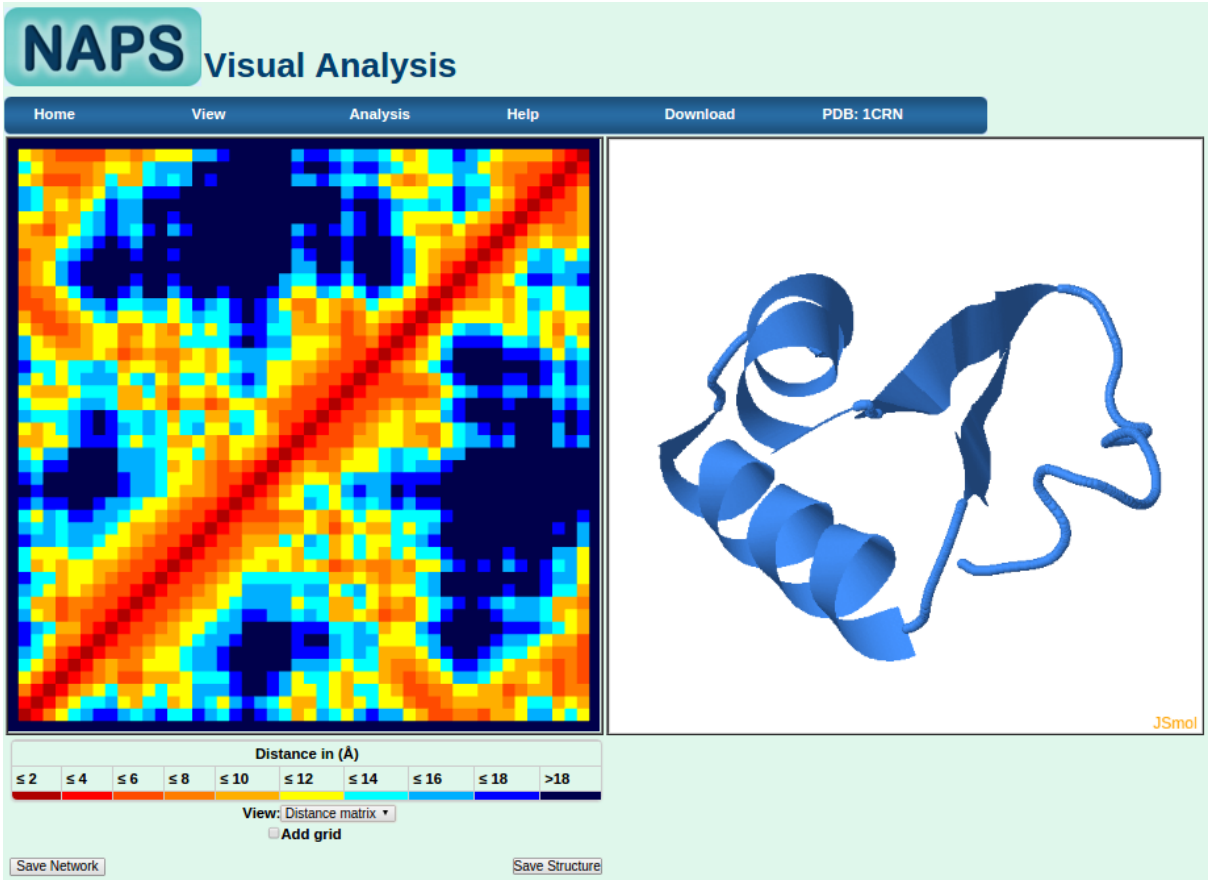

Figure 7: Distance matrix representation of the network.

Sub-network view

Visual analysis of a sub-network based on physicochemical properties of the residues can be performed. The network 3D view provides an option to select specific residue types: hydrophobic, hydrophilic and charged. On selecting one of the residue type, all the residues with the selected physicochemical property are highlighted in the network 3D view and the JSmol applet, and edges are shown only between the highlighted residues, as shown in Figure 8 for hydrophobic residues.

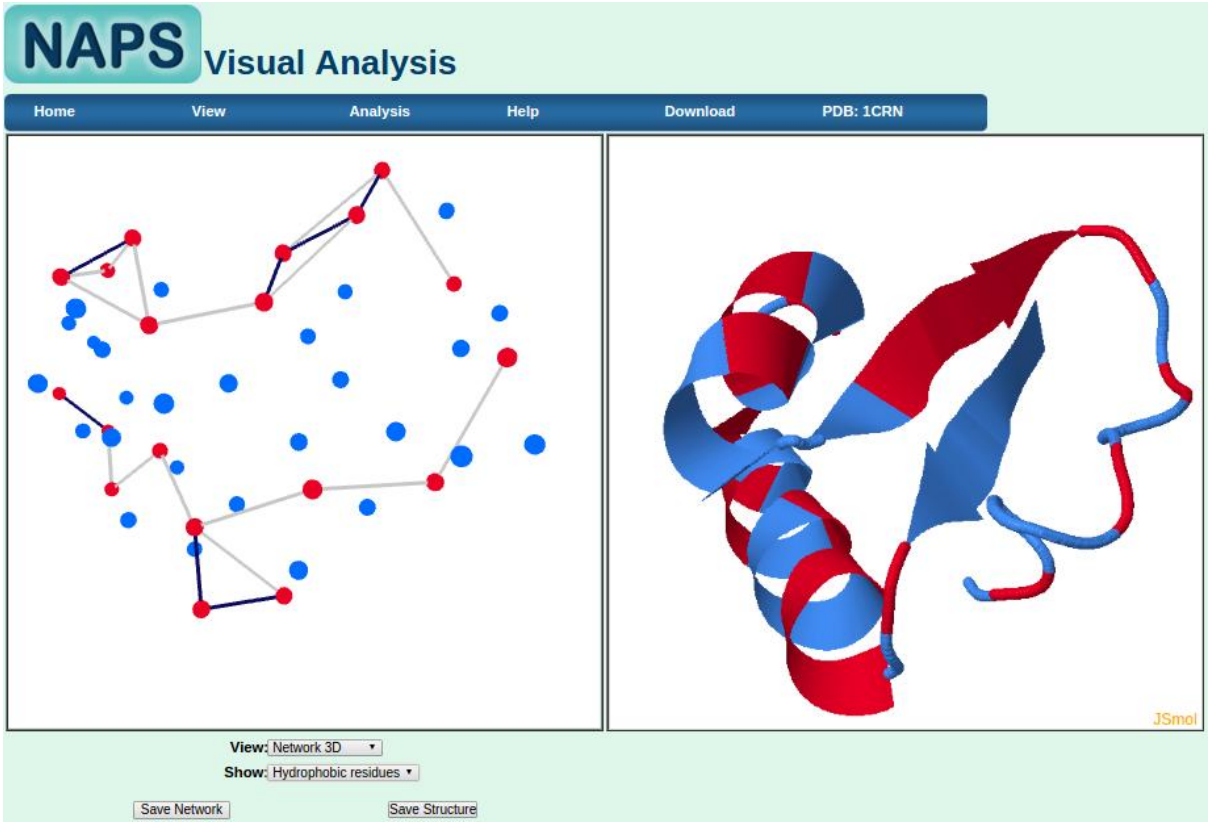

Figure 8: Sub-network representation of network 3D view showing Hydrophobic residues and edges between them.

6. Node centrality analysis

Centrality measure of a node provides a quantification of the topological importance of the node in the network. Different centrality measures have been proposed for ranking the nodes in a complex network and quantifying their relative importance. The analysis of centrality measures for a PCN can be performed by selecting the 'Node centrality' option from the pull-down menu of 'Analysis' tab at the top menu bar. Here we provide option to compute seven node-based centrality measures:

a) **Degree:** Number of direct neighbors of a node.

b) **Closeness:** It is the inverse of total shortest path distance of the node to all other nodes of the network.

c) **Betweenness:** It is the ratio of shortest paths passing through the node.

d) **Clustering coefficient:** It is the ratio of number of connected neighbors to the total number of connections possible between the neighbors.

e) **Eigenvector centrality:** It is the eigenvector component corresponding the largest eigenvalue of the adjacency matrix.

f) **Eccentricity:** Shortest path distance of the node to the farthest node in the network.

g) **ANN degree:** Average of degree of its immediate neighbors.

h) **Strength:** Weighted degree represented by cumulative weights of all the edges connected to a node. This is applicable only for weighted networks.

On the centrality analysis page, a table lists the centrality value for each residue for a chosen centrality measure, on the right panel. The interactive network and 3D structure view are displayed on the left panel. The user can select the centrality measure from a drop down menu, can sort it based on the residue id or in descending order of the centrality value. One can also analyze the centrality value of any node by selecting it on the interactive network. The corresponding residue in JSmol applet is simultaneously highlighted along with the corresponding entry in the table, in red color. Snapshot of the interactive analysis of centrality is shown in Figure 9.

