ProSNEx: a web-based application for exploration and analysis of protein structures using network formalism

Rasim Murat Aydınkal1,2,†, Onur Serçinoğlu1,3,† and Pemra Ozbek1,∗

1Department of Bioengineering, Faculty of Engineering, Marmara University, Kadıköy, Istanbul 34722, Turkey, 2Ali Nihat Gokyigit Foundation, Etiler, Istanbul 34340, Turkey and 3Department of Bioengineering, Faculty of Engineering, Recep Tayyip Erdoğan University, Rize 53100, Turkey

ABSTRACT
ProSNEx (Protein Structure Network Explorer) is a web service for construction and analysis of Protein Structure Networks (PSNs) alongside amino acid flexibility, sequence conservation and annotation features. ProSNEx constructs a PSN by adding nodes to represent residues and edges between these nodes using user-specified interaction distance cutoffs for either carbon-alpha, carbon-beta or atom-pair contact networks. Different types of weighted networks can also be constructed by using either (i) the residue-residue interaction energies in the format returned by gRINN, resulting in a Protein Energy Network (PEN); (ii) the dynamical cross correlations from a coarse-grained Normal Mode Analysis (NMA) of the protein structure; (iii) interaction strength. Upon construction of the network, common network metrics (such as node centralities) as well as shortest paths between nodes and k-cliques are calculated. Moreover, additional features of each residue in the form of conservation scores and mutation/natural variant information are included in the analysis. By this way, tool offers an enhanced and direct comparison of network-based residue metrics with other types of biological information. ProSNEx is free and open to all users without login requirement at http://prosnex-tool.com.

INTRODUCTION
Proteins mediate a great number of functions in a living cell. The folded state (i.e. the shape) plays a major role in determining thermostability (1–3), dynamics (4,5) and thereby the function of the protein. The shape, in turn, is highly dependent on the sequence of amino acid residues, their physicochemical properties and the various types of chemical interactions (bonded or non-bonded) they are involved in within the folded state (5–12). Sequence conservation and protein stability as well as dynamics have been found to be closely connected in several protein families (6,7,13–17).

In order to better understand the role played by each residue in shaping protein function, it is proper to consider the protein structure as a network of amino acids that interact with each other in the three-dimensional space via various chemical interactions. In the past few years, network formalism has been a popular approach for studying individual proteins as well as protein complexes with the aim of understanding the underlying structural organization within the protein structure and elucidating the importance and functional roles of individual residues (18–23). In this approach, a Protein Structure Network (PSN) is constructed by taking residues within the structure as nodes and adding edges between them using distance cutoffs or other more advanced criteria. Weights can also be given to edges or residues to construct weighted networks in order to emphasize the interaction strength, e.g. by using force-field based interaction energies, atom-atom contacts or pairwise residue dynamic coupling. Dynamic coupling between residues can be obtained by constructing Dynamical Cross-Correlation Maps (DCCMs) from Molecular Dynamics (MD) simulation trajectories (24–29). If interaction energies are used for edge weight assignment, the network becomes a ‘Protein Energy Network’ (PEN) (30–33). Once the network is constructed, an analysis of residue-based local or global network metrics including centrality measures can be useful to detect non-evident functions of residues such as relative importances for protein stability (19,30,34–36), allosteric communication between parts of protein (37–42) and other family-specific functions (e.g. catalytic sites in enzymes) (19,43–45).

The network approach is very useful for a fast initial characterization of protein dynamics as well. To this end, Elastic Network Models (ENMs) have found widespread usage.

†To whom correspondence should be addressed. Tel: +90 216 418 23 57 / 1697; Email: pemra.ozbek@marmara.edu.tr
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
among researchers (46–52). In an ENM, usually a selected set of atoms from each residue (e.g. carbon-alpha atoms) are connected to each other with hypothetical springs, yielding a harmonic interaction network. A Normal Mode Analysis (NMA) is then performed to generate harmonic vibrational modes of motion and flexibility profiles. Despite the assumption of harmonic motion around the native protein structure, flexibility predictions by the ENM approach has been found to correlate well with the global cooperative modes of motions generated from experimentally elucidated conformation ensembles or atomistic MD simulations (52). Following the first application by Tirion on all-atom systems (53), it has been also found that reducing the resolution by coarse-graining the protein structure and using only alpha carbons to represent the topology of the protein structure is sufficiently enough for accurate predictions. There are two basic types of ENMs: the Gaussian Network Model (GNM) (47) and the Anisotropic Network Model (ANM) (48,50,54). In GNM, fluctuations are assumed to be isotropic (i.e. distributed equally along different directions in the coordinate system) whereas in ANM, directionality of motions, and thus the three-dimensionality, is included.

