

ProSNEx USER MANUAL

About ProSNEx

We have developed a web-based application called ProSNEx (Protein Structure Network Explorer) which provides a novel and enhanced analysis of protein structures using network formalism. It provides a wide range of data and visualization options.

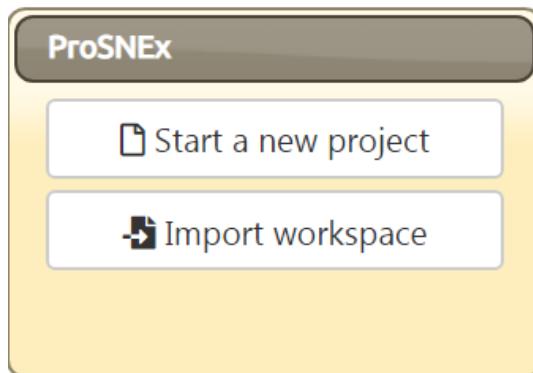
System Requirements

There are no system requirements. There is no installation or login requirement.

Getting Started

Opening

The user is asked to choose from the following two options: Start a new project or Import a workspace that has been already studied previously. The first time users can continue with ‘Start a new project’.



Select PDB File

Network construction is based on a PDB file that is either provided with a PDB ID can alternatively uploaded by the user.

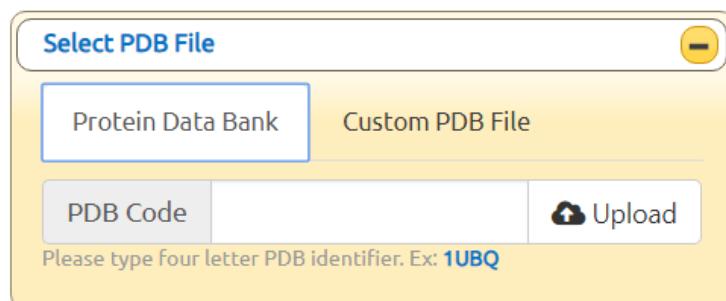


Figure 1. Select PDB File

Once a PDB is provided by the user, a chain and network selection window will appear. An example is provided below for PDB ID 1OF2. The user can type sub-portions of the structure if desired as well.

Once the chain selection and network selection process is completed, the user can click on ‘create’ to view the network.

Sub-portion selection: As to give an example, for a chain having 276 residues starting from 1 to 277, the user can choose 3 sub-portions within this chain by typing: 5,65;150,180;220,250 where the sub-portions are separated by semicolons.

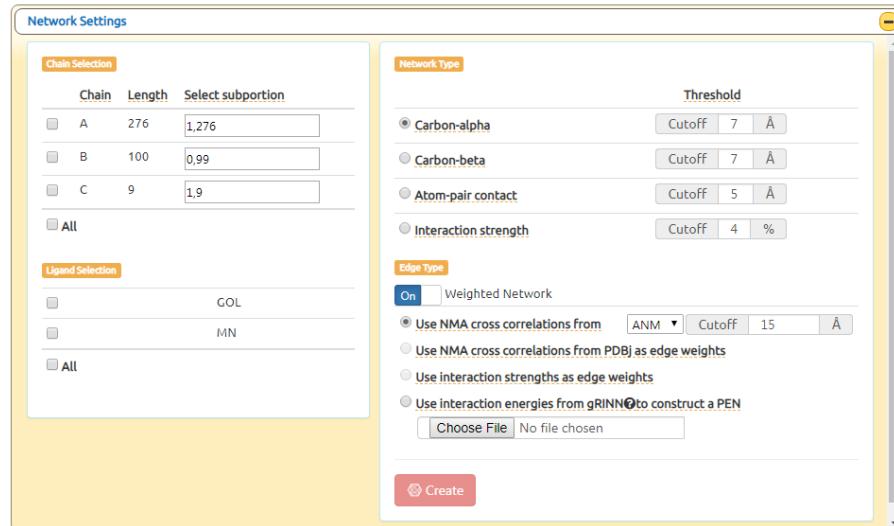


Figure 2. Chain and network selection window

Network Construction Settings

There are various network construction options provided for the users.

Network Type

- **Carbon-alpha**

C-alpha atoms of each amino acid are assigned as nodes. Edges are constructed between a pair of C-alpha atoms based on the threshold chosen by the user, where the default value is 7Å.

- **Carbon-beta**

C-beta atoms of each amino acid are assigned as nodes. Edges are constructed between a pair of C-beta atoms based on the threshold chosen by the user, where the default value is 7Å.

- **Atom-pair contact**

This network representation provides analysis of protein structure at the atomic level by capturing interactions between any pair of atoms based on the users' threshold choice. The default value of the threshold is 5Å.

- **Interaction Strength**

Edges between nodes are added only if the interaction strength is more than the user-specified threshold, where the default value is 4%. The interaction strength calculations are based on Brinda et al [1]*.

NOTE! The user can choose the threshold value, where the default values are specified as 7 Å for C-alpha and C-beta and 5 Å for atom-pair contact networks. In the case of a low threshold value, fewer interactions will be present within that specified range leading to less edge formation. In the case of a higher threshold value, more interactions will be present resulting in a crowded network with many edges.

Weight types

- ***Unweighted Network***
In this option, all edges are considered equally important.
- ***Edge weights from NMA cross correlations from ANM/GNM***
Weights are extracted from Dynamical Cross-Correlation Maps (DCCMs) calculation of GNM/ANM[2-4] simulations. The user can specify the threshold. Both these methods are Carbon-alpha based methods.
- ***Use NMA cross correlations from PDBj as edge weights***
This option provides the usage of Normal Mode Analysis (NMA) information found in PDBj [5] website for the construction of a network. For weight, normalized and time-averaged cross-correlations, as reported at the Promode-elastic database of PDB Japan, are used. This option is not available for custom PDB files.
- ***Use interaction strengths as edge weights***
In this option, the interaction strength values are taken from Kannan et al [6].
- ***Edge weights from gRINN [7]***
The tool provides the option of using Interaction Energy Matrices (IEM) given by gRINN [7] to construct a network. Here, average interaction energies in the file format returned by gRINN is accepted.

A demonstration is provided below regarding the choice of cut-off values and their effects on the network metrics.

Choosing cut-off values: Here's a short demonstration on a small protein (1UBQ)

As to demonstrate the differences in the betweenness centrality and closeness centrality parameters, we have constructed a Carbon-alpha network using three types of edge weights: Unweighted, ANM with a default cut-off of 15 Å and GNM with a default cut-off of 7 Å.

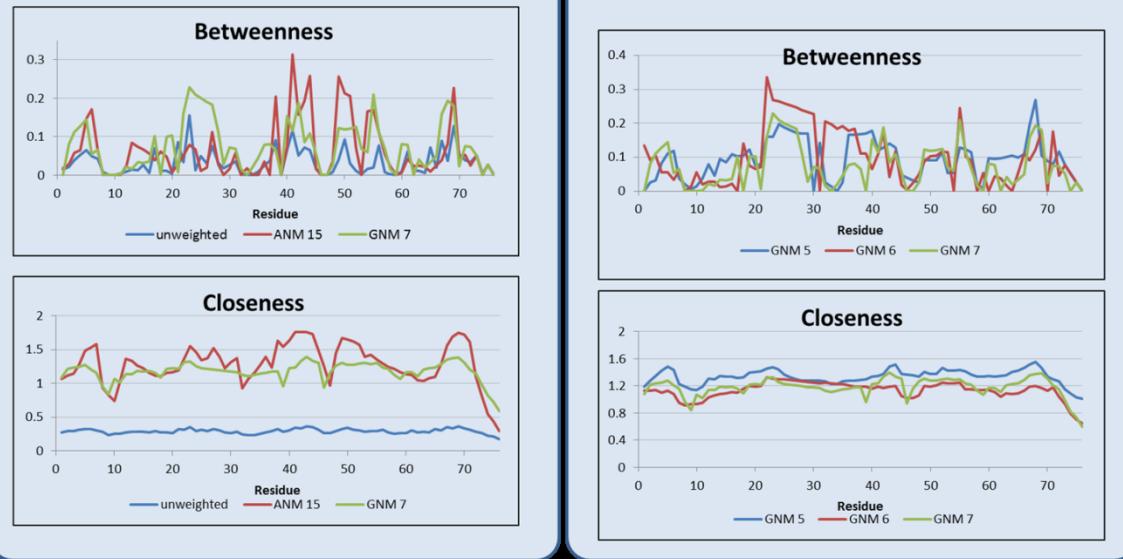


Figure 3. Choosing cut-off values and their effects on the network metrics

Main Screen: 3D Structure of the Molecule, Residue Interaction Network and Analysis Tools

Once the network creation method is selected, 3 main window panels will appear on the screen.