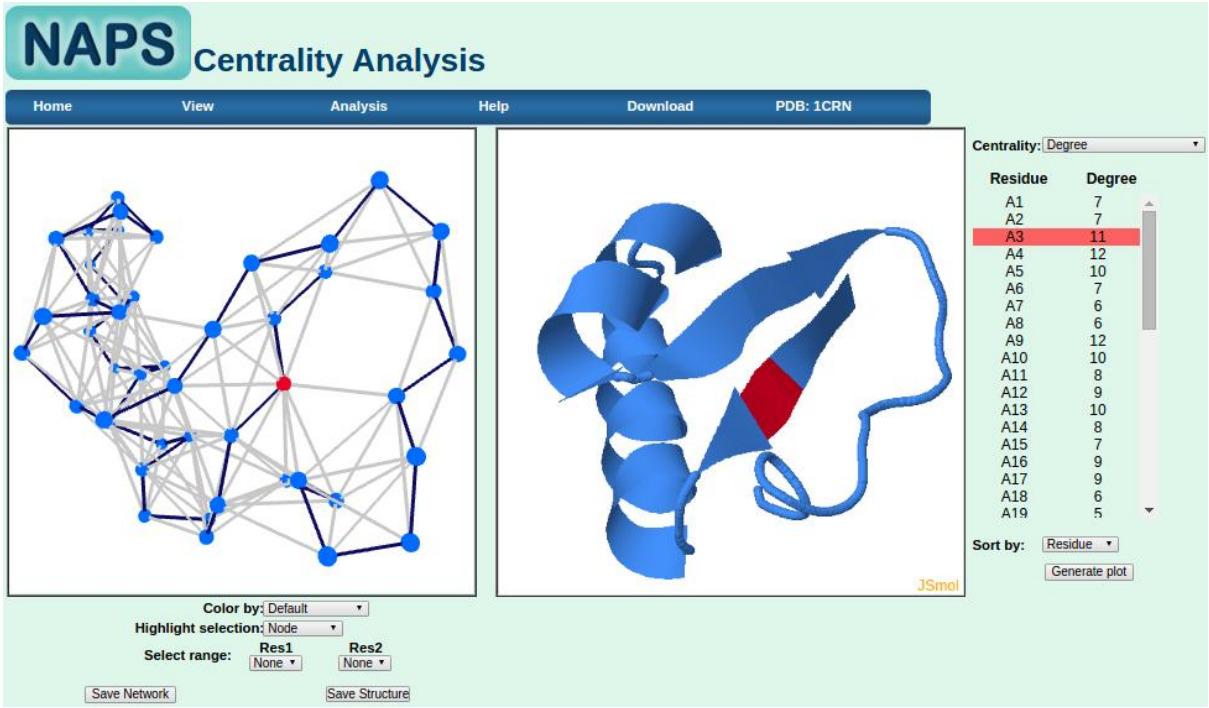

Figure 9: Centrality page showing degree centrality with residue A2 highlighted.

For each centrality measure, the following options are provided to aid in the analysis of centrality measures:

a) **Color by centrality:** By changing the color option from default to centrality, the nodes in the network view and the corresponding residues in the JSmol applet are colored according to the centrality values in a gradient of red (maximum) to yellow (minimum), as shown in Figure 10. This helps in easy identification of residues with high centrality values, which are likely to be the residues important for the 3-dimensional fold of the protein, or functionally important (16).

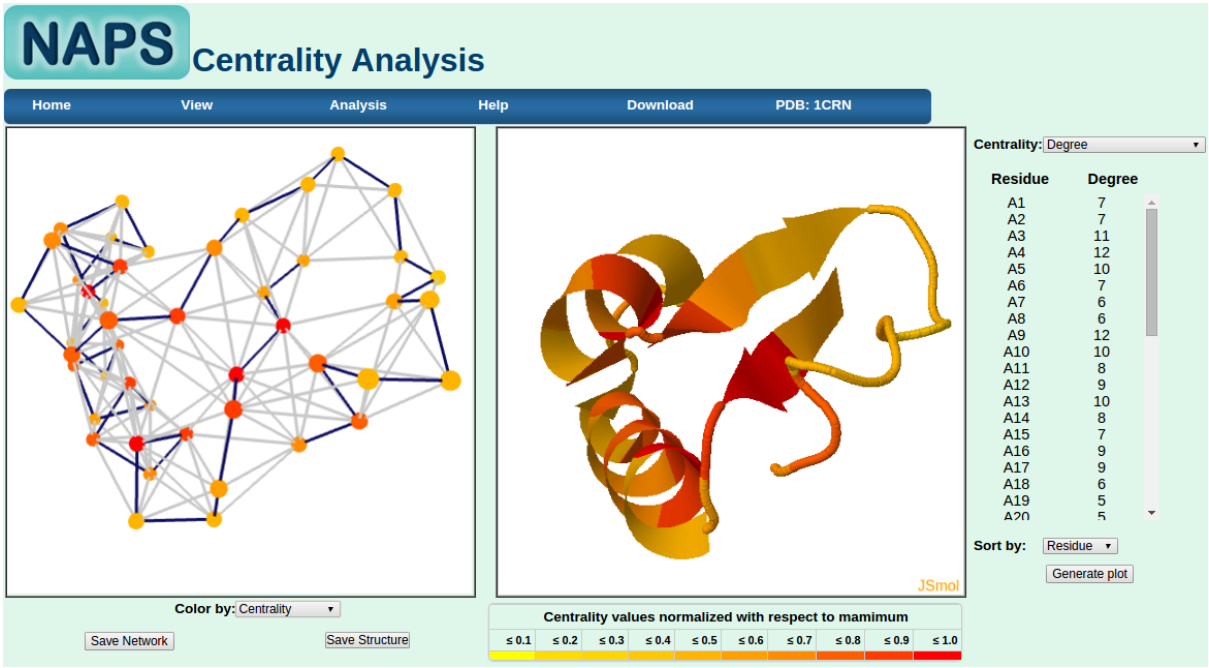

Figure 10: Nodes colored according to centrality values.

b) **Color by Hydrophobicity:** By changing the 'Color by' option to 'Hydrophobicity', the nodes in the network view and the corresponding residues in the JSmol applet are colored according to the hydrophobicity values of the amino acid residue. The hydrophobicity indices for 20 amino acids given by Kyte and Doolittle range between -4.5 to 4.5 (17). These values are colored in gradients of red to show hydrophobic residues and gradients of blue to show hydrophilic residues, as shown in Figure 11.

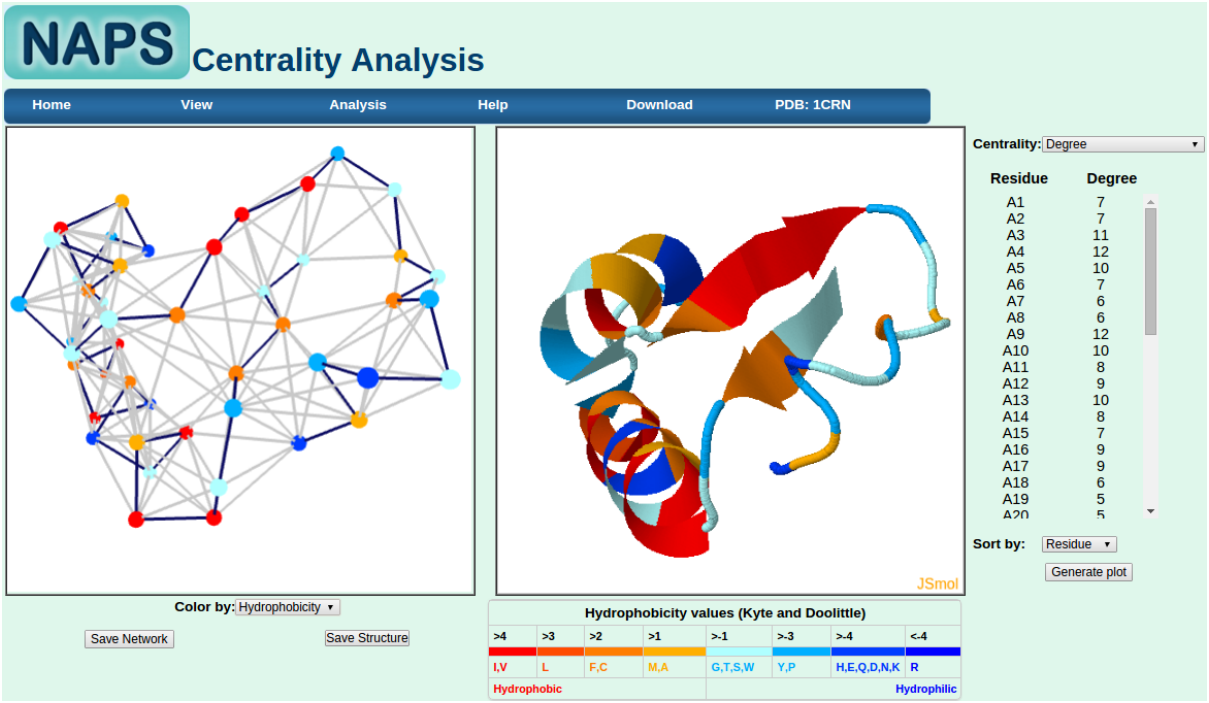

Figure 11: Nodes colored according to hydrophobicity.

c) **Centrality Plots:** By clicking the 'Generate plot' button, one can obtain a plot of the centrality measure along the length of the protein, as shown in Figure 12. A new page opens where the user can select the file type, line color and resolution of the plot. The user can generate high quality images of the centrality plot for publication purposes and save it by clicking on the download button.

d) **Highlight residue range:** A set of specific residues or a range of residues can be highlighted by selecting the residue ids or the range from the drop down menu named 'Select range' provided with the 'Color by' option as 'default'. The high centrality nodes obtained from the sorted centrality table can be highlighted using this option.

e) **Highlight neighbor:** It is desirable to identify the immediate neighbors of high centrality nodes of a network. Clicking a node with highlight selection as 'neighbor', the node is highlighted by red and its neighbors are highlighted by yellow color. The feature is described in detail in Network Visualization section.