Taking only a single protein structure and cutoff distances as input, simple PSN and ENM approaches are fast and powerful for getting valuable initial insights regarding the structural and dynamical behavior of a protein. To this end, several freely accessible web-services have been offered to researchers for the construction, visualization and analysis of PSNs and ENMs. For PSNs, NAPS (55) server offers an analysis based on a variety of weighted and unweighted network types. RING 2.0 (56) offers a similar service in addition to identifying different types of chemical interactions between residues and the results can be visualized by using Cytoscape (57). Similarly, RINalyzer (58) and structure-Viz integrate Cytoscape (57) and UCSF Chimera (59) to create and visualize RINS, CSU (60), PSN-Ensemble (61), NetworkAnalyzer (62) and PyMOL (63) plug-ins xPyder (64) and PyInteraph (65) can also be used for similar purposes. The Protein Contact Atlas (66) offers a rich visualization interface for examining different types of contacts within protein structures, including residue-centric network metrics alongside mapping of custom metrics such as sequence conservation or thermodynamics stability changes that can be supplied by the user. VERMONT 2.0 (67) is another web-server for performing network analysis on protein structures while integrating additional features such as sequence conservation, residue physicochemical parameters and solvent accessibility. For ENM-based protein dynamics analysis, eNemno (68), the ANM 2.1 (69,70), iGNM 2.0 (71) and DynOmis ENM (72) servers are available. Bio3D-web also offers user-friendly interface for performing NMA (Bio3D-NMAweb) on single protein structures (73). DynaMut server offers NMA-based prediction of stability changes upon amino-acid mutations (74). WebPSN web-server combines ENM and NMA approaches for finding allosteric pathways in protein structures (75).

In a typical research workflow involving an investigation of relationships between sequence, structure, dynamics and function of a given protein of interest; all relevant features, such as conservation scores for the amino-acid sequence, residue-centric network metrics, flexibility profiles and other functional annotations must be obtained to facilitate the analysis. Researches with sufficient technical expertise may choose to utilize customized workflows by utilizing multiple software packages for this purpose. For non-experts, however, the options are limited. Although some of the aforementioned web-servers offer an integration of more than one type of feature to some extent, there is currently no web-service available offering extensive comparative analysis features involving PSN analysis, sequence conservation and protein flexibility profiles as well as other functional annotations.

Aiming to fill this gap, we have developed the ProSNEx (Protein Structure Network Explorer) web-server. ProSNEx offers enhanced PSN-based protein structure analysis by integrating sequence conservation profiles, protein annotations and flexibility profiles from different types of ENM into a single user interface.

MATERIALS AND METHODS

Protein structure networks and elastic network models

ProSNEx offers the construction of several types of un-weighted and weighted networks. Unweighted networks can be constructed by selecting one of the four methods for edge assignment between nodes (carbon-alpha, carbon-beta, atom-pair interaction and interaction strength). In the carbon-alpha and carbon-beta networks, only carbon-alpha or carbon-beta atom positions from the input PDB structure are used to specify the edges in the network, respectively. Edges between nodes are added only if two atoms are closer than the user-specified threshold (cutoff) distance. In the atom-pair interaction network, edges between nodes are added if at least one atom from a residue is closer than the user-specified distance threshold to at least one atom from another residue. In the interaction strength network, edges between nodes are added only if the interaction strength is higher than the user-specified threshold. The interaction strength calculations are based on (34).

Once the network edge assignment type is selected, ProSNEx can also assign weights to edges, yielding a weighted network. Here, four options are offered to the user. Using either one of the first two options; the user can construct a weighted PSN by extracting weights from DCCMs calculation from of GNM/ANM simulations or the NMA method of Wako et al. (76–78). In the latter case, normalized and time-averaged cross-correlations, as reported at the Promode-elastic database of PDB Japan, are used (77,79). Hence, this option is not available for custom PDB files. The GNM and ANM calculations are performed by ProSNEx server.

In the third option, the user is given the option to choose average force-field based interaction energies from a MD simulation trajectory as edge weights to construct a PEN. Here, average interaction energies in the file format returned by gRNN (31) is accepted. In the last option, the interaction strength values are taken from (80).

Following network construction, global and local network metrics are calculated. Specifically, average degree, path length, network density, clustering coefficient as well as node (residue)-centric metrics including node degrees,
betweenness-centrality and closeness-centrality are calculated and reported. Additionally, shortest path between two selected nodes and $k$-cliques are calculated as well.