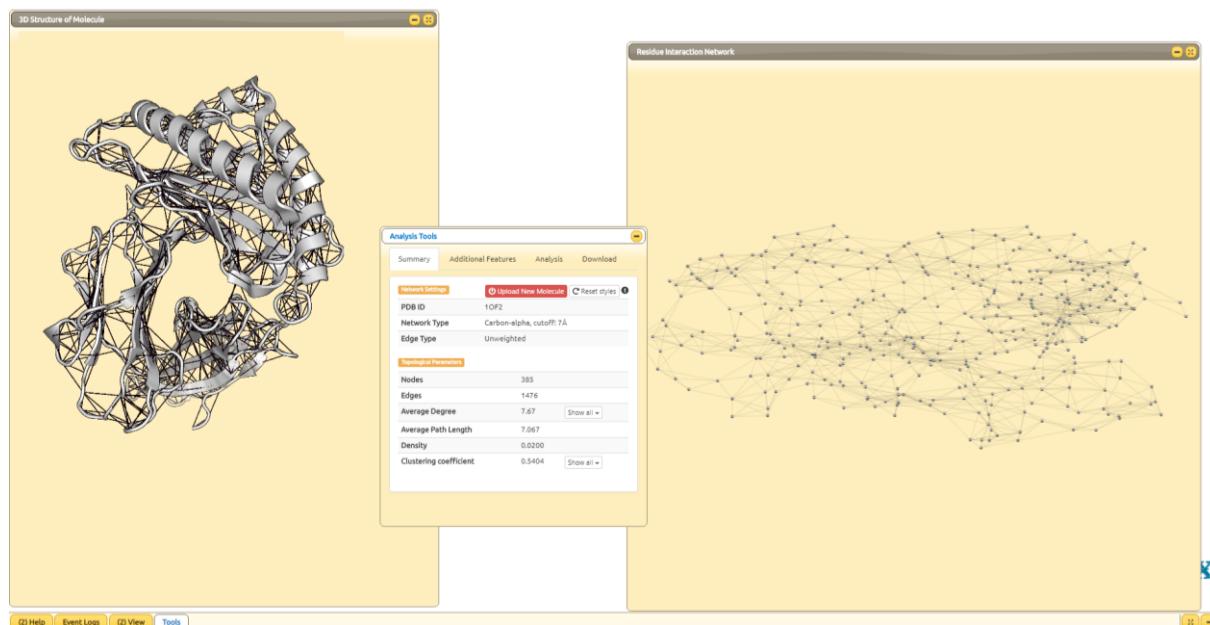


Figure 4. Three main window panels of ProSNEEx

3D Structure of the Molecule: In this panel, 3D structure of the protein can be visualized. On hovering the mouse over a residue, the residue is colored in red and is also highlighted on the network.

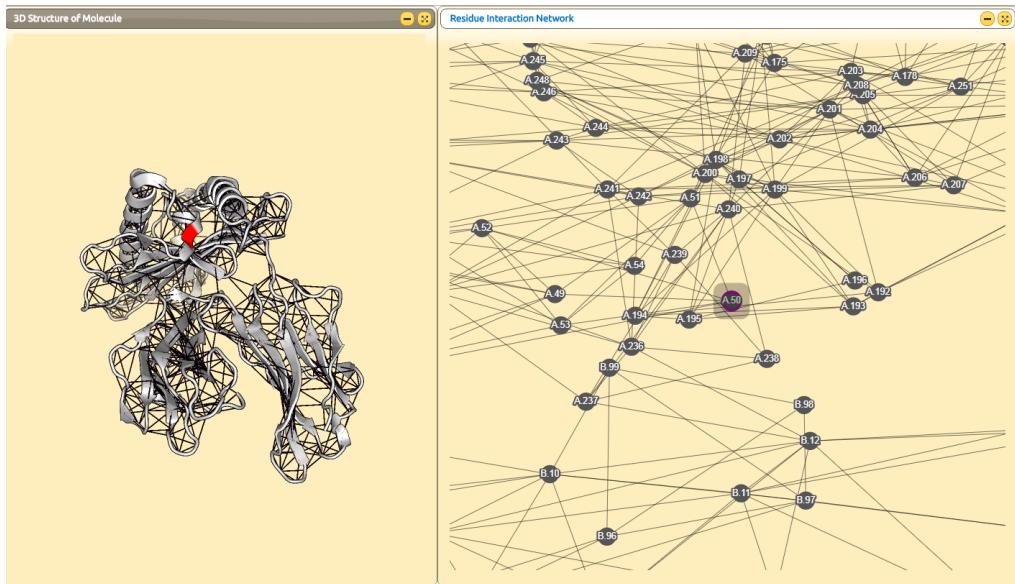


Figure 5. 3D structure of the molecule

Residue Interaction Network: In this panel, the constructed network can be visualized. When a node is selected by clicking, the corresponding residue also gets highlighted on the 3D structure of the protein.

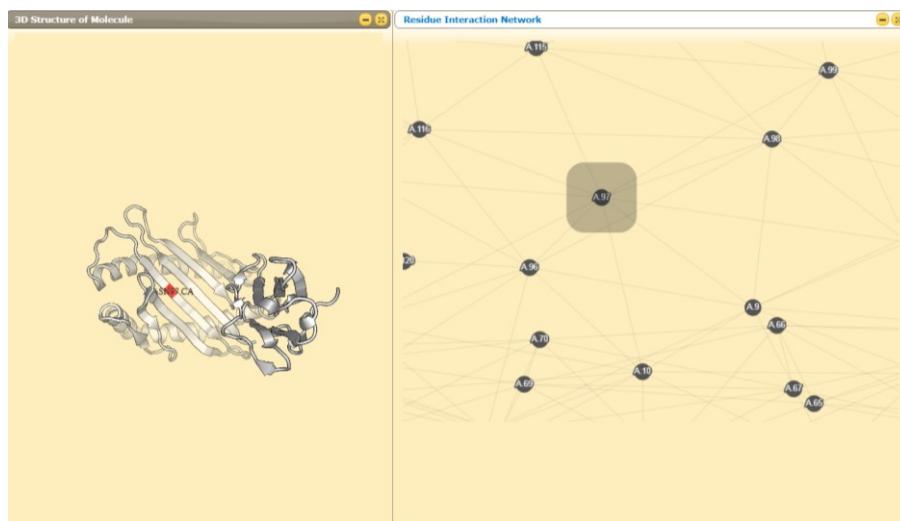


Figure 6. Residue Interaction network

Analysis Tools: This is the main window where you can reach all the network analysis results and visualization options of these results. A detailed explanation of this window and contents are provided below.

There are 4 tabs at the bottom of the main window.

Help: This is the section where you can reach the citation of the tool. It also contains a HELP section for the users.

Event logs: you can follow the log from this panel.

View panel: this panel will aid the user to navigate between the 3D structure of the molecule and the Residue Interaction Network windows.

Tools: this panel contains the analysis tools panel.

Analysis Tools Panel

“Analysis Tools Panel” is mainly composed of 4 tabs: Summary, Additional Features and Analysis and Download.

Summary section provides a summary of the input and the user selections as well as details of the topological parameters.

Network Setting

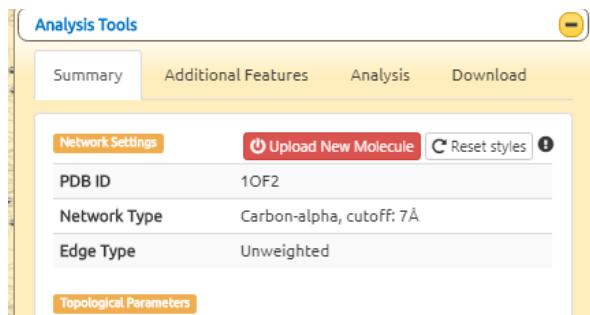


Figure 7. Network Settings

This section displays the selections of the user including the PDB ID (when available), network and edge types.

- Reset styles button can be used to reset all the changes and to display the default view.

Topological Parameters

Topological Parameters		
Nodes	385	
Edges	1476	
Average Degree	7.67	Show all ▾
Average Path Length	0.664	
Density	0.0200	
Clustering coefficient	0.5404	Show all ▾

Figure 8. Topological parameters

Nodes: number of nodes in the constructed network, which is equivalent to the number of residues in the structure.

Edges: number of edges in the network.

Average Degree: It is the average degree of all the nodes in the network. Individual degrees of all residues are also available. The values can be chosen to be sorted by residue or value, where txt. download option of the values is also available. Alternatively the hub residues can be shown on the structure or the values can be plotted to be examined in more detail. A hub residue is defined as a residue that has a degree centrality value higher than the average degree.

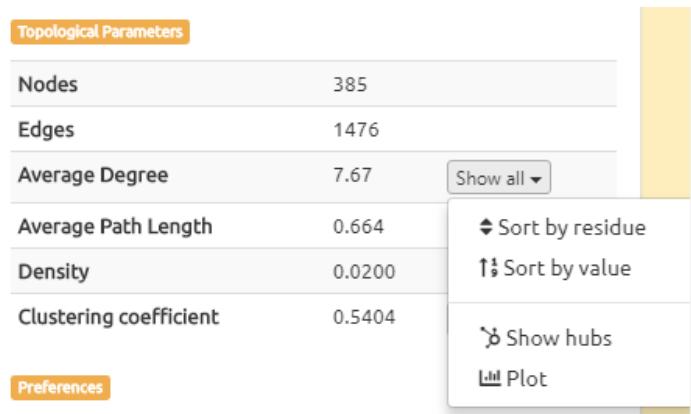


Figure 9. Average degree tab and the drop-down menu of this property

Average path length: It is the average of the shortest paths between all node pairs in the network. It is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of efficiency of information or mass transport on a network.

Density: Density of the network is defined as a ratio of the number of edges (E) to the number of possible edges in a network within N nodes.

Clustering coefficient: Clustering coefficient of the network is a critical metric for describing the characteristic structure of a protein. Individual clustering coefficients of all residues are also available. The values can be chosen to be sorted by residue or value, where xlsx. download option of the values is also available. The values can also be plotted to be examined in more detail.

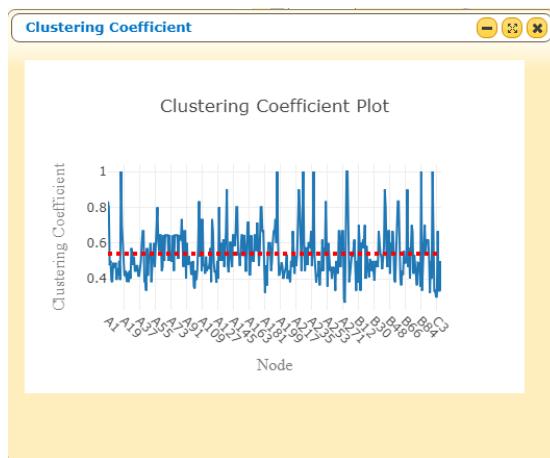


Figure 10 . Clustering coefficient plot. The value of each residue is shown and the average value is marked with a red line.