From the download option on the menu bar of this page the user can download:

a) Centrality values in tab separated text file.

b) Plots of Centrality measures in PNG format.

c) Network image in PNG format.

d) Structure image in PNG format.

e) Edgelist in tab separated text file.

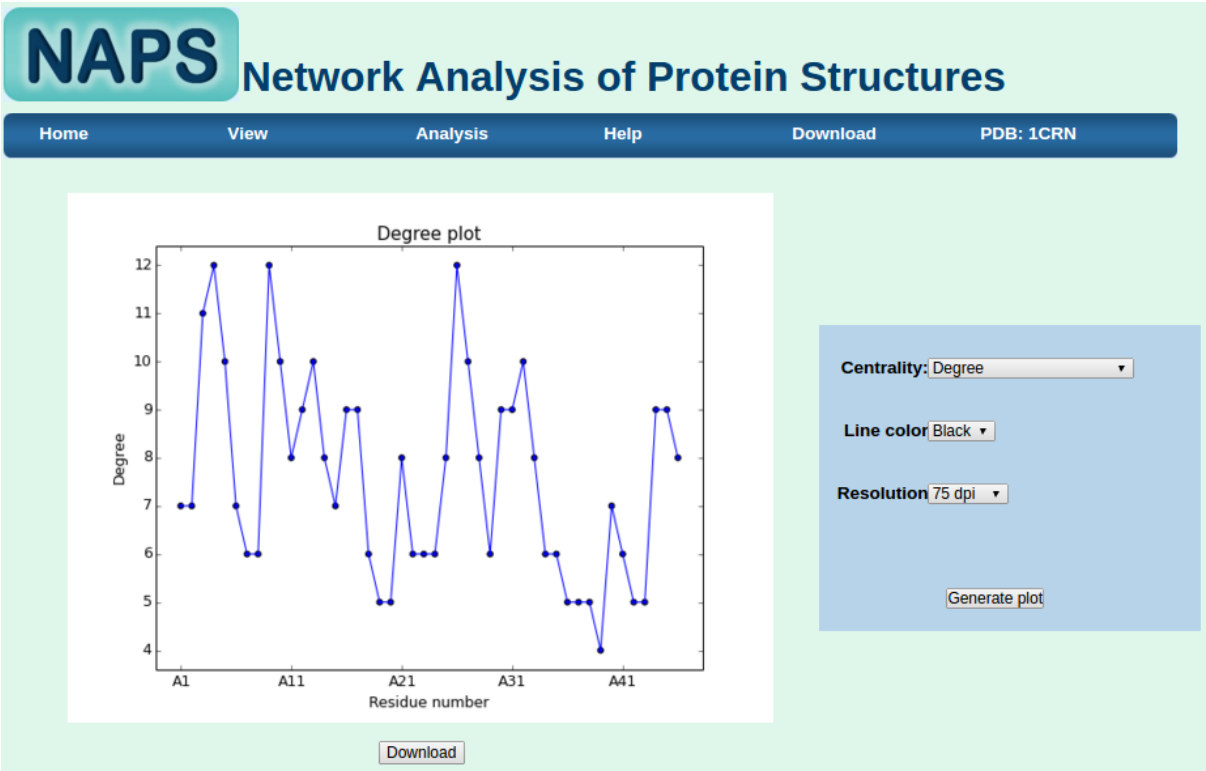

Figure 12: Plot of degree centrality.

7. Edge centrality:

The analysis of edge centrality measures for a PCN can be performed by selecting the ‘Edge centrality’ option from the pull-down menu of ‘Analysis’ tab at the top menu bar. The contact map view of the network is displayed along with the protein 3D structure and a table showing the edge betweenness values. On selecting an edge on the contact map, the corresponding edge in the contact map, the two residues in the JSmol applet and the edge betweenness value in the table get highlighted by red color as shown in Figure 13.

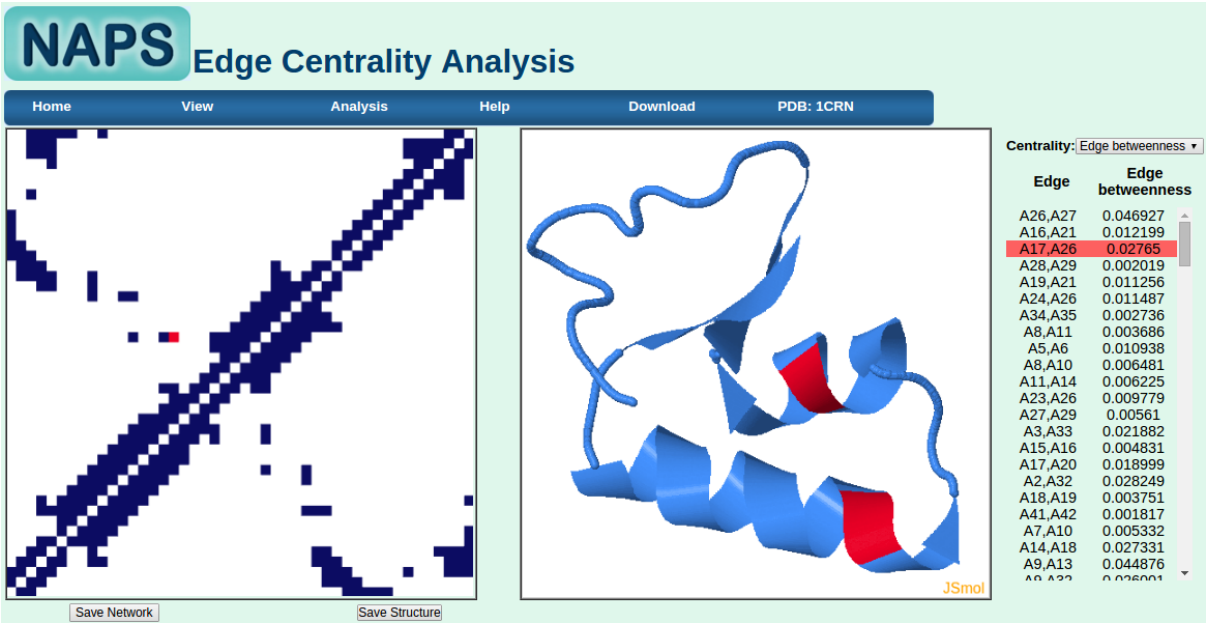

Figure 13: Edge centrality page showing edge betweenness.

8. Shortest path analysis

Shortest path distance between two nodes in a network is the minimum number of nodes that need to be traversed in order to reach from one node to the other. The shortest path analysis can be performed by selecting the 'Shortest path' option from the pull-down menu of 'Analysis' tab at the top menu bar. The user can select two residues from the drop down list, all shortest paths (if more than one) between the two nodes are identified and listed with radio buttons as shown in Figure 14. Selecting any path by clicking a radio button highlights the nodes and edges along that path in orange color in both the network view and JSmol applet (Figure 14).

