Bioinformatics, 33(6), 2017, 935–937
doi: 10.1093/bioinformatics/btw756
Advance Access Publication Date: 16 December 2016
Applications Note

Structural bioinformatics

The PPI3D web server for searching, analyzing and modeling protein–protein interactions in the context of 3D structures

Justas Dapkūnas[1], Albertas Timinskas[1], Kliment Olechnovič[1,2], Mindaugas Margelevičius[1], Rytis Dičiūnas[1] and Česlovas Venclovas[1,*]

[1]Institute of Biotechnology, Vilnius University, Vilnius LT-10257, Lithuania and [2]Faculty of Mathematics and Informatics, Vilnius University, Vilnius LT-03225, Lithuania

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on September 30, 2016; revised on November 7, 2016; editorial decision on November 19, 2016; accepted on November 22, 2016

Abstract

Summary: The PPI3D web server is focused on searching and analyzing the structural data on protein–protein interactions. Reducing the data redundancy by clustering and analyzing the properties of interaction interfaces using Voronoi tessellation makes this software a highly effective tool for addressing different questions related to protein interactions.

Availability and Implementation: The server is freely accessible at http://bioinformatics.lt/software/ppi3d/.

Contact: ceslovas.venclovas@bti.vu.lt

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Using modern techniques it is relatively easy to find out whether some proteins interact. However, these data have only limited value without knowing how the proteins interact. The interaction details can be obtained from the three-dimensional (3D) structures of protein complexes. Unfortunately, determining protein structures is often difficult. Although the overall number of the structures of protein complexes in the Protein Data Bank (PDB) is steadily growing (Berman *et al.*, 2013), there is a large gap between the number of known protein interactions and the number of solved structures of corresponding protein complexes.

Similarly to individual proteins, protein complexes tend to preserve their 3D structures in the course of evolution (Aloy *et al.*, 2003). Due to this tendency, the information about protein interactions can be inferred using structures of their homologs. Such homology-based analysis may be widely applicable, and a number of software tools exploiting the principle of structural conservation have been developed (Esmaielbeiki *et al.*, 2016; Kawabata, 2016; Szilagyi and Zhang, 2014).

Dealing with the structural data on protein interactions requires addressing at least two important issues. The first one is the definition of the interaction interface. Oftentimes, interfaces are defined based on distances between atoms and their solvent-accessible surface areas. However, the tessellation-based definition of interfaces is more advantageous. Protein structure tessellation naturally defines pairs of interacting residues, makes it possible to directly calculate their contact areas and provides a more detailed representation of the interaction interface (Ban *et al.*, 2006; Cazals *et al.*, 2006).

The second issue is the highly redundant nature of the structural data in the PDB (Berman *et al.*, 2013). Clustering structures by sequence similarity can solve this problem for individual proteins. However, such approach is not suitable for retrieving a non-redundant set of protein–protein interfaces as the same proteins might interact through alternative binding sites (Hamp and Rost, 2012; Kundrotas and Vakser, 2013).

Here we present the PPI3D web server, which can be used for searching, analyzing and modeling pairwise protein–protein interactions (Fig. 1). The definition of interaction interfaces and their detailed analysis in PPI3D are based on the Voronoi tessellation. All the pairwise protein interactions are clustered according to both the protein sequence and the interaction interface similarity. The latter is

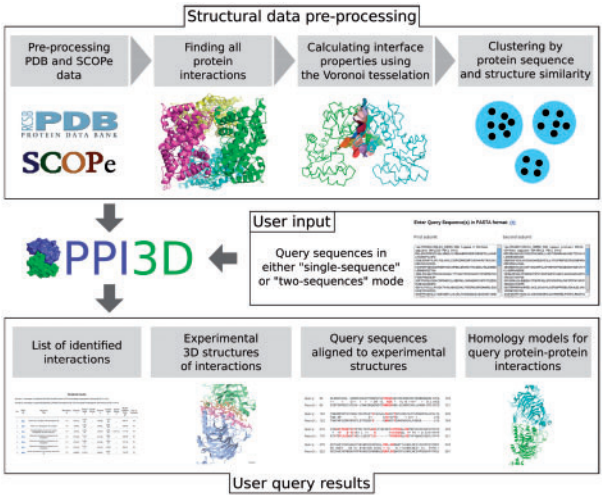

Fig. 1. Schematic workflow of the PPI3D web server

computed using a superposition-free contact area-based approach. Clustering reduces the data redundancy over sevenfold while keeping the alternative binding modes. Thus, PPI3D enables users to get both detailed and condensed view of protein interactions in the context of structural data.

2 Pre-processing of the structural data

The source of structural data for PPI3D is PDB biological assemblies. Protein structures are extracted from all PDB entries having resolution better than 4 Å. They are divided into proteins and peptides. Additionally, structural domains are identified (Fox *et al.*, 2014) in order to provide a possibility to also explore the domain–domain interactions (see Supplementary data for more details on data processing).