Additional Features section is composed of Preferences, Additional features, 3D structure coloring, Edge coloring, Network layout and Comparison sections.

Preferences

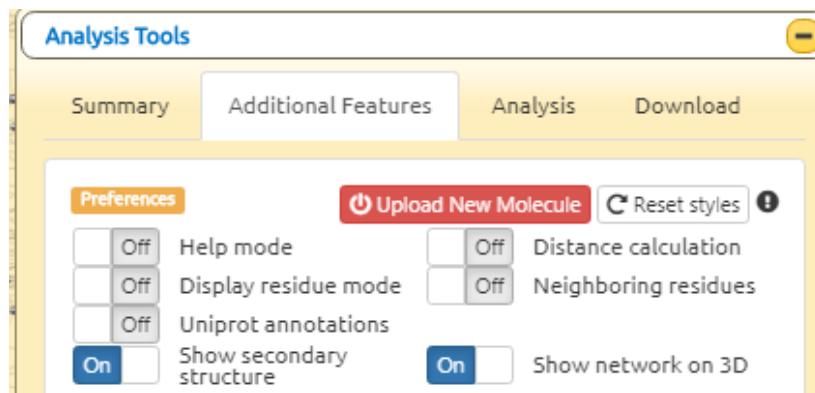


Figure 11. Preferences section

- **Help mode:** when this mode is ‘ON’, orange underlines appear for the parameters present in the tool. These parameters/terms are described upon mouse hovering.

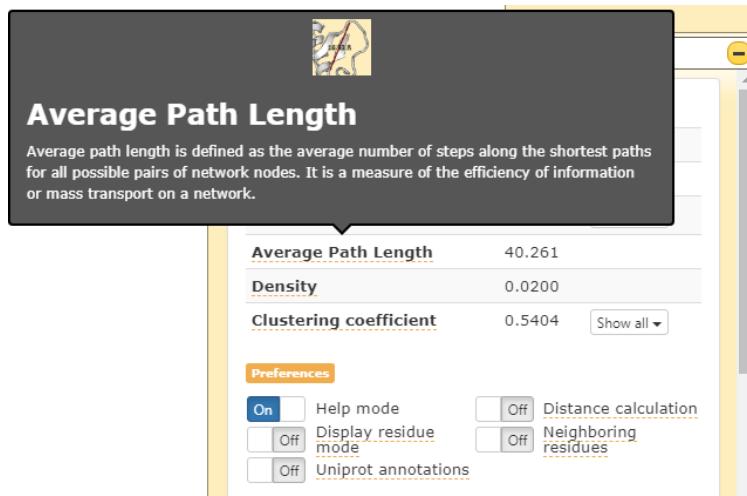


Figure 12. An example of the help mode

- **Display residue mode:** when this mode is ‘ON’, residue properties residue number and type are shown on the 3D structure panel and the chosen residue is also highlighted on the network.

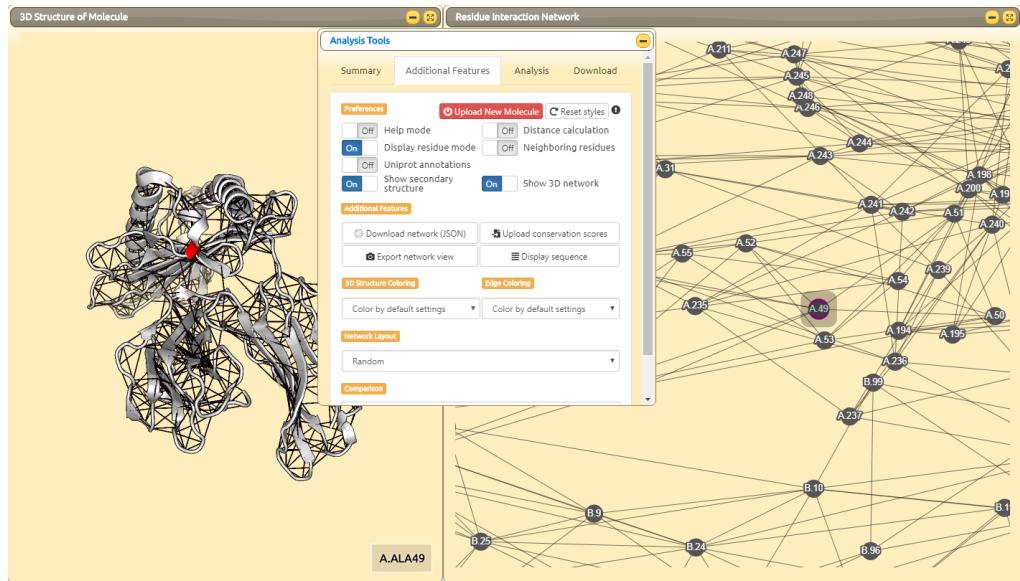


Figure 13. Display Residue mode

- **Distance calculations:** when this mode is ‘ON’, distance between any two selected residues is calculated and shown on the structure. The chosen residues are also highlighted in the network view.

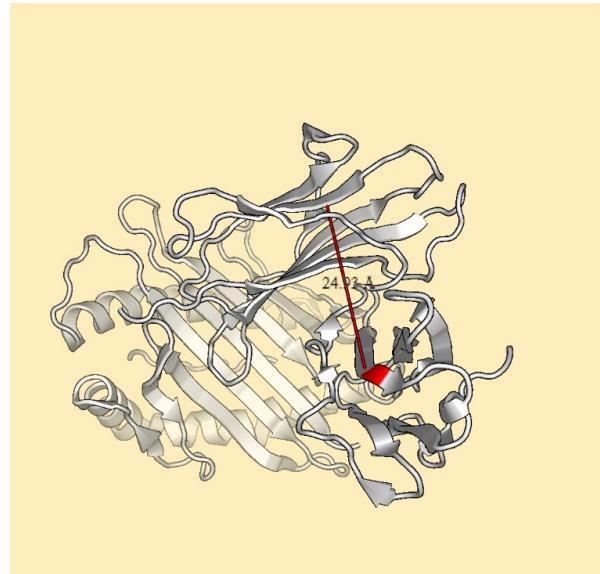


Figure 14. Distance calculation between residues B.ALAA 79 and A.ALAA245 for 1OF2.

- **Neighboring residues:** When this mode is ‘ON’, the neighboring residues of a selected node can easily be identified. The neighbors are highlighted in the network view and are also shown on the 3D structure. The node is shown in red and the neighboring residues are shown in purple.

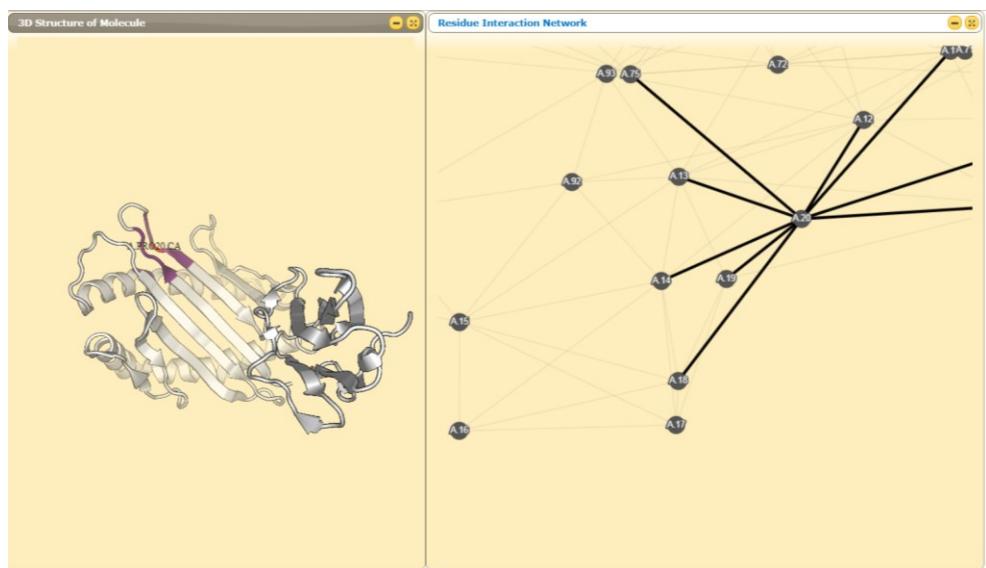


Figure 15. Display of the neighboring residues mode. Residue 20 is selected from the network panel, the corresponding neighbors are displayed on the 3D structure.

- **Uniprot annotations:** This unique feature displays the natural variant and disease mutations data taken from Uniprot for the given PDB. The annotations are shown on the 3D protein structure and the details are listed on the screen.