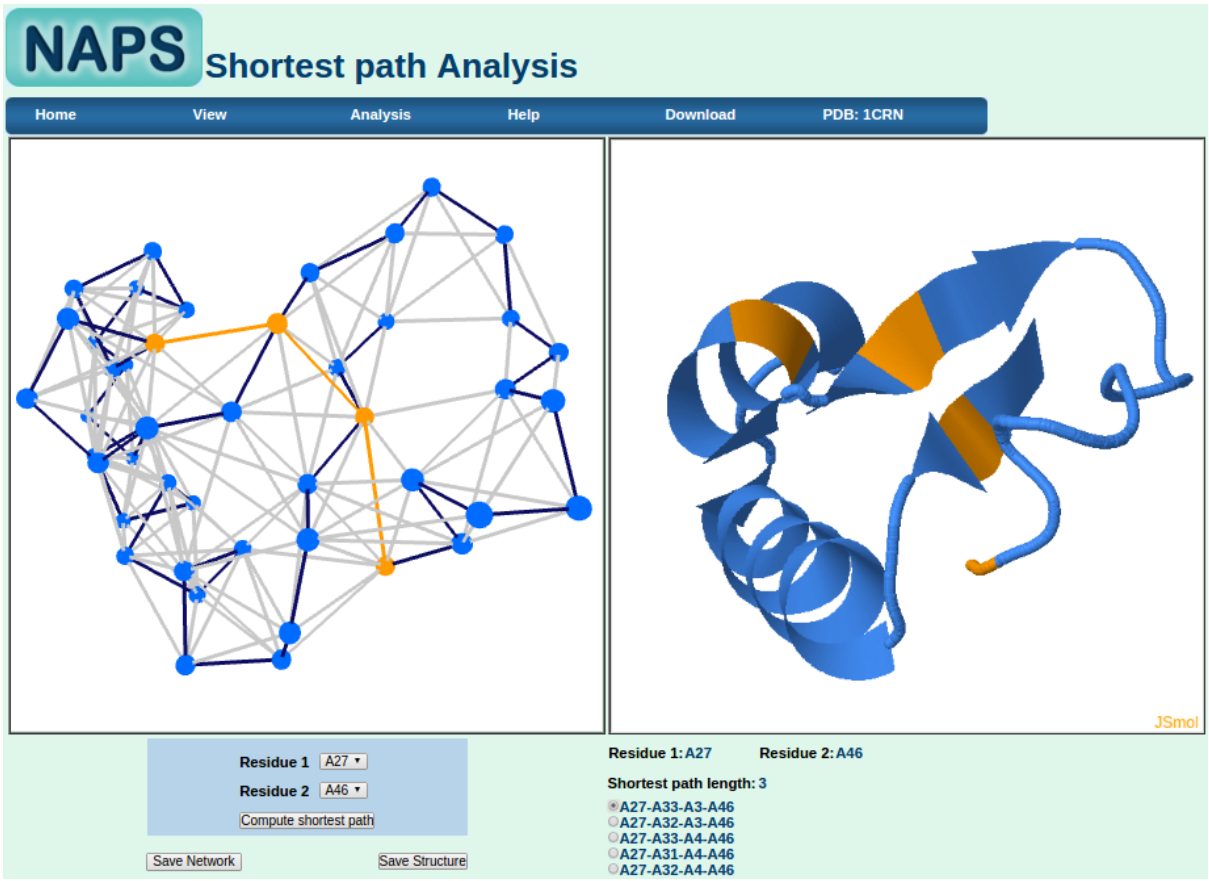

Figure 14: Shortest path analysis page highlighting one of the paths between residue A27 and A46.

From the download option on the menu bar of this page the user can download:

a) Shortest path between selected nodes in text format.

b) Network image in PNG format.

c) Structure image in PNG format.

d) Edgelist file.

9. *k*-clique analysis

A *k*-clique is a sub-network of *k* nodes with all the *k* nodes are connected to each other. These nodes may have edges to nodes outside the sub-network. The clique analysis can be performed by selecting the '*k*-clique' option from the pull-down menu of 'Analysis' tab at the top menu bar. On selecting *k*, from the drop down menu, all the cliques of size *k* are displayed as radio buttons. Selecting a radio button, the nodes and edges of that *k*-clique are highlighted in orange color in both the network view and JSmol applet as shown in Figure 15.

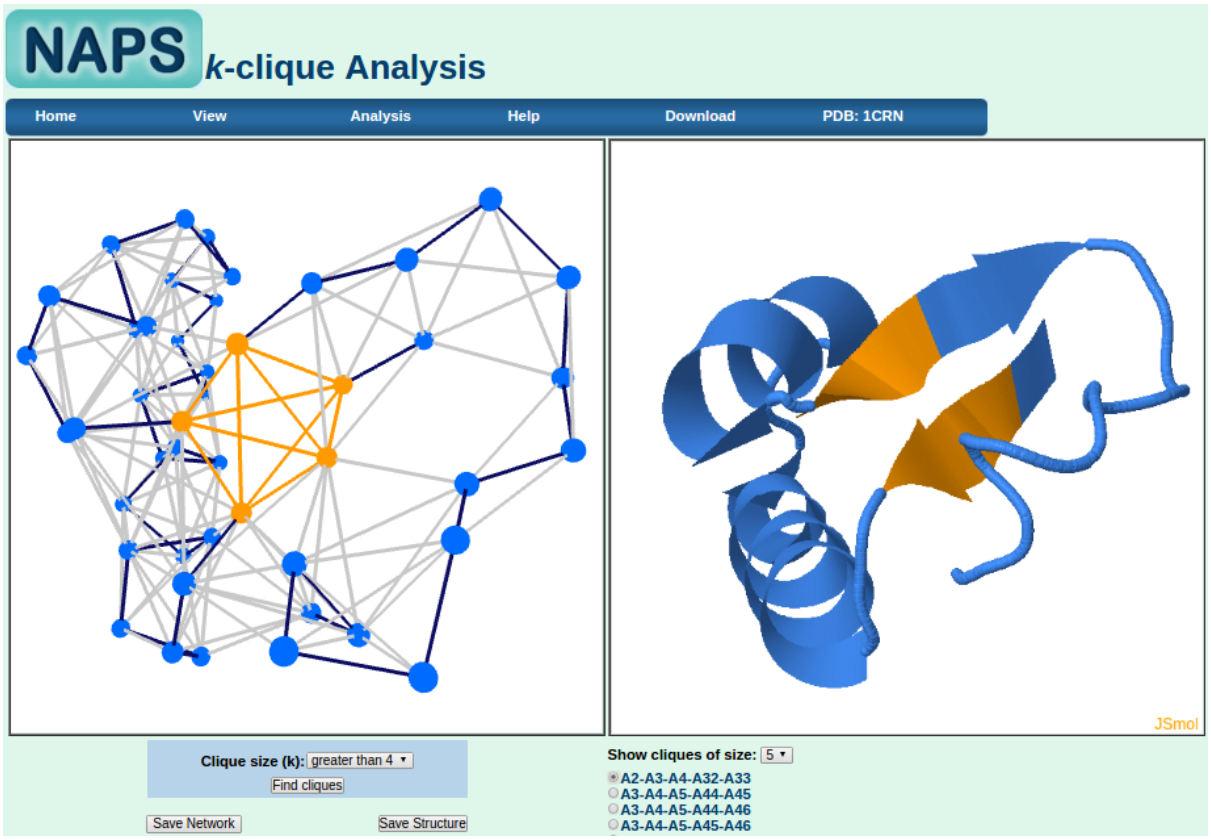

Figure 15: Clique of size 5 shown.

From the download option on the menu bar of this page the user can download:

a) All clique of selected size in text format.

b) Network image in PNG format.

c) Structure image in PNG format.

10. Graph spectral analysis

The graph spectral analysis provides spectral analysis of adjacency and laplacian matrices. It can be performed by selecting the 'Graph spectra' option from the pull-down menu of 'Analysis' tab at the top menu bar. The eigenvector component corresponding the largest eigen value of the adjacency matrix and the eigenvector corresponding second smallest eigen value of the laplacian matrix can be analyzed. All the view options discussed for node centrality analysis are available for graph spectral analysis. An example case with nodes in the network view and residues in JSmol colored according to the eigenvector corresponding second smallest eigen value of the laplacian matrix are shown in Figure 16.

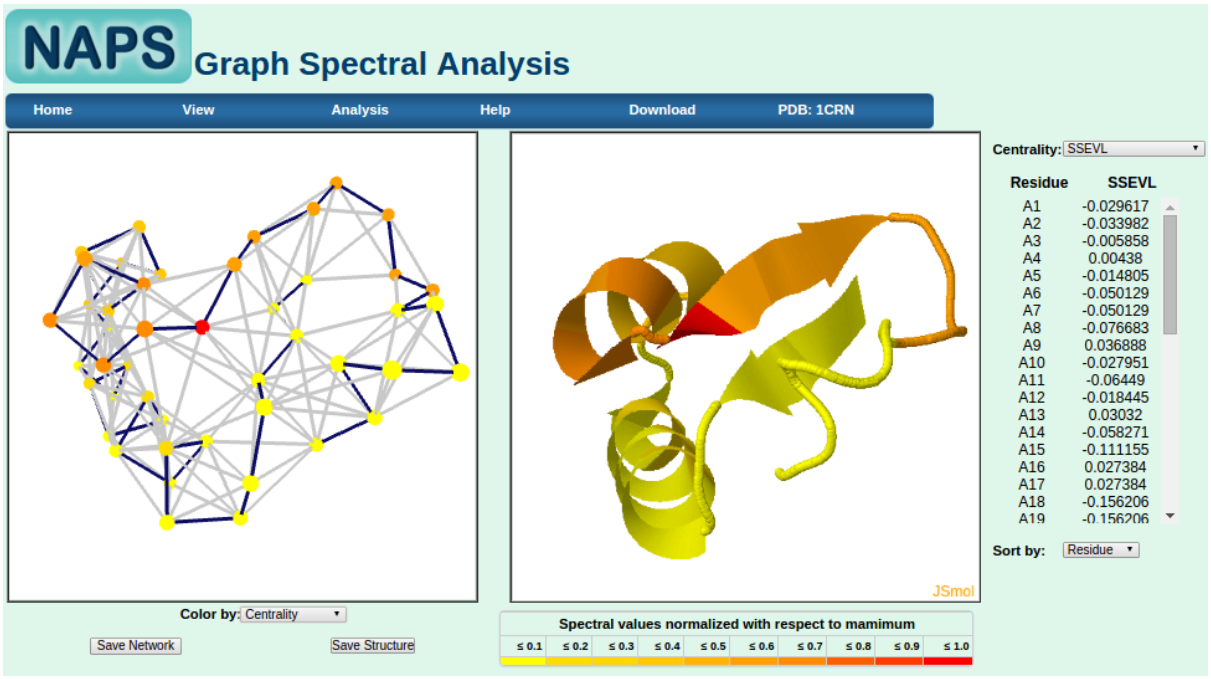

Figure 16: Nodes in network view and residues in JSmol applet colored according to the eigenvector corresponding second smallest eigen value of the laplacian matrix.

