

Structural bioinformatics

WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules

Michele Seeber, Angelo Felling, Francesco Raimondi, Simona Mariani and Francesca Fanelli*

Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 183, 41125 Modena, Italy

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on 28 July 2014; revised on 21 October 2014; accepted on 23 October 2014

Abstract

Summary: We developed a mixed Protein Structure Network (PSN) and Elastic Network Model-Normal Mode Analysis (ENM-NMA)-based strategy (i.e. PSN-ENM) to investigate structural communication in biomacromolecules. The approach starts from a Protein Structure Graph and searches for all shortest communication pathways between user-specified residues. The graph is computed on a single preferably high-resolution structure. Information on system's dynamics is supplied by ENM-NMA.

The PSN-ENM methodology is made of multiple steps both in the setup and analysis stages, which may discourage inexperienced users. To facilitate its usage, we implemented WebPSN, a freely available web server that allows the user to easily setup the calculation, perform post-processing analyses and both visualize and download numerical and 3D representations of the output. Speed and accuracy make this server suitable to investigate structural communication, including allosterism, in large sets of bio-macromolecular systems.

Availability and implementation: The WebPSN server is freely available at <http://webpsn.hpc.unimore.it>.

Contact: fanelli@unimo.it

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Concepts and methods borrowed from graph theory are being increasingly used to study several aspects of structural biology, including structural communication that comprises allosterism. The graph-based approach defined as Protein Structure Network (PSN) has been implemented in the freely available Wordom software (Seeber *et al.*, 2011). It computes network features (e.g. nodes, hubs, links, etc.) and shortest communication pathways on ensembles of structures generally derived from Molecular Dynamics (MD) simulations. With this approach, information on system's dynamics participates both in building the Protein Structure Graph (PSG) and in searching for the shortest communication pathways.

In a recent development of the method intended for high-throughput investigations, the PSG is computed on a single high-resolution structure rather than an MD trajectory and information on system's dynamics (i.e. cross-correlation of atomic motions) is supplied by Elastic Network Model-Normal Mode Analysis (ENM-NMA) (Raimondi *et al.*, 2013) (see [Supplementary Methods](#) for detailed information). The approach, hereafter indicated as PSN-ENM, is powerful and fast but complex in the setup and analysis steps. To foster wide usage, the method was therefore implemented in a web server (WebPSN) that provides the user with a friendly graphical interface for setting the input and visualizing the output, while hiding all computation and data management steps.

2 Features of the WebPSN server

2.1 Workflow of the PSN-ENM method

The structural communication between two distal sites is inferred from the computation of the shortest paths (i.e. chains of pairwise linked nodes) between pairs of non-linked nodes (by default, the two extremities must be separated by at least five nodes). Figure 1 shows a workflow of the PSN-ENM approach implemented in the WebPSN server (see [Supplementary Methods](#) for more detailed information). The first step consists in performing the PSN analysis on a single high-resolution structure, which means building the PSG. The latter is searched for the shortest paths between pairs of nodes. The algorithm first defines all possible communication paths between selected node pairs and then it filters the results according to cross-correlation of atomic motions, as derived from ENM-NMA. Filtering consists in retaining only the shortest path(s) that contains at least one residue correlated (i.e. with a cross-correlation value ≥ 0.6) with either one of the two extremities (i.e. the first and last amino acids in the path). Meta paths made of the most recurrent nodes and links in the path pool (i.e. global meta paths) are worth computing to infer a coarse/global picture of the structural communication in the considered system ([Raimondi et al., 2013](#)). In detail, meta paths are made of nodes present in $\geq 5\%$ of the considered path pool (i.e. ‘frequent nodes’), and of links satisfying both conditions of being present in $\geq 5\%$ of the paths and of connecting ‘frequent nodes’.

Selection of single pathways can be additionally done on the basis of a number of indices (see below the ‘Output’ section). To facilitate path analysis, the WebPSN server allows the grouping of similar paths into clusters. The user can then focus on the structural communication described by a given cluster, e.g. the most populated one, by considering the cluster meta path or the cluster center.

2.2 Input

The WebPSN server needs a protein structure in the Protein Data Bank (PDB) format as an input. The user can either specify the PDB ID for retrieval from the data bank (<http://www.rcsb.org>) or upload a structure of interest from the local computer. The server allows also to upload more than one structure, which is instrumental in the parametrization of non-yet parametrized residues (see [Supplementary Methods](#) and [Supplementary Fig. S1](#)).

Those residues selected for the computation of PSG and communication pathways are interactively highlighted on the protein structure in the Jmol applet (implemented using the javascript version of Jmol, i.e. JSmol, [Supplementary Fig. S1](#)).

Depending on the users’ expertise, a basic or an advanced input mode can be chosen.

It is possible to follow the whole PSN-ENM procedure, which leads to the computation of both PSG and communication pathways, or compute just the PSG. Thus, the ENM method is activated only if the ‘PSN-Path’ option is selected (see [Supplementary Methods](#) and [Supplementary Fig. S1](#)).