The interfaces of all protein–protein, protein–peptide and domain–domain interactions are analyzed by means of the Voronoi tessellation (Olechnovič and Venclovas, 2014). To reduce the data redundancy, all protein interfaces are first clustered by the sequence similarity (Li and Godzik, 2006). The resulting sequence clusters are additionally split according to the similarity of protein binding sites and interaction interfaces defined by CAD-score (Olechnovič *et al.*, 2013) and its derivatives (see Supplementary data for the description and examples of clustering).

The pre-processed data including the computed interface properties and clustering results are stored in a database. New PDB structures are appended every week using the same pipeline. This ensures that the PPI3D users can find and analyze new structural data on protein–protein interactions soon after their release in PDB.

3 Using the PPI3D web server

The PPI3D server has three modes for exploring protein interactions: (i) the search with a single sequence (sequence group) detects any protein and peptide binding sites in the query and its homologs; (ii) the search with two sequences (sequence groups) finds interactions between a given pair of sequences and their homologs and (iii) the search with the PDB code queries all protein interactions within a single PDB entry.

The sequence search modes accept protein sequences in FASTA format or corresponding UniProt Accession Numbers (ACs). Every input sequence is searched against protein sequences associated with the structural protein interaction data using either BLAST (for the detection of close homologs) or PSI-BLAST (for both close and remote homologs) by the BLAST+ package (Camacho *et al.*, 2009). After the sequence search finishes, PPI3D provides a summary and a detailed list of the interactions found for each query sequence or each pair of sequences. The results are clustered by default and only the representatives are displayed. Clustering levels may be interactively modified.

The users can also visually compare interactions from multiple clusters. Aligned sequences offer a condensed view of binding sites from different clusters at the residue level, whereas superimposed structures show the variability of protein interactions at the structural level.

Every interaction interface can be explored in detail including its total area, interface-forming residues and the extent of their burial, residue–residue contacts across the interface (hydrogen bonds, disulphide and salt bridges are labeled) and ligands. Details of interaction interfaces can be visualized using JSmol or PyMOL. The interface residues are also mapped onto the sequence alignment, which can be used to generate a homology model for the complex.

4 Conclusions

The PPI3D web server features a simple and user-friendly interface. The server provides both summarized and detailed information on protein interactions in different forms including both protein sequence and structure. PPI3D may be useful for various research tasks such as analyzing all available structural data on specific protein interactions, inferring interaction properties from structures of homologous protein complexes or detecting binding partners and interaction sites for a given protein. The server may be useful for anyone interested in protein interactions at the level of 3D structures including beginners and experts, experimental and computational biologists.

Funding

This work has been supported by the intramural funds of Vilnius University and by postdoctoral fellowship (to J.D.) funded by European Union Structural Funds project 'Postdoctoral Fellowship Implementation in Lithuania'.

Conflict of Interest: none declared.

References

Aloy,P. *et al.* (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332**, 989–998.

Ban,Y.E.A. *et al.* (2006) Interface surfaces for protein–protein complexes. *J. ACM*, **53**, 361–378.

Berman,H.M. *et al.* (2013) Trendspotting in the Protein Data Bank. *FEBS Lett.*, **587**, 1036–1045.

Camacho,C. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cazals,F. *et al.* (2006) Revisiting the Voronoi description of protein–protein interfaces. *Protein Sci.*, **15**, 2082–2092.

Esmaielbeiki,R. *et al.* (2016) Progress and challenges in predicting protein interfaces. *Brief. Bioinform.*, **17**, 117–131.

Fox,N.K. *et al.* (2014) SCOPe: Structural Classification of Proteins–extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.*, **42**, D304–D309.

Hamp,T. and Rost,B. (2012) Alternative protein–protein interfaces are frequent exceptions. *PLoS Comput. Biol.*, **8**, e1002623.

Kawabata,T. (2016) HOMCOS: an updated server to search and model complex 3D structures. *J. Struct. Funct. Genomics*, doi:10.1007/s10969-016-9208-y.

Kundrotas,P.J. and Vakser,I.A. (2013) Protein–protein alternative binding modes do not overlap. *Protein Sci.*, **22**, 1141–1145.

Li,W. and Godzik,A. (2006) CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Olechnovič,K. *et al.* (2013) CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins*, **81**, 149–162.

Olechnovič,K. and Venclovas,Č. (2014) Voronota: a fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J. Comput. Chem.*, **35**, 672–681.

Szilagyi,A. and Zhang,Y. (2014) Template-based structure modeling of protein–protein interactions. *Curr. Opin. Struct. Biol.*, **24**, 10–23.