11. Multi domain analysis

In large multi-domain proteins, it is desirable to analyze the interaction patterns within and between the domains, which is provided in the 'Domain view' page of NAPS. Multi domain analysis can be performed by selecting the 'Domain view' option from the pull-down menu of 'Analysis' tab at the top menu bar. NAPS provides options to select up to 5 domains together. The user can provide the desired number of domains along with the coordinates, which will be used to color the nodes in the network 3D view and the residues in JSmol view. The analysis can be carried out for both contiguous and non-contiguous domains, which is one of the most useful feature required for visual analysis of multi-domain proteins. An example protein, 16PK with two domains is shown in Figure 17.

The coordinates of a contiguous domain can be given by first selecting the chain from the drop down menu and then providing the range. For example, input range '5:192' means residues 5 to 192. The coordinates of a non-contiguous domain can be given by providing the ranges separated by ','. For example, '1:10,25,30:35' means residues 1 to 10, residue 25 and residues 30 to 35.

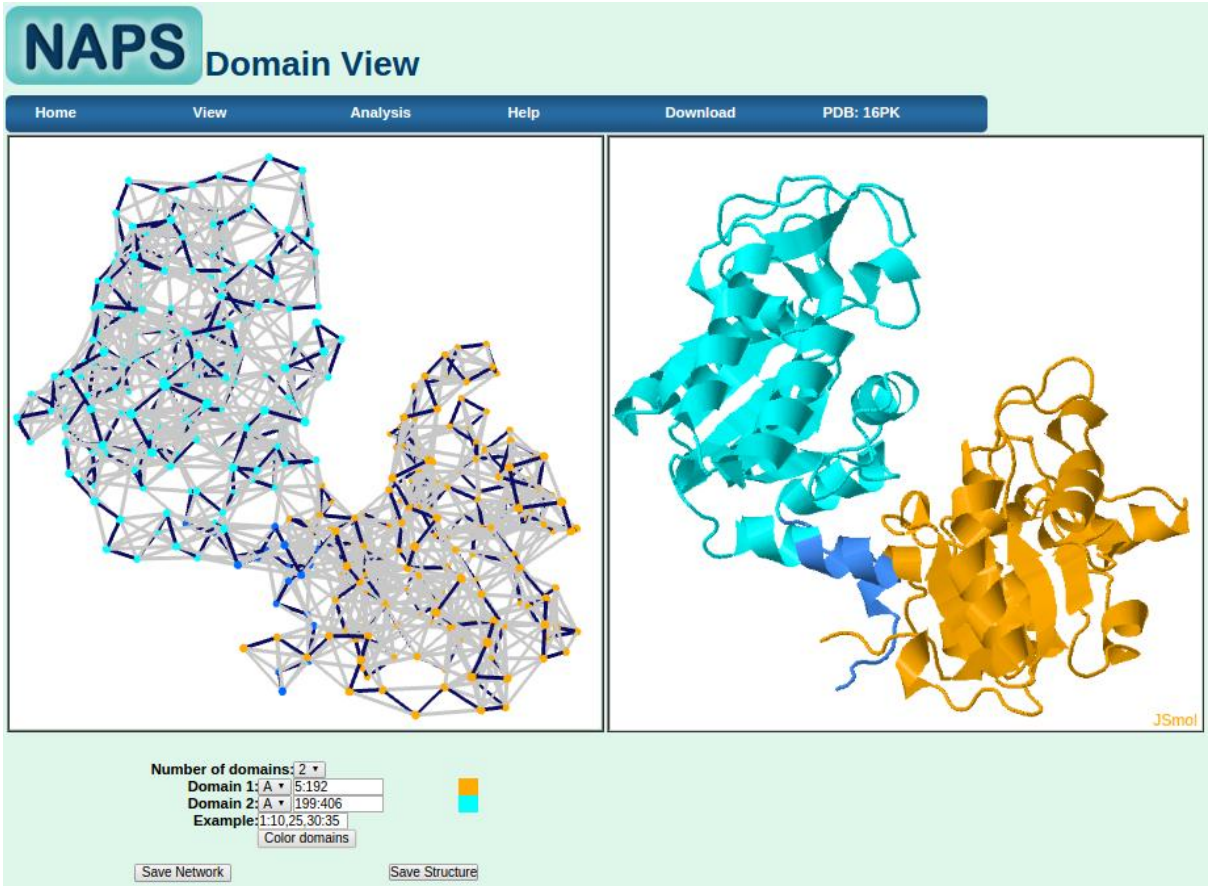

Figure 17: Domain view showing two domains of the protein 16PK.

The betweenness centrality analysis of the two domain protein is shown in Figure 18. It can be observed that the residues in the interface of the two domains have high betweenness values indicating their importance in the inter-domain communications.

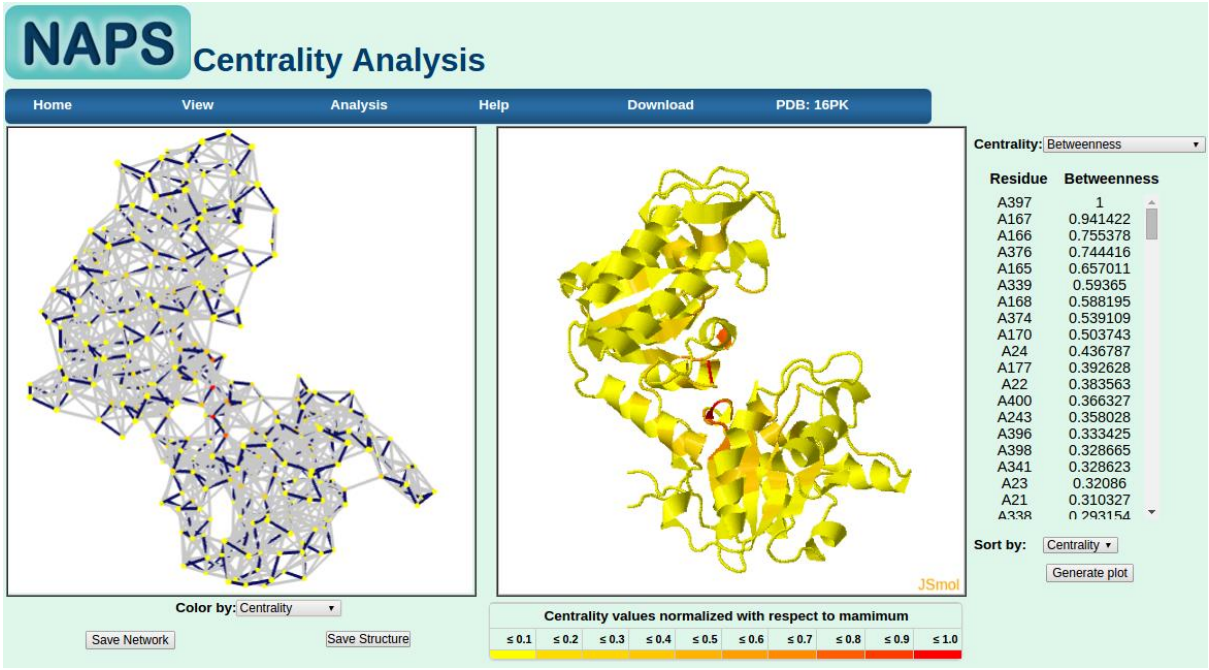

Figure 18: Betweenness centrality analysis of multi-domain protein, 16PK.

12. Analysis of weighted network

A weighted network can be constructed by selecting ‘weight’ option on the home page as ‘weighted’. The definition of edge weight for all 5 network types are described in Section 3 (Network construction). All the analysis for a weighted network can be performed similar to the unweighted network as discussed in the previous sections. In this section, we show the analyses pages which are different for a weighted network. The edge lines in 3D network view and the cells in contact map view are color coded according to the edge weights, as shown in Figure 19 and Figure 20.