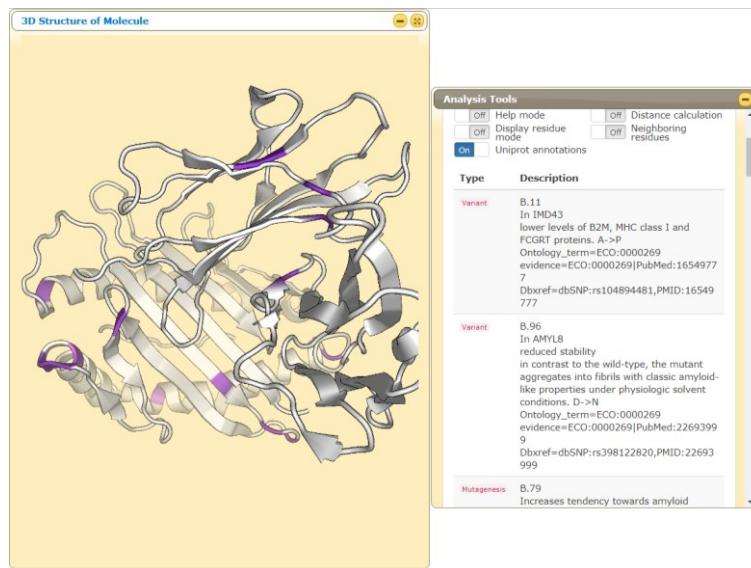


Figure 16 UNIPROT Annotations are marked on the 3D structure and are also listed.

Show secondary structure: this option is ‘on’ as default. The user can change it to hide the secondary structure.

Show 3D network: this option is ‘on’ as default. The user can change it to hide the network shown on the 3D structure.

Additional Features

Additional Features



Figure 17. Some additional features of the preferences section

- **Download network:** This button enables the users to download the network constructed by ProSNEx in a .json format.
- **Export network view:** This button enables the user to capture to constructed network that is present on the screen.
- **Upload conservation scores:** The user can upload the conservation scores for the PDB. For multi chain structures, corresponding values should be uploaded chain basis. The format should be the same as CONSURF-DB format. An example format is provided when the option is chosen.
- **Display sequence:** This button enables the users to display the chosen residue both on the 3D structure of the molecule and the residue interaction network panels.

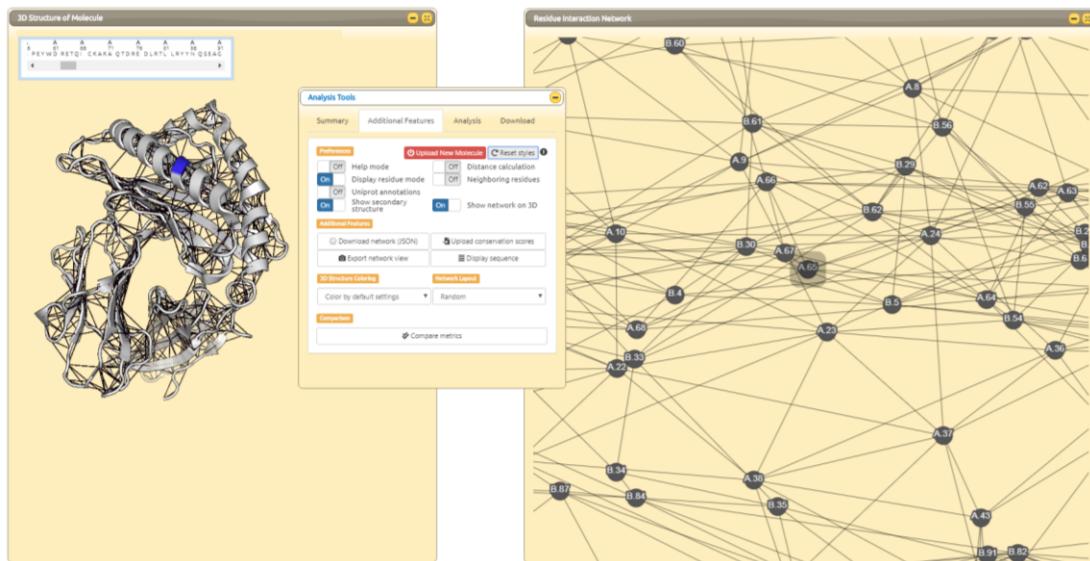


Figure 18. Display sequence mode

3D Structure Coloring

Structure coloring based on chain, secondary structure, b-factor (taken from PDB), conservation scores (taken from CONSURF DB [9]) and network metrics (degree, closeness centrality and betweenness centrality) is possible. Both the 3D structure of the molecule and the residue interaction networks will be colored based on the choice.

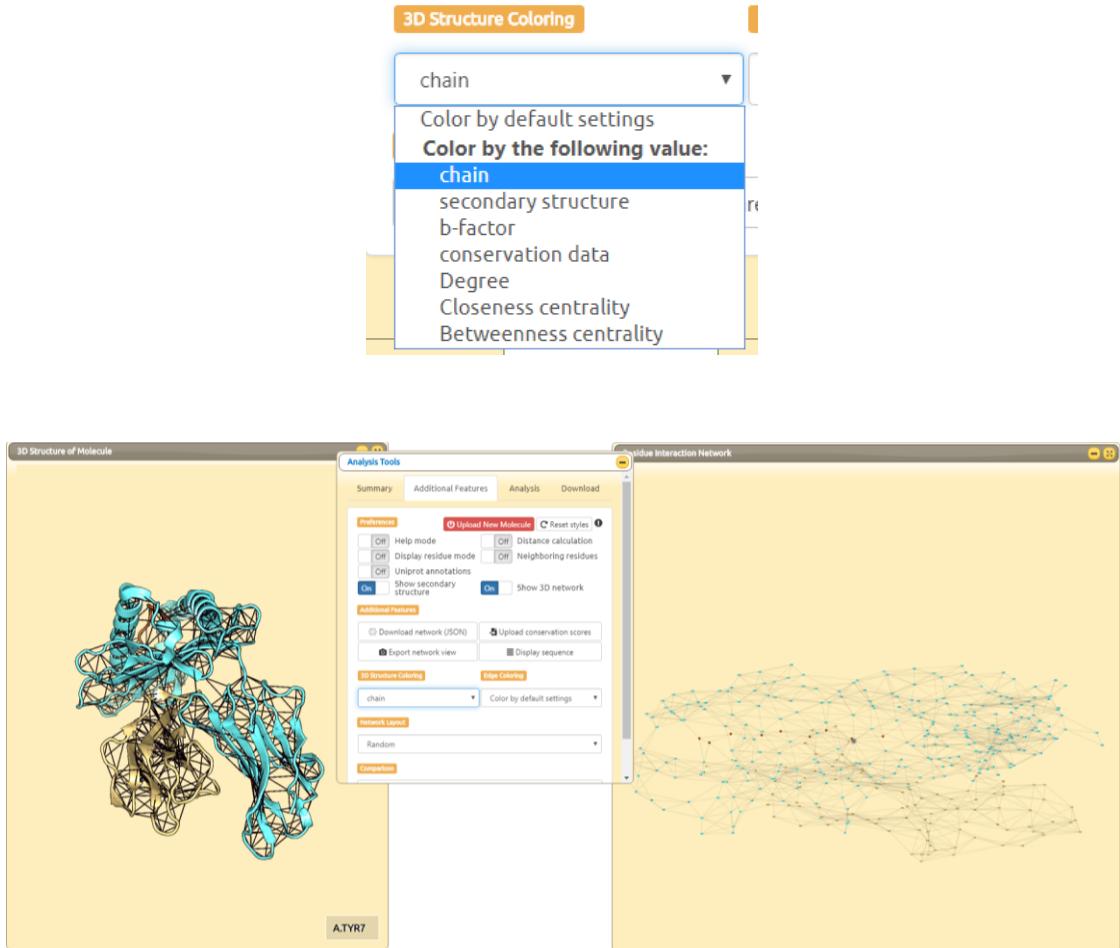


Figure 19. Coloring of both the 3D structure and the residue interaction network based on chains.

Edge Coloring

Different edge coloring options are present. The data is retrieved from Arpeggio Server [8].

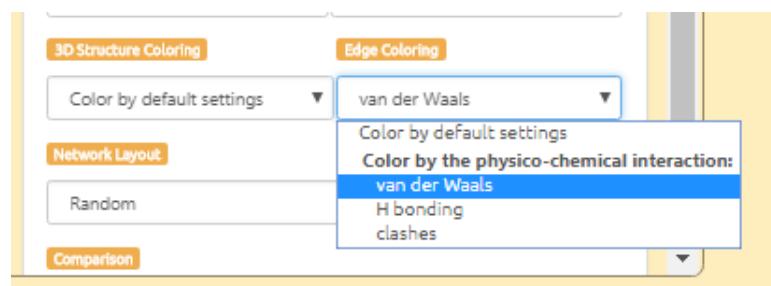


Fig 20. Display of edge coloring on the structure

Network Layout

Different network layout options are present.

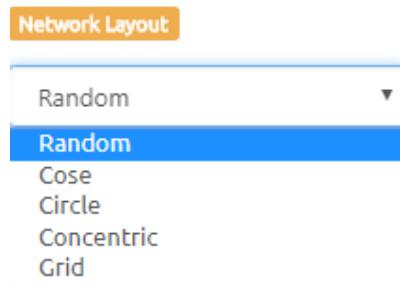


Figure 21. Network layout options

Comparison

Comparison section enables the users to compare and visualize certain metrics (b-factor, conservation data, degree, closeness centrality and betweenness centrality) on the structures. When 'compare metrics' is clicked, a new window appears. There are two drop-down menus available for the user. 3D structure of the protein is colored according to the choice of the user.