Advanced setting is available as well for PSN, ENM and PSN-Path for expert users (see [Supplementary Methods](#)).

2.3 Output

Results can be both interactively visualized on the web page and downloaded. Web 3D visualizations (i.e. by the JSmol applet) include PSG, hub distribution, meta paths and centers of the three most populated clusters, which can be selected through a predisposed box (see [Supplementary Methods](#) for an explanation of the graphical representations).

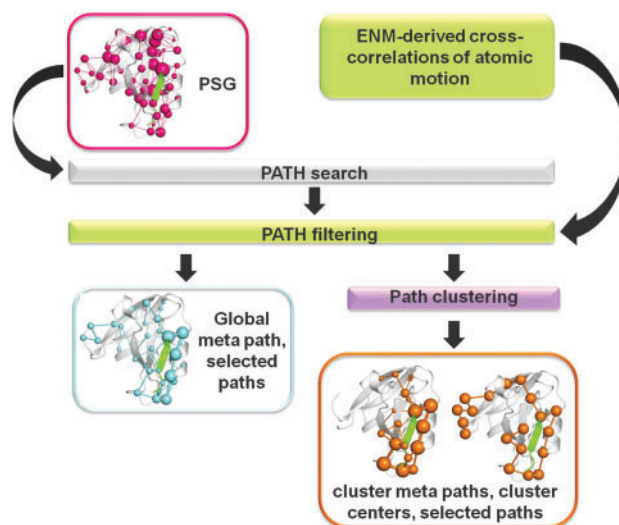


Fig. 1. Flowchart of the approach. The top image is the 3D representation of a PSG. Nodes are spheres centered on the C α -atoms, whose diameter is proportional to the number of links made by the node. The middle image is the 3D representation of a global meta path. Sphere (i.e. node) diameters and link thickness are, respectively, proportional to node and link frequencies in the pool of paths. The bottom image shows the 3D representation of a cluster meta path (left) and of the center of a path cluster. Whereas in the cluster meta path (left) sphere diameters and link thickness are proportional to node and link frequencies in the cluster paths, in the cluster center (right), all spheres and links are equi-dimensional. The 3D protein representations have been drawn by the PyMOL 0.99rc6 software

An interactive table lists all pathways with a number of associated indices, such as: (i) path length; (ii) mean square distance fluctuations between all node pairs in a path (p) (MSDF p), which is a measure of path stiffness (see [Supplementary Methods](#); [Raimondi et al., 2013](#)); (iii) hub content; (iv) fraction of user-defined relevant amino acid nodes; (v) correlation score, i.e. the ratio between number of correlated residues and path length (the latter excludes the two extremities); (vi) average similarity score (according to the clusterization method) between the path and each path in the cluster; and (vii) number of the cluster the path belongs to (notice that such number increases with decreasing cluster population). The table can be sorted by any of these indices, or filtered according to index cutoff values chosen by the user. Any predicted pathway can be visualized on the 3D structure by clicking the path itself in the table.

The output page shows also the distribution of (i) path length; (ii) MSDF p index; (iii) hub content; (iv) correlation score; (v) node frequency in paths; and (vi) frequency of four amino acid fragments in the predicted paths. A table is shown as well listing the number of nodes, hubs and links in the PSG or in each node cluster as well as the number of paths, nodes and links in those paths.

Downloads are available both for the 3D representations of the output (as pymol and VMD scripts) and the numerical data (as csv and plain text files). The downloadable zip files also contain an output manual explaining the meaning of output files and numerical data.

2.4 Tutorial: a PDZ2 domain as a case study

The PSN-ENM approach has been validated on a PDZ2 domain in its free state and in complex with a C-terminal peptide ([Raimondi et al., 2013](#)).

Collectively, predictions by the PSN–ENM method were significant and valuable. The statistical significance of such predictions was verified through a two-tailed Fisher's exact test, which gave a P -value of 6.4×10^{-3} , and the computation of sensitivity, which resulted equal to 0.786.

The tutorial and almost all default settings of WebPSN are based on the study on PDZ2.

3 Concluding remarks

The WebPSN server implements the PSN–ENM method for a fast and user-friendly systematic scanning of the shortest communication pathways between all residue pairs in a biomolecular system.

High speed and accuracy make the WebPSN server suitable to high-throughput investigation of the structural communication,

including allosterism, in biomacromolecular systems in different functional states.

Funding

This study was supported by an Airc-Italy grant [IG10740] and a Telethon-Italy grant [S00068TELC to F.F.].

Conflict of Interest: none declared.

References

- Raimondi, F. *et al.* (2013) A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems: the PDZ2 domain from tyrosine phosphatase 1E as a case study. *J. Chem. Theory Comput.*, **9**, 2504–2518.
- Seeber, M. *et al.* (2011) Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.*, **32**, 1183–1194.