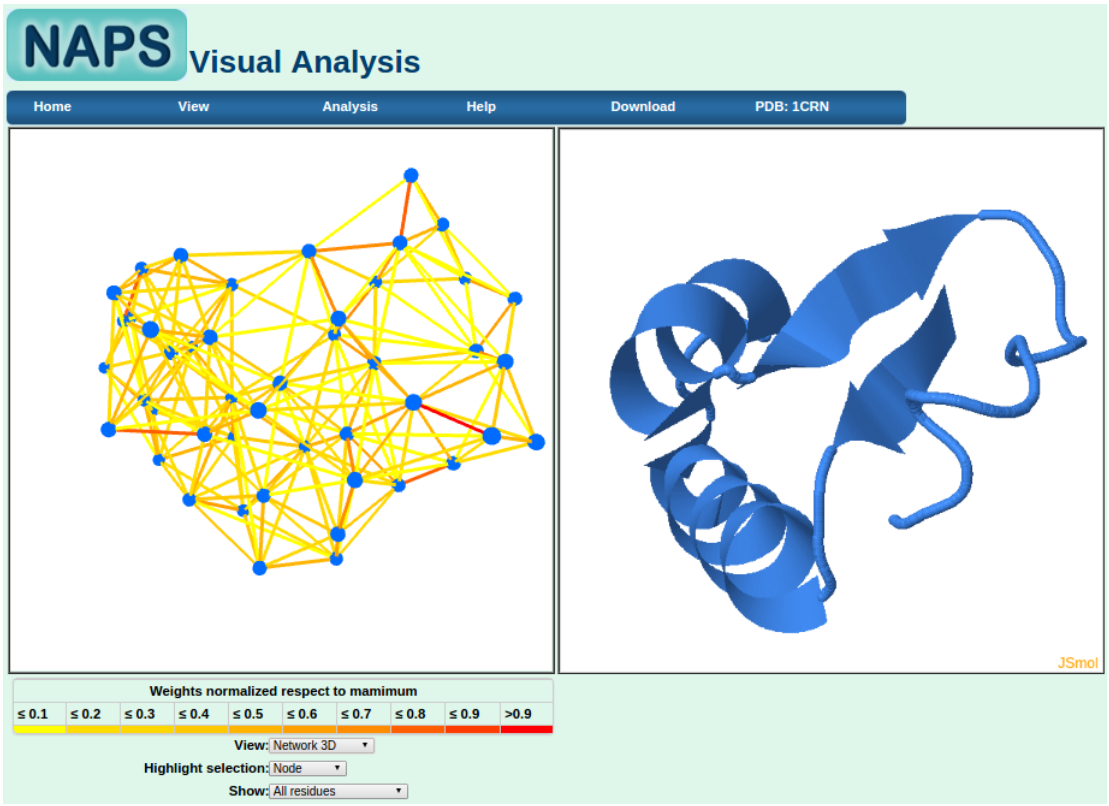

Figure 19: Network 3D view showing edge color based on weights.

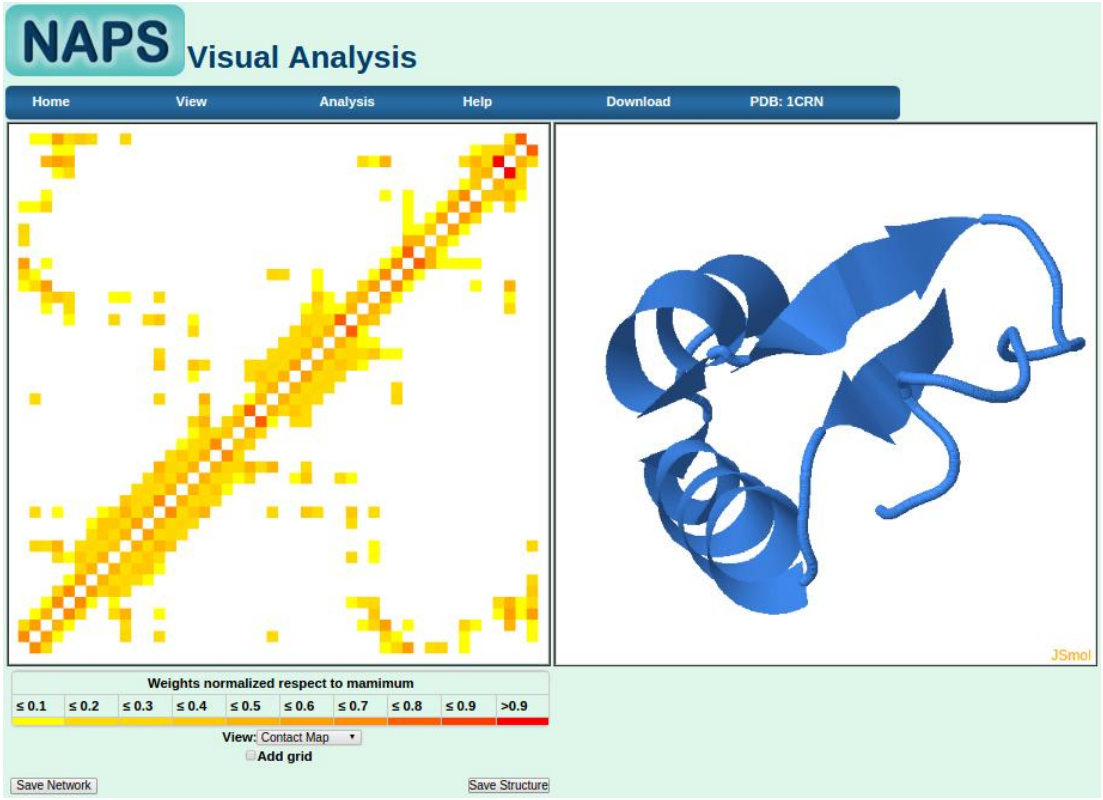

Figure 20: Contact map view with cell color showing the edge weights.

Shortest path analysis in weighted network

The inferences from shortest path analysis are based on the principle that the most important path (route) between a pair of nodes is the one with the minimum distance, i.e. the edges with lower distances (weights) are more important. In a weighted Protein Contact Network (Section 3), the weights are defined in such a way that the edges with more weights are more important. In order to make meaningful inferences from shortest path analysis of weighted PCNs, the distance between a pair of nodes i and j is taken as $(1/w_{ij})$ for all shortest path analyses: average shortest path, closeness centrality, betweenness centrality and shortest path analysis between a pair of nodes.

13. Analysis of Protein complex

All the analysis that can be performed on one protein can be performed on a protein complex with two chains representing either a dimer of the same protein, or a complex of two interacting proteins. The 3D network view shows the nodes corresponding the two protein chains in blue and green colors with backbone edges in the corresponding colors. Inter-chain edges, i.e. the edges between nodes of different chains are represented in magenta color while all intra-chain edges are colored grey. The snapshots of the tool for different analysis pages are shown in following figures.

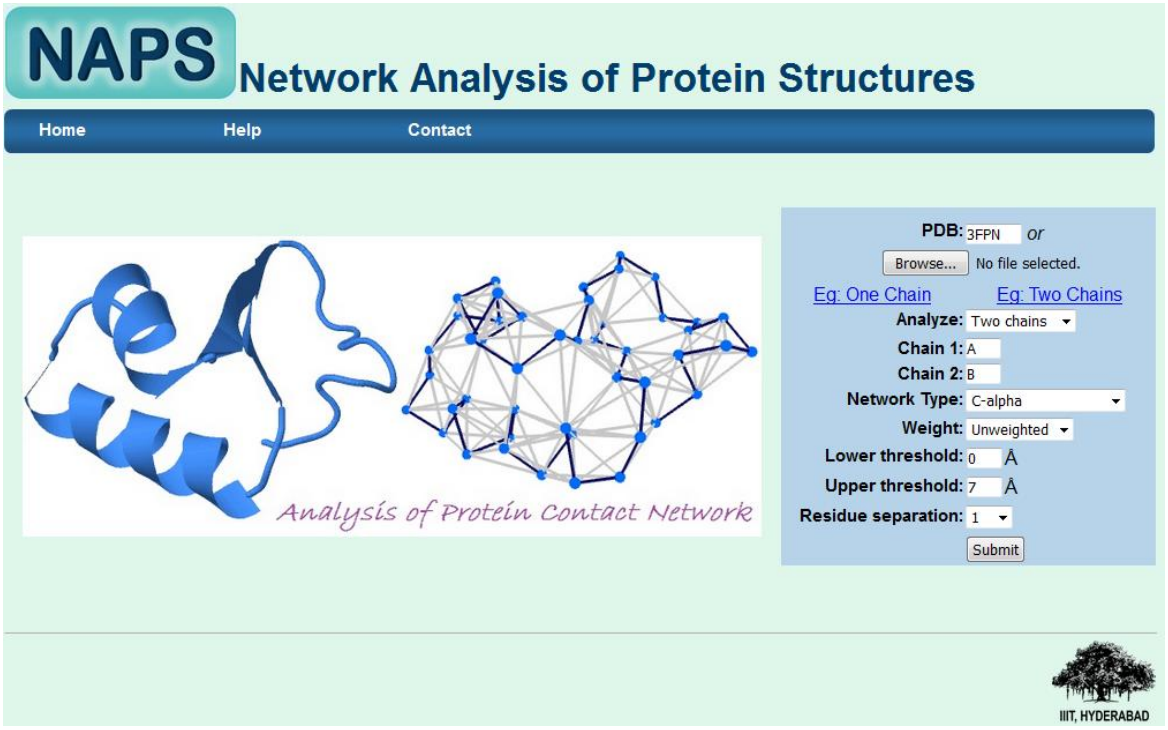

Figure 21: Input form for protein complex at the home page.