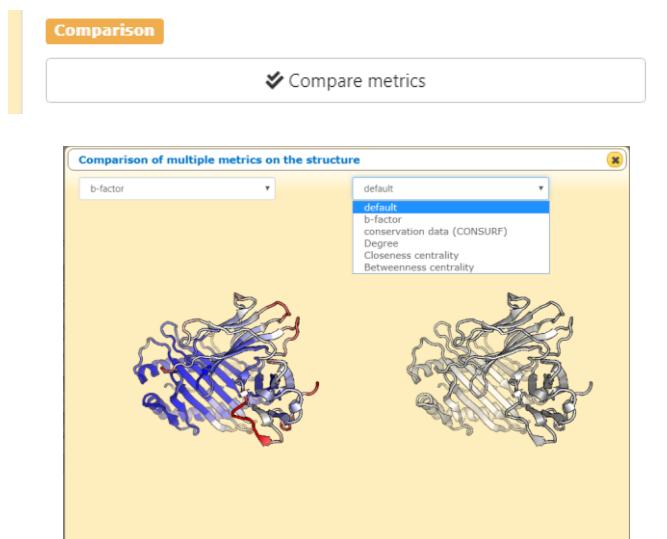


Figure 22. Comparison of multiple metrics on the same structure can be done using the 'compare metrics' button

Shortest Path Algorithm

Shortest path between two vertices is the path that has the minimum sum of edge weights. There exist two options: single pair and multiple pairs. The user can display the shortest path(s) by choosing the residues of interest from the drop-down menu for a single pair. Alternatively, the users who would like to find out the shortest paths for multiple pairs can compute it by typing multiple sources and targets in different lines and will get the results individually.

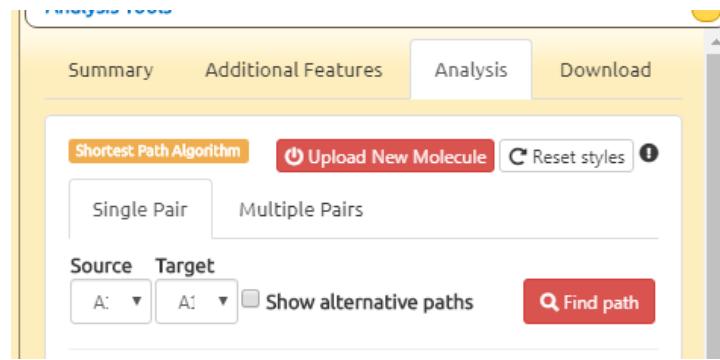


Figure 23. Shortest Path menu

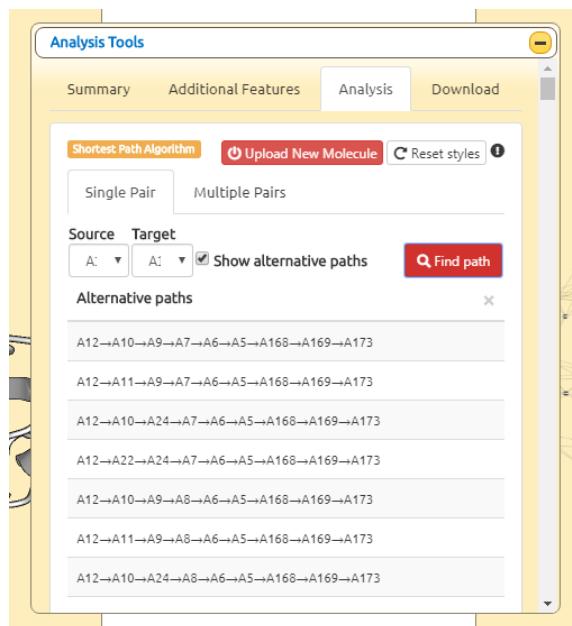
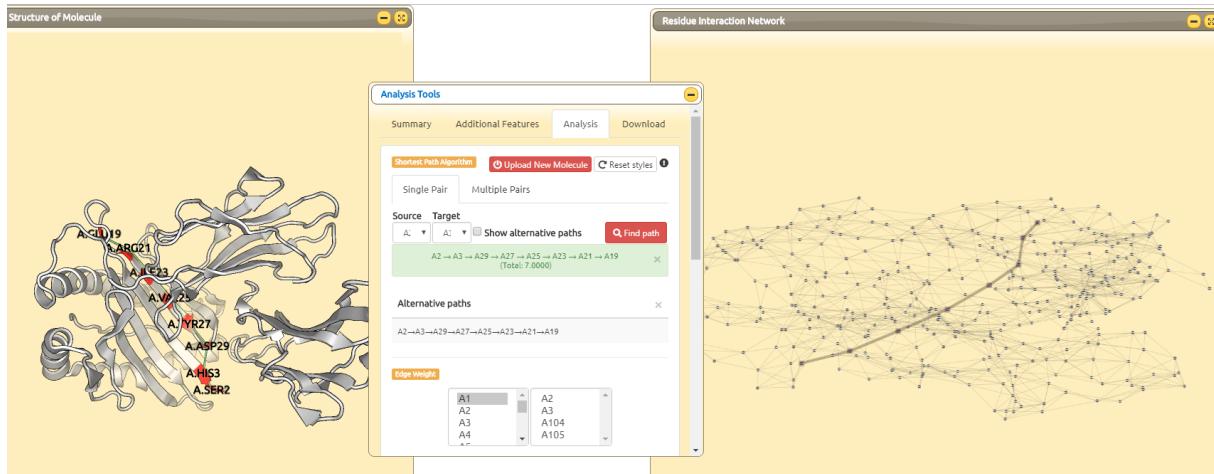


Figure 24. An example shortest path analysis result is displayed for 1OF2, between residues A2 and A19. The total edge weight is given and the path is displayed on both the 3D structure and the network view panels. For displaying alternative paths, user can tick show alternative paths button and a list of paths will be displayed on the screen.

A case study regarding the use of the shortest paths is also provided below for the interested users:

A CASE STUDY: Caspase-1 (PDB ID: 2FQQ)

Caspases-1 are known to be allosteric. It has been determined that binding of a substrate at one active site enhances catalysis at another location (Scheer J.M., Romanowski M.J., Wells J.A. (2006) A common allosteric site and mechanism in caspases Proc. Natl Acad. Sci. USA, 103, 7595-7600). Through experimental studies, it is also suggested that in this allosteric pathway, nine side chains are involved: **Glu390, Thr388, Thr334, Ser333, Ser332, Arg286, Asp336, Asn337, Ser339**, which reside in the B chain of this molecule (Datta D, Scheer JM., Romanowski MJ., Wells JA.,(2008) An Allosteric Circuit in Caspase-1, Journal of Molecular Biology, 381, 5, 1157-1167).

List of shortest paths computed between Glu390 and Ser339 using some selective network types are provided in the table below. The pathways are generated after both chains (A, B) of 2FQQ are used to construct the network. The residues that are actually present in the experimentally determined pathway are marked in bold.

As expected, upon construction of the network, when edge weight is added, the path does not change. The only factor that affects the paths generated is the initial network type and the thresholds chosen at that stage, since this selection specifies the nodes and the general infrastructure of the network.

As can be deducted from the results, the networks constructed using C-alpha are much more useful if the user is interested in shortest path generation.

C-alpha	Unweighted Cutoff:7 Å	B390 → B388 → B333 → B335 → B337 → B339	In the actual path instead of 335, 334 and 336 are present, which are the near neighbors of the residue that is predicted by our shortest path prediction.
C-alpha	Unweighted Cutoff:6 Å	B390 → B389 → B388 → B387 → B333 → B334 → B335 → B336 → B337 → B338 → B339	389, 387, 335 and 338 are near neighbors of the residue that are determined experimentally.
C-alpha	Unweighted Cutoff:5 Å	B390 → B389 → B388 → B387 → B332 → B333 → B334 → B335 → B336 → B337 → B338 → B339	389, 387, 335 and 338 are near neighbors of the residue that are determined experimentally.
C-alpha	Cutoff:7 Å ANM cutoff:15 Å (Total: 4.6460)	B390 → B388 → B333 → B335 → B337 → B339	335 is the near neighbor of 334 and 336.
C-alpha	Cutoff:7 Å GNM cutoff:7 Å (Total: 4.1720)	B390 → B388 → B333 → B335 → B337 → B339	335 is the near neighbor of 334 and 336.
Atom-pair contact	Unweighted Cutoff: 5 Å	B390 → B333 → B335 → B337 → B339	335 is the near neighbor of 334 and 336.
Interaction Strength	Unweighted Cutoff: 4%	No path can be generated (due to disconnected nodes)	
Interaction Strength	Unweighted Cutoff: 1%	B390 → B329 → A282 → A243 → A236 → A237 → A179 → A283 → B336 → B383 → B339	This pathway is not consistent with the experimentally determined allosteric pathway.

Edge Weight

For the chosen network construction type, edge weights are calculated accordingly. For any 2 residues, value of the edge weight is displayed and these residues are highlighted on the network view panel.

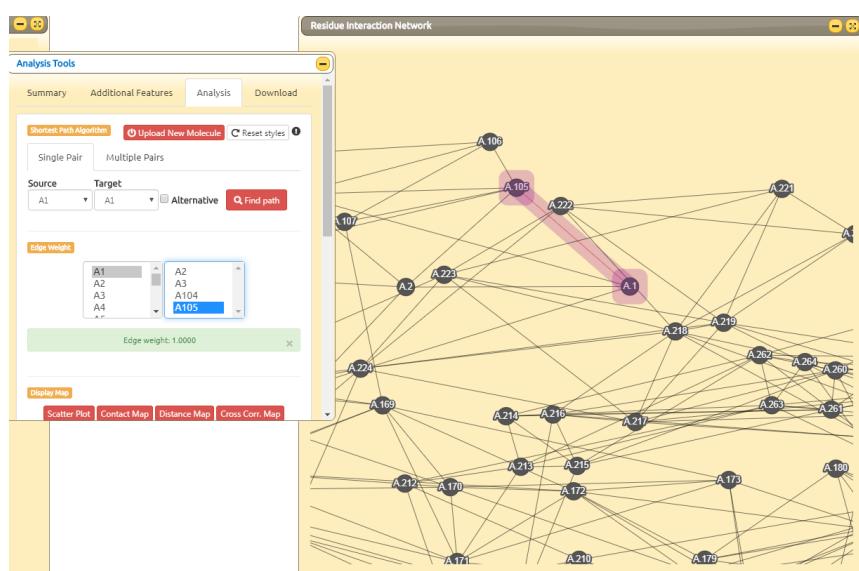


Figure 25. Display of Edge Weight option

Display Map

In this section, there are 3 maps that can be displayed.