**Sequence conservation profiles, protein annotations and interatomic interactions**

If available, ProSNEx retrieves sequence conservation from the CONSURF database (CONSURF-DB), which includes conservation scores calculated by the CONSURF method (81–83). If the user supplies a custom PDB file instead of a PDB access code or if no CONSURF score is available at CONSURF-DB for the given PDB access code, an option to import conservation scores in the format returned by the CONSURF server is available following network construction. ProSNEx also retrieves annotations (if available) from the Uniprot database, including sequence variants, mutagenesis experiment results, etc. (84,85). Finally, interatomic interactions within the input structure are retrieved by using Arpeggio and annotated on the network structure (86).

**RESULTS**

**Use case: investigating the sequence–structure–dynamics relationships in TEM-1 $\beta$-lactamase**

We demonstrate the use of ProSNEx server for investigating the sequence–structure–dynamics relationships in a TEM-1 $\beta$-lactamase enzyme from *Escherichia coli*.

ProSNEx main page includes a simple interface for entering a PDB code, selection of protein chains (if applicable) and network settings, respectively. (Figure 1A-B). Starting from a TEM-1$\beta$-lactamase structure (PDB code: 1ZG4), a weighted PSN (carbon-alpha network, threshold distance: 7 Å) was constructed using cross-correlations from an ANM simulation (threshold distance: 15 Å) as edge weights.

Figure 1C–E gives an overview of outputs from ProSNEx. Upon finishing calculations, the tool presents three major windows for displaying the 3D structure of the input molecule (Figure 1C), a 2D network representation (Figure 1E) and a window titled ‘Analysis Tools’ (Figure 1D). The Analysis Tools window is the access point for investigating further features included in ProSNEx analysis. In Figure 2, a collection of plots, all accessible from the Analysis Tools window, is given. Figure 2C and Figure 2C shows two scat-
Figure 2. (A) Comparison of multiple node (residue) metrics on protein structure. (B) An ‘all-in-one’ Scatter Plot shows scatter plots between degree, betweenness-centrality, closeness-centrality, fluctuations (if NMA is used for edge-weight assignment), sequence conservation scores, B-factors and clustering coefficients. (C, D) Selected scatter plots between closeness-centrality and fluctuation and betweenness-centrality and sequence conservation scores. (E) Cross correlation plot when NMA is used for edge-weight assignment.

For plots of particular interest: in the first one, the closeness centrality is seen to be correlated to fluctuation profiles from ANM, confirming its usefulness as a rigidity descriptor (87). In Figure 2D, high betweenness-centrality values are seen to coincide well with highly conserved residues, highlighting a relationship between sequence evolution and dynamic cross-talk within the β-lactamase structure.

IMPLEMENTATION

The tool is implemented in JavaScript and jQuery. The web-interface is built using Bootstrap CSS style. PSN is constructed and analysed using by JSNetworkX (88) library. Cytoscape.js (57) framework is used for network visualization. PV is used for protein structure visualization (89). For GNM and ANM calculations, ProDy is used (90).

CONCLUSION

We have developed ProSNE server which provides a novel and enhanced analysis of protein structures using network formalism. The tool is designed to be very user friendly and easily adaptable for all researchers in the field of protein structural biology. The major novelty of the tool lies in the presence of features such as: (i) comparison of multiple single residue metrics from network analysis as well as other additional information such as sequence conservation scores and Uniprot annotations and (ii) usage of dynamic cross-correlations between pairs of amino acids from NMA in the weighted PSNs.

DATA AVAILABILITY

ProSNE server is free and open to all users without login requirement. The server does not store any data submitted by the user. It is compatible with major web browser including Chrome, Firefox and Safari.

ACKNOWLEDGEMENTS

O.S. conceived the idea of server; R.M.A. designed the user interface, constructed network analysis computations and retrieval of external data; O.S. constructed the backend GNM and ANM computations; P.O. wrote the help documents; O.S. and P.O. provided critical input on analysis methods. O.S. and P.O. wrote the manuscript. P.O. supervised the project and oversaw the manuscript preparation. We acknowledge the helpful discussion with Fatma Gizem Avci on the structural details of TEM-1 β-lactamase. TUBITAK 218M861 is acknowledged.

FUNDING

Funding for open access charge: Self funded. Conflict of interest statement. None declared.

REFERENCES

1. Robertson, A.D. and Murphy, K.P. (1997) Protein structure and the energetics of protein stability. Chem. Rev., 97, 1251–1268.
2. Baldwin, R.L. (2007) Energetics of protein folding. J. Mol. Biol., 371, 283–301.