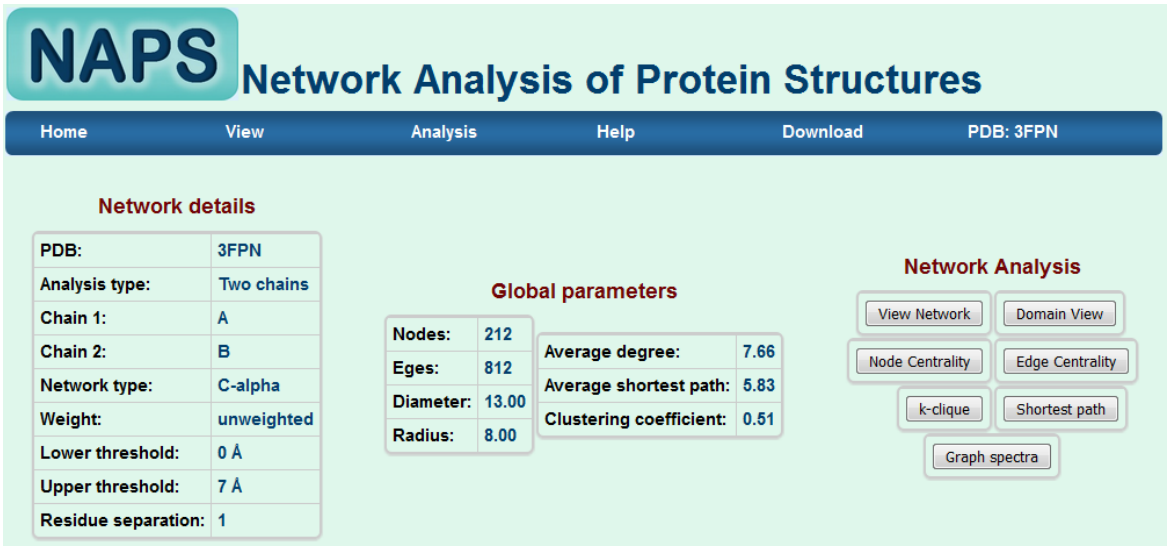

Figure 22: Network properties for protein complex.

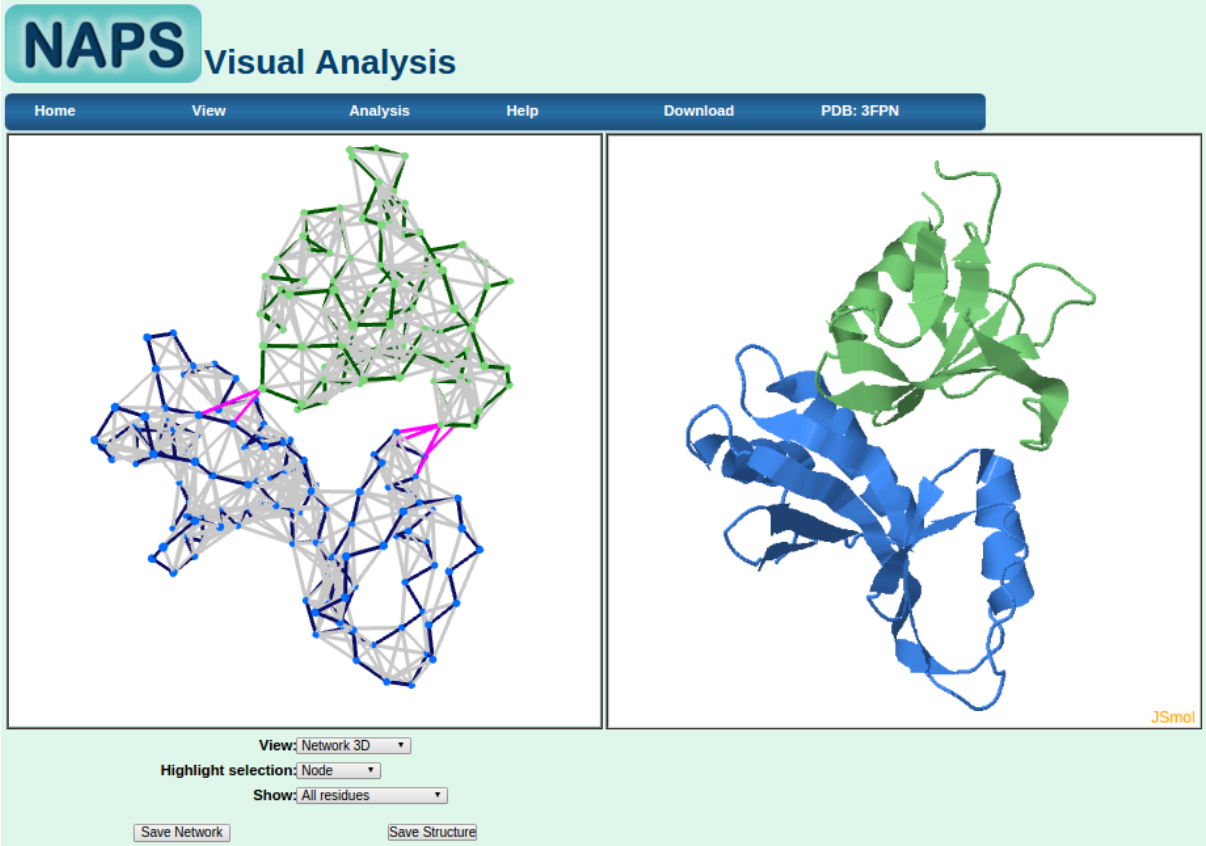

Figure 23: Network 3D view for protein complex. The two chains are shown in blue and green color in JSmol applet. The corresponding nodes in the network view are also shown in respective colors. Intra-chain edges are represented by grey color while inter-chain edges are represented by magenta color.

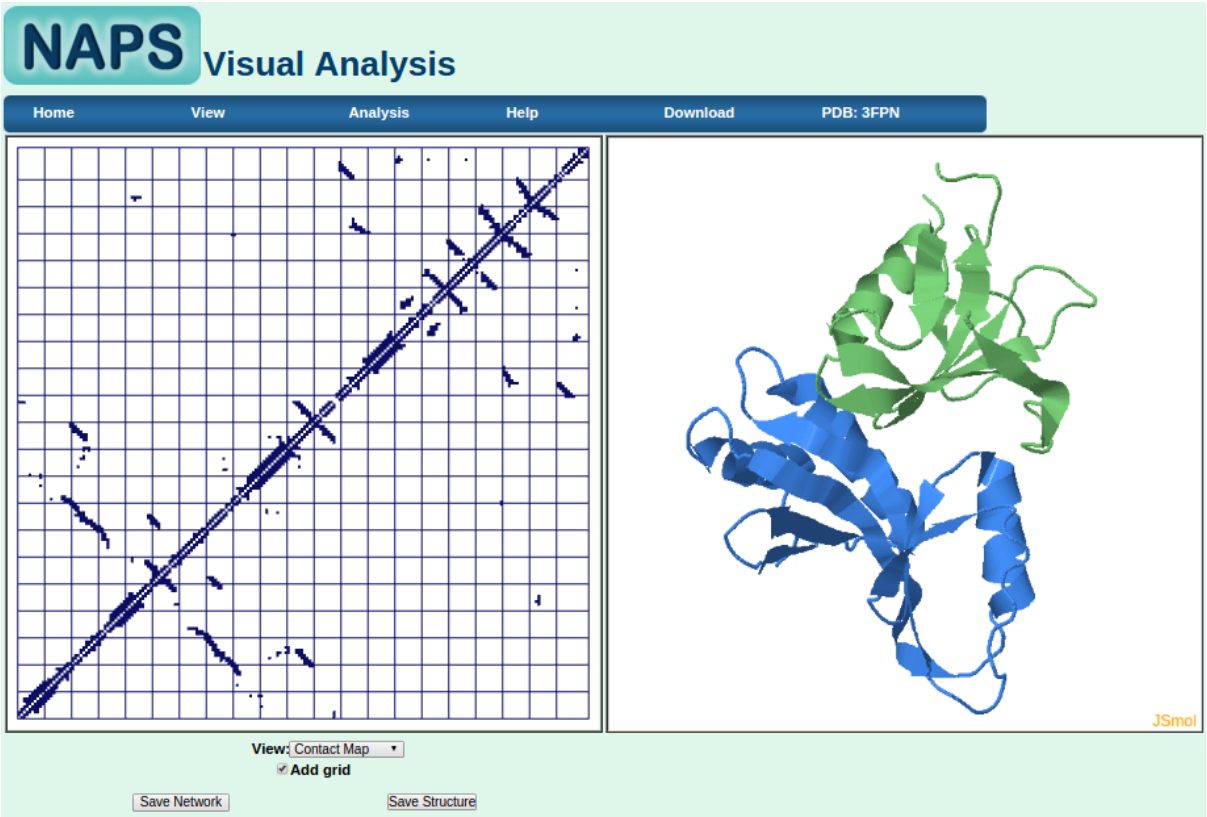

Figure 24: Contact map view of the protein complex.