Figure 26. Display map section

Scatter Plot Matrix: This highly informative scatter plot matrix displays scatter plots of all the network metrics and parameters (degree, betweenness, closeness, clustering coefficient, b-factor, conservation score and fluctuations) against each other. Data regarding the value of each can also be downloaded in a .txt and .xlsx formats.

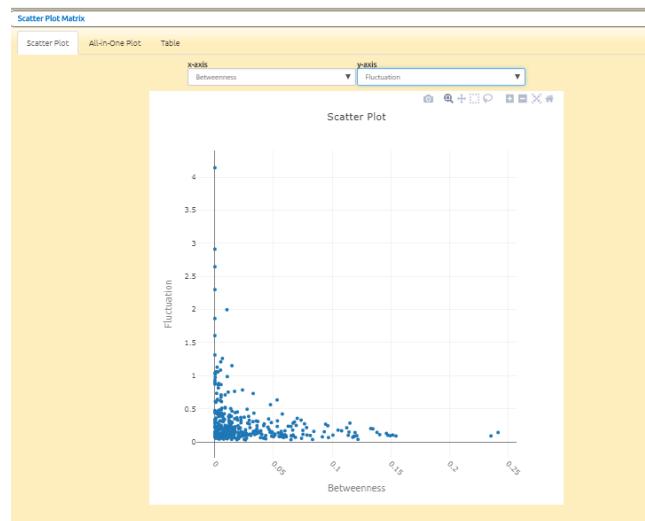


Figure 27 Scatter Plot matrix, single view

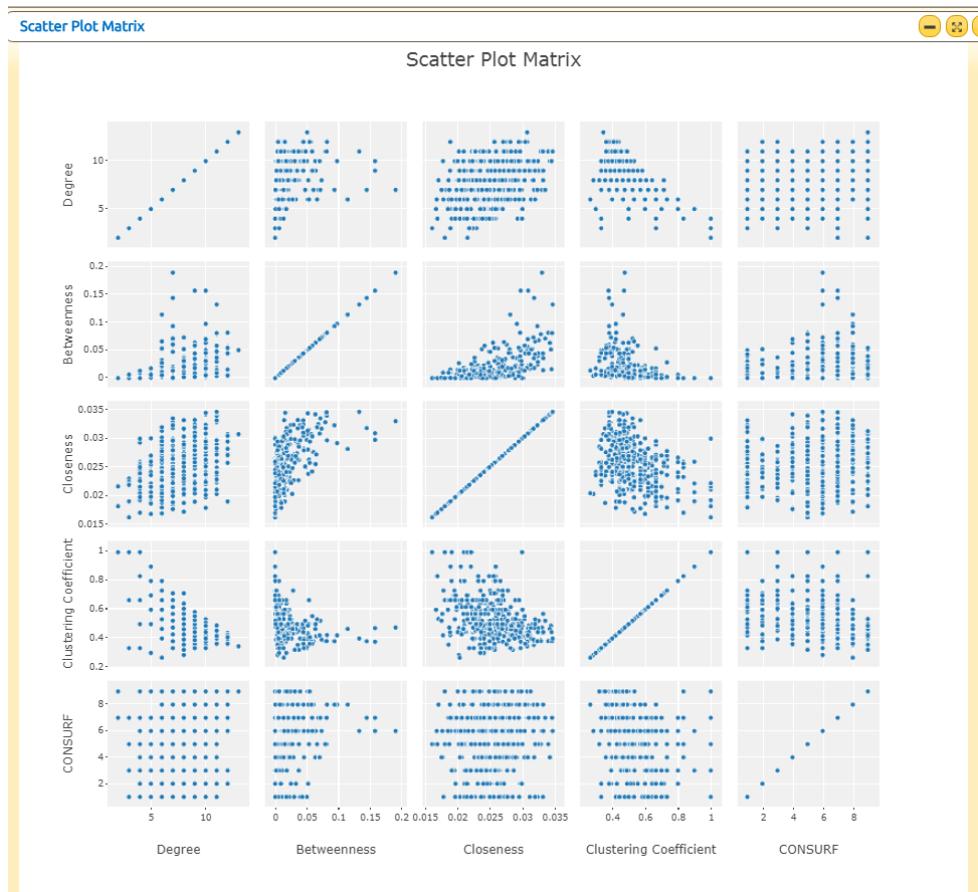


Figure 28. Display of the Scatter Plot Matrix for 1OF2, all-in-one view

Contact Map: Contact map is a display of the network in a 2D matrix view. Each dot represents an edge between the i^{th} and j^{th} residues. 2D representation of the matrix is available for download in both .txt data format and .png formats. When a cell is chosen in the contact map, it is also highlighted on the 3D structure and the network view in order to ease the analysis for the user.

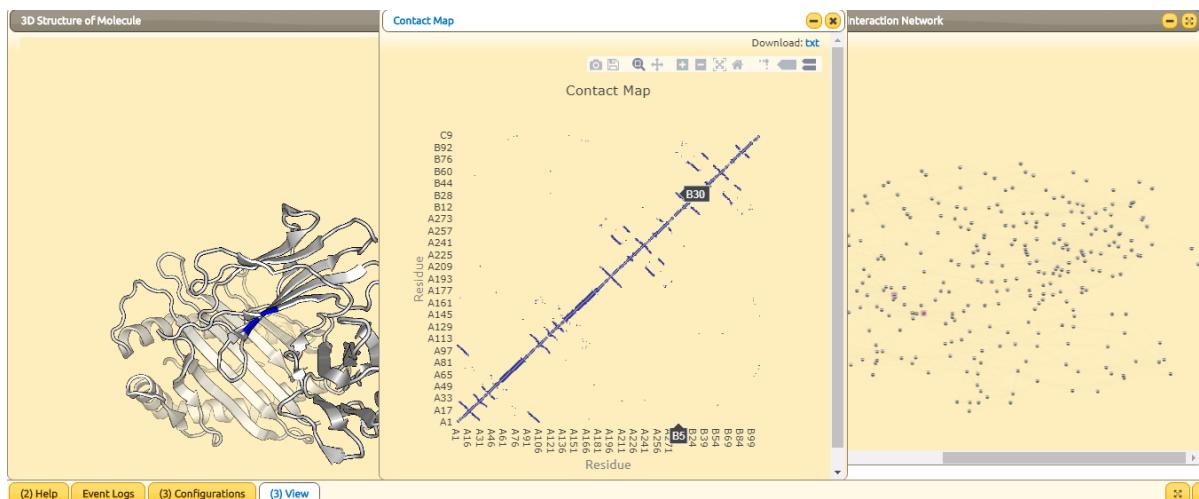


Figure 29. Display of the Contact map for 1OF2

Distance Map: Displays the distance map between the amino acid residues. It is a matrix of elements representing the distances between any two residues. 2D representation of the matrix is available for download in both *.txt* data format and *.png* formats.

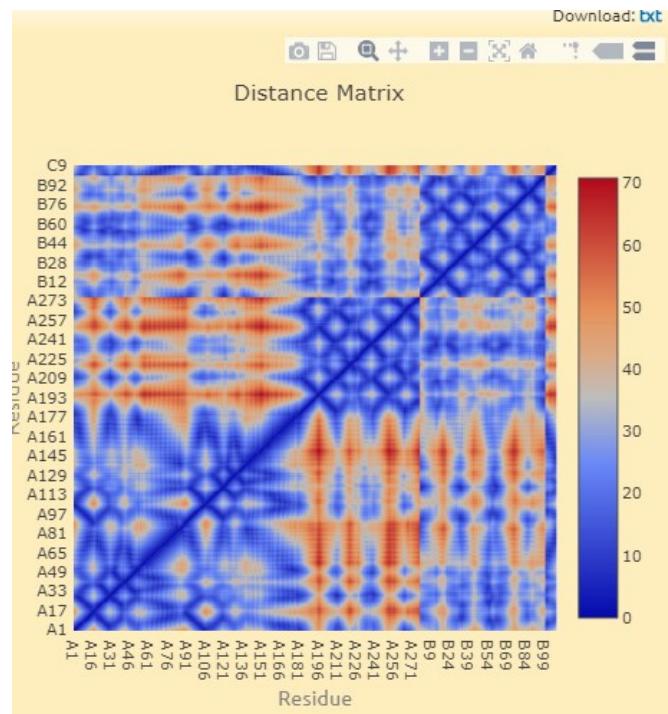


Figure 30. Display of the distance map for 1OF2

Cross Correlation Map: Displays the dynamical cross correlation data obtained from NMA. This view is available only when the network is constructed either by using ANM/GNM data or PDBj data. 2D representation of the matrix is available for download in both *.txt* data format and *.png* formats.

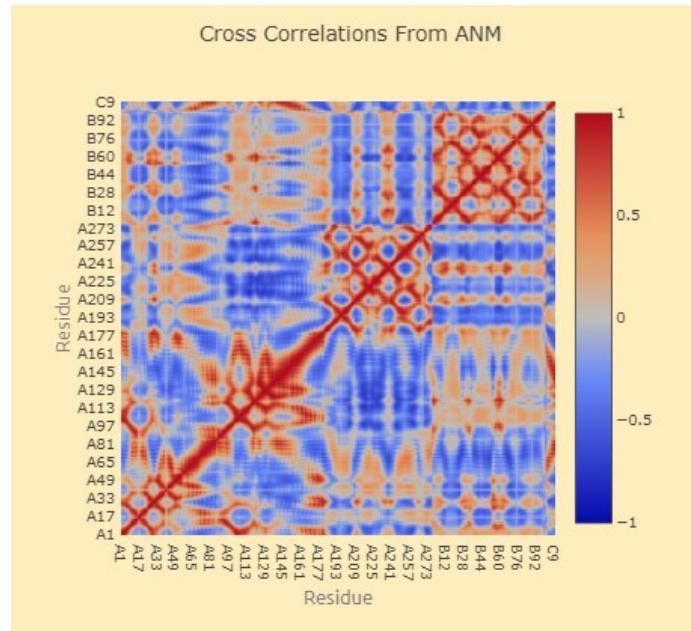


Figure 31. Display of the Cross correlation data for 1OF2 obtained from ANM

Residue Metrics

Residue metrics analysis section includes the data of square fluctuations (available when a weighted network is constructed using NMA data), sequence conservation scores and b-factor. Centrality measures are also included including degree, closeness and betweenness calculations. The results can be normalized or can be visualized directly. The values can be displayed by sorting them according to the residue number or metric value.

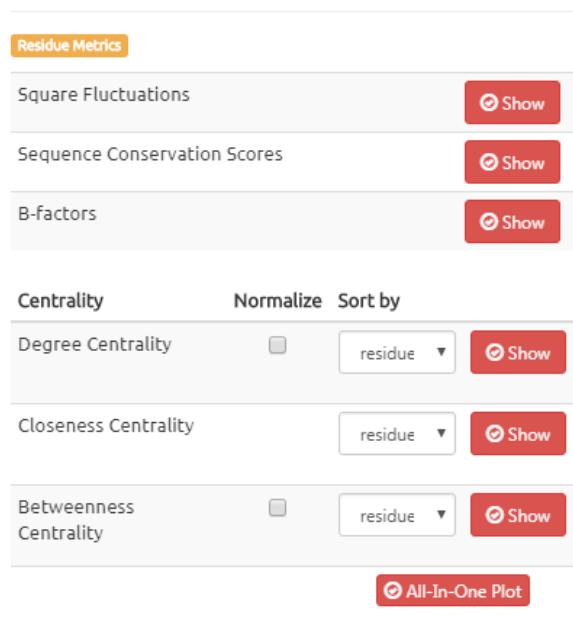


Figure 32. Residue metrics section

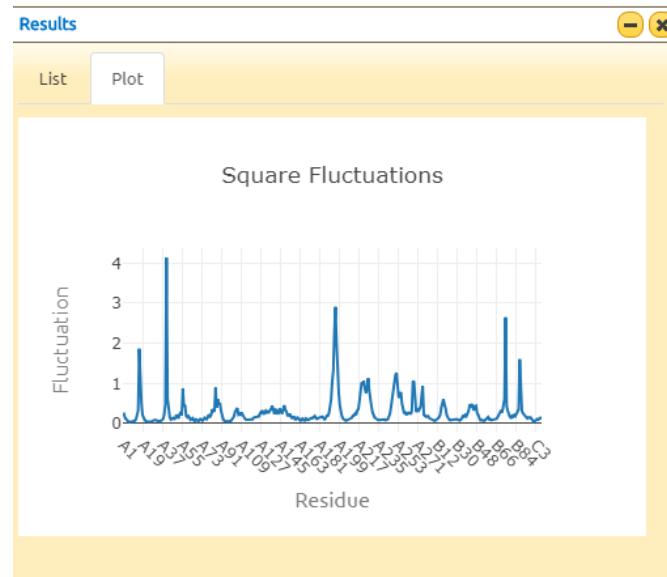


Figure 33. Square fluctuations for 1OF2 obtained from ANM calculations.

Degree: Degree is the measure of the direct neighbors of a node. For a carbon-alpha network constructed for chain A of 1OF2, normalized degree values are given in the following figure. The values can be listed, the list can be downloaded in .txt and .xlsx formats. When a certain residue is clicked in the list, it is also shown on the 3D structure. In the network panel, nodes are also colored and sized according to the degree values. A plot option for the degree results is also available for analysis.



Figure 34. Degree centrality results of a network constructed for chain A of 1OF2.

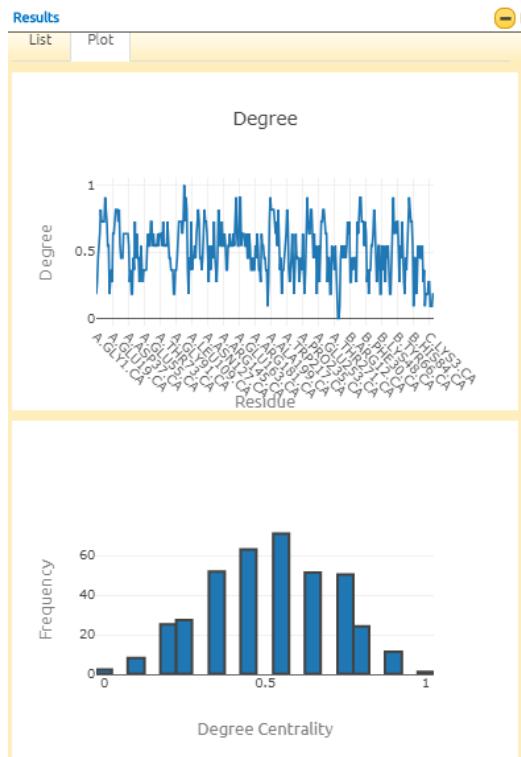


Figure 35. Plot of degree centrality values

Closeness centrality: Closeness centrality is the inverse of the total shortest path distance of the node to all other nodes of the network. For a carbon-alpha network constructed for chain A of 1OF2, closeness centrality values are given in the following figure. The values can be listed, the list can be downloaded in .txt and .xlsx formats. When a certain residue is clicked in the list, it is also zoomed in the 3D structure. In the network panel, nodes are also colored and sized according to the degree values. A plot option is also available for further analysis.

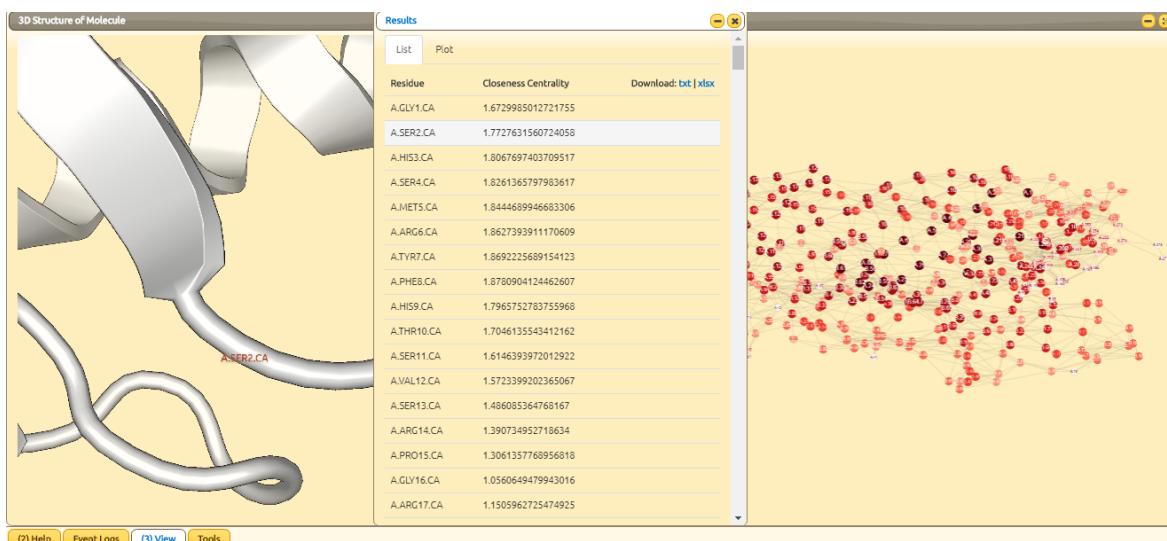


Figure 36. Closeness centrality results of a network constructed for chain A of 1OF2.

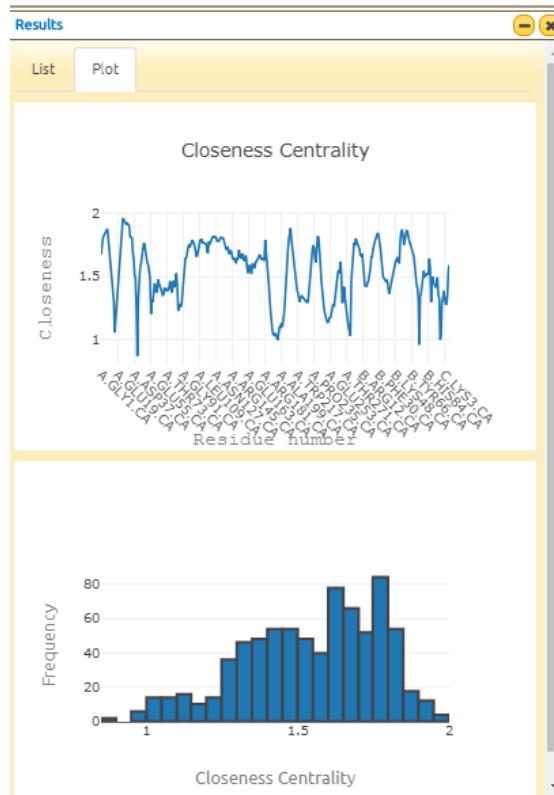


Figure 37. Plot of closeness centrality results for a network constructed for chain A of 1OF2

Betweenness centrality: Betweenness centrality is the ratio of shortest paths passing through the node. For a carbon-alpha network constructed for chain A of 1OF2, normalized betweenness centrality values are given in the following figure. The values can be listed, the list can be downloaded in .txt and .xlsx formats. When a certain residue is clicked in the list, it is also shown by zooming in the 3D structure. In the network panel, nodes are also colored and sized according to the degree values. A plot option is also available for further analysis.