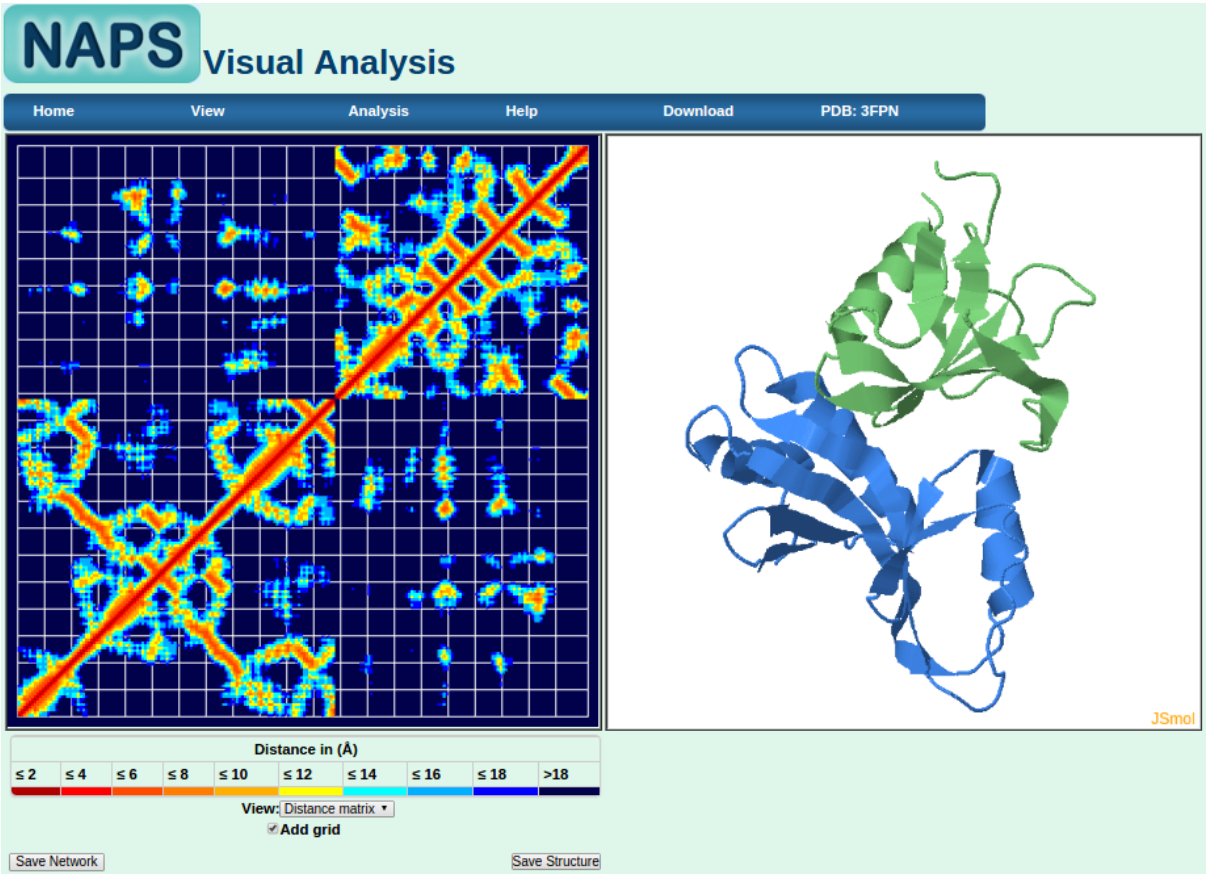

Figure 25: Distance matrix of the protein complex.

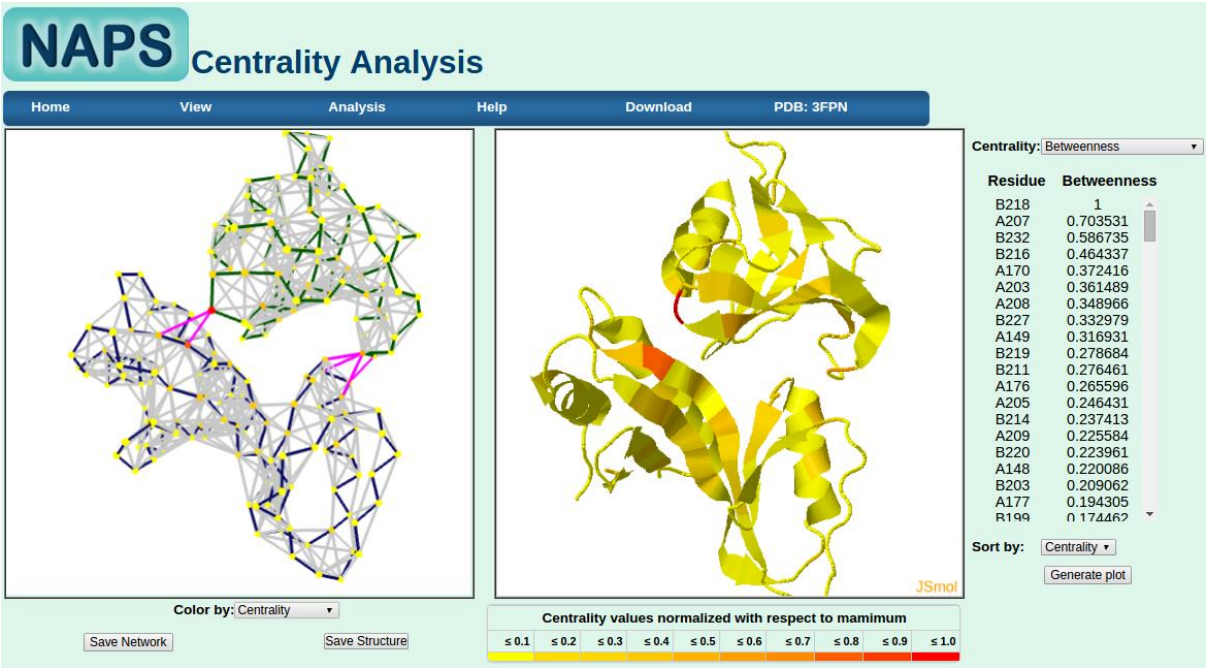

Figure 26: Betweenness centrality analysis of the protein complex. The interface residues with magenta edges showing protein-protein interaction can be observed to have high betweenness values indicating their importance in the inter-protein communications.

14. References

1. Kannan,N. and Vishveshwara,S. (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.*, **292**, 441–464.

2. Brinda,K.V. and Vishveshwara,S. (2005) A network representation of protein structures: implications for protein stability. *Biophys. J.*, **89**, 4159–4170.

3. Bagler,G. and Sinha,S. (2005) Network properties of protein structures. *Physica A: Statistical Mechanics and its Applications*, **346**, 27–33.

4. Bartoli,L., Fariselli,P. and Casadio,R. (2007) The effect of backbone on the small-world properties of protein contact maps. *Phys Biol*, **4**, L1–5.

5. Bagler,G. and Sinha,S. (2007) Assortative mixing in Protein Contact Networks and protein folding kinetics. *Bioinformatics*, **23**, 1760–1767.

6. Vishveshwara,S., Ghosh,A. and Hansia,P. (2009) Intra and inter-molecular communications through protein structure network. *Curr. Protein Pept. Sci.*, **10**, 146–160.

7. Patra,S.M. and Vishveshwara,S. (2000) Backbone cluster identification in proteins by a graph theoretical method. *Biophys. Chem.*, **84**, 13–25.

8. Chakrabarty,B. and Parekh,N. (2012) Analysis of graph centrality measures for identifying Ankyrin repeats. In *2012 World Congress on Information and Communication Technologies (WICT)*.pp. 156–161.

9. Chakrabarty,B. and Parekh,N. (2014) Identifying tandem Ankyrin repeats in protein structures. *BMC Bioinformatics*, **15**, 6599.

10. Chakrabarty,B. and Parekh,N. (2014) PRIGSA: protein repeat identification by graph spectral analysis. *J Bioinform Comput Biol*, **12**, 1442009.

11. Atilgan,A.R., Akan,P. and Baysal,C. (2004) Small-world communication of residues and significance for protein dynamics. *Biophys. J.*, **86**, 85–91.

12. Greene,L.H. and Higman,V.A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.*, **334**, 781–791.

13. Aftabuddin,M. and Kundu,S. (2007) Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein. *Biophys J*, **93**, 225–231.

14. Alves,N.A. and Martinez,A.S. (2007) Inferring topological features of proteins from amino acid residue networks. *Physica A: Statistical Mechanics and its Applications*, **375**, 336–344.

15. JSmol: an open-source HTML5 viewer for chemical structures in 3D.

16. Yan,W., Zhou,J., Sun,M., Chen,J., Hu,G. and Shen,B. (2014) The construction of an amino acid network for understanding protein structure and function. *Amino Acids*, **46**, 1419–1439.

17. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.