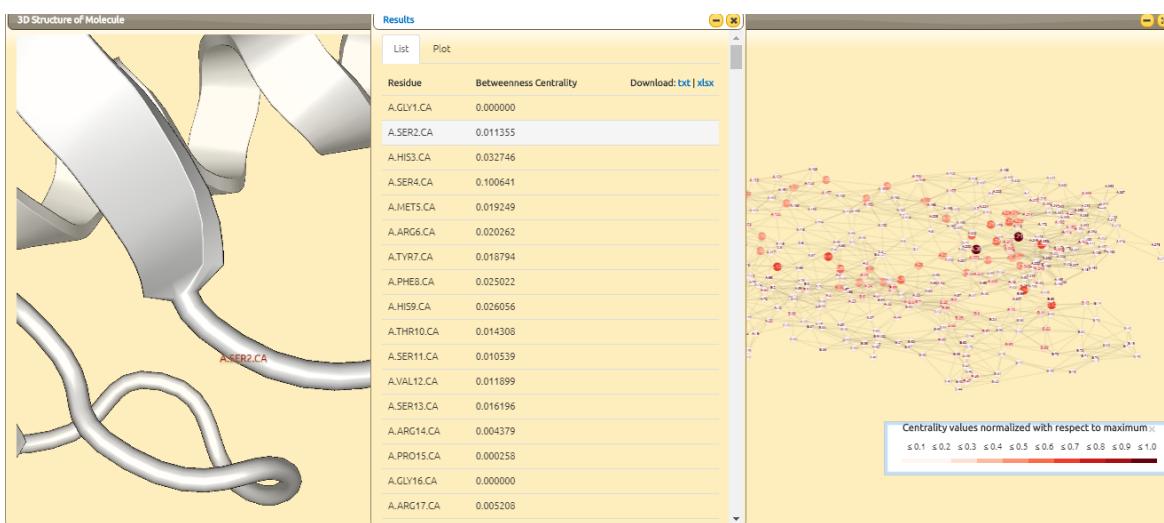


Figure 38. Betweenness centrality results of a network constructed for chain A of 1OF2.

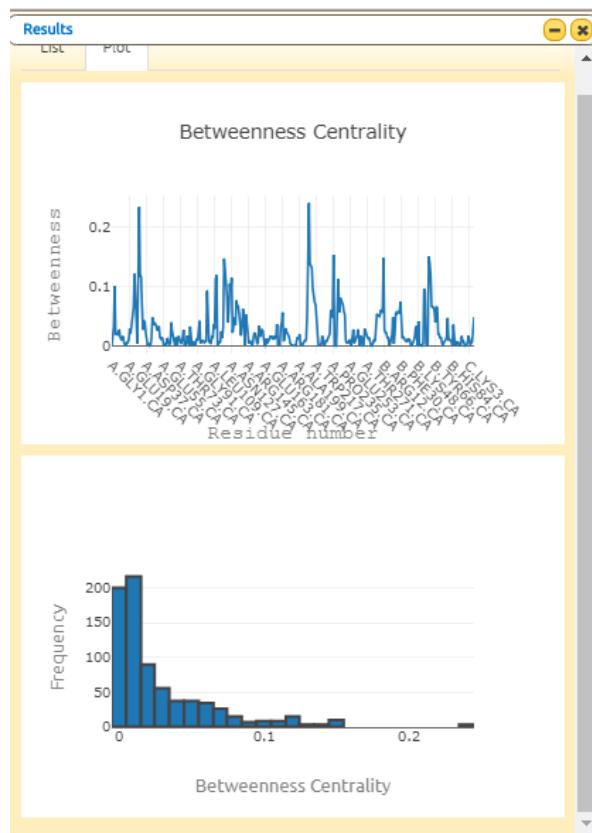


Figure 39. Plot of betweenness centrality results for a network constructed for chain A of 1OF2

An all-in-one plot is available containing all residue metrics plots as well.

k-clique Analysis

K-clique defines the sub-network of k nodes with all the k nodes connected to each other. The size of the k can be entered by the user. The resulting k -cliques will be listed and the selected one will be highlighted in both the 3D protein structure and the network view. The list of cliques can also be downloaded in a .txt format.

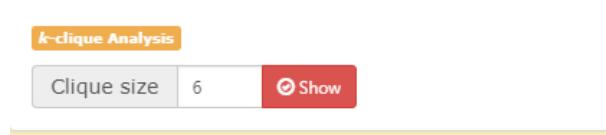


Figure 40. k-clique analysis

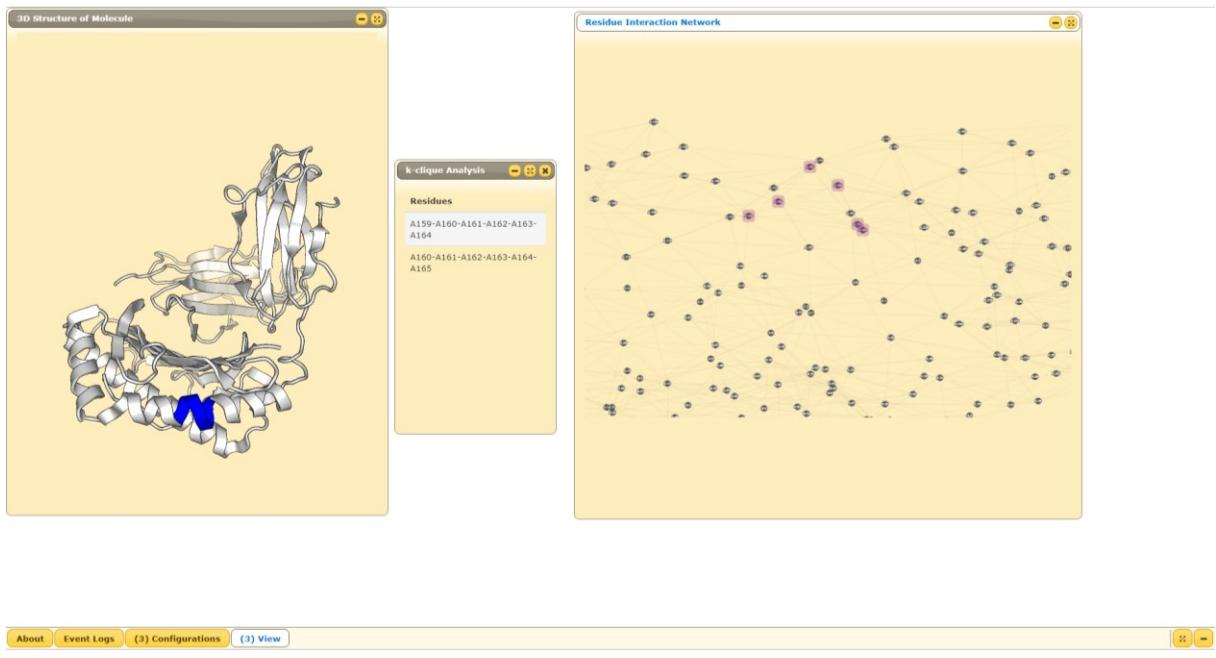


Figure 37. Display of k-clique analysis results for a clique size of 6 on 1OF2

References

1. Brinda,K.V. and Vishveshwara,S. (2005) A Network Representation of Protein Structures: Implications for Protein Stability. *Biophysical Journal*, 89, 4159–4170. <https://doi.org/10.1529/biophysj.105.064485> <http://www.ncbi.nlm.nih.gov/pubmed/16150969>
2. Bahar,I., Atilgan,A.R. and Erman,B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2, 173–181.[https://doi.org/10.1016/S1359-0278\(97\)00024-2](https://doi.org/10.1016/S1359-0278(97)00024-2)
3. Atilgan,A.R., Durell,S.R., Jernigan,R.L., Demirel,M.C., Keskin,O. and Bahar,I. (2001) Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model.
4. Haliloglu,T., Bahar,I. and Erman,B. (1997) Gaussian Dynamics of Folded Proteins.
5. Akira R. Kinjo, Gert-Jan Bekker, Hirofumi Suzuki, Yuko Tsuchiya, Takeshi Kawabata, Yasuyo Ikegawa, Haruki Nakamura; Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures, *Nucleic Acids Research*, Volume 45, Issue D1, 4 January 2017, Pages D282–D288, <https://doi.org/10.1093/nar/gkw962>
6. Kannan,N. and Vishveshwara,S. (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *Journal of Molecular Biology*, 292, 441–464. <https://doi.org/10.1006/JMBI.1999.3058>
7. Onur Serçinoğlu, Pemra Ozbek; gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W554–W562, <https://doi.org/10.1093/nar/gky381>
8. Harry C Jubb, Alicia P Higuero, Bernardo Ochoa-Montaña, Will R Pitt, David B Ascher, Tom L Blundell (2017) Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures, *Journal of Molecular Biology*, Volume 429, 3, 365-371, <https://doi.org/10.1016/j.jmb.2016.12.004>.
9. Meytal Landau, Itay Mayrose, Yossi Rosenberg, Fabian Glaser, Eric Martz, Tal Pupko, Nir Ben-Tal; ConSurf 2005: the projection of evolutionary conservation scores of

residues on protein structures, Nucleic Acids Research, Volume 33, Issue suppl_2, 1 July 2005, Pages W299–W302, <https://doi.org/10.1093/nar/gki370>