

# PIMADb: A Database of Protein–Protein Interactions in Huge Macromolecular Assemblies



Oommen K. Mathew<sup>1,2</sup> and Ramanathan Sowdhamini<sup>1</sup>

<sup>1</sup>National Centre for Biological Sciences (TIFR), GKVK Campus, Bangalore, India. <sup>2</sup>SASTRA University, Tirumalaisamudram, Thanjavur, Tamil Nadu, India.

**ABSTRACT:** Protein–protein interactions play a very important role in the process of cellular functionality. Intricate details about the interactions between the proteins in a macromolecular assembly are important to understand the function and significance of protein complexes. We are reporting about a database of protein–protein interactions in huge macromolecular assemblies (PIMADb) that records the intrinsic details of 189,532 interchain interactions in 40,049 complexes from the Protein Data Bank. These details include the results of the quantification and analysis of all the interactions in the complex. The availability of interprotomer interaction networks can enable the design of point mutation experiments. PIMADb can be accessed from the URL: <http://caps.ncbs.res.in/pimadb>

**KEYWORDS:** protein–protein interaction, macromolecule, protein assembly, interface residues

**CITATION:** Mathew and Sowdhamini. PIMADb: A Database of Protein–Protein Interactions in Huge Macromolecular Assemblies. *Bioinformatics and Biology Insights* 2016;10:105–109 doi: 10.4137/BBI.S38416.

**TYPE:** Original Research

**RECEIVED:** March 15, 2016. **RESUBMITTED:** May 11, 2016. **ACCEPTED FOR PUBLICATION:** May 21, 2016.

**ACADEMIC EDITOR:** Thomas Dandekar, Associate Editor

**PEER REVIEW:** Three peer reviewers contributed to the peer review report. Reviewers' reports totaled 357 words, excluding any confidential comments to the academic editor.

**FUNDING:** This study was supported by the Centre of Excellence (COE) grant (BT/01/COE/09/01) by Department of Biotechnology (DBT), Government of India. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**CORRESPONDENCE:** [mini@ncbs.res.in](mailto:mini@ncbs.res.in)

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE). Provenance: the authors were invited to submit this paper.

Published by Libertas Academica. Learn more about this journal.

## Introduction

Protein–protein interactions are defined as specific physical contacts between protein pairs that occur by selective molecular docking in a particular biological context.<sup>1</sup> Protein–protein interactions play a crucial role in biological systems, varying from enzymatic involvement to signal transduction. Many of the important biological functions involve huge multi-component complexes of proteins such as ribosomes. Proteins execute the genetic programme encoded in the genome of an organism, and the total collection of proteins produced by a cell at any given time constitutes the proteome. The highly dynamic nature of the proteome is well observed in the developmental stages and in the presence of external stimuli. The members of a proteome provide a network of interactions in which they support and regulate each other.<sup>2</sup> These interactions are specific physical contacts established between two or more proteins as a result of biochemical events and/or electrostatic forces and can be permanent (obligatory) or temporal/transient (nonobligatory) in nature.

Hydrophobic interactions have been observed as one of the dominant driving forces in the formation of protein–protein associations.<sup>3</sup> However, electrostatic interactions have been identified to be strongly affecting the rate at which the protein–protein association is occurring.<sup>4</sup> The average size of protein–protein interfaces is approximately 10% of a typical

monomer surface and the packing of the atoms at the recognition site and the interface is very tight.<sup>5</sup> A study on dimeric protein complexes reveals that the proteins existing in the multimeric form have larger hydrophobic interfaces than the ones that exist independently.<sup>6</sup> Similarly, previous analysis has shown that the obligatory complexes have a closely packed but less planar interface, with fewer inter-subunit hydrogen bonds than the nonobligatory complexes.<sup>7</sup> Previous studies have focused on the analysis of protein complexes, where the formation of complexes with another type of proteins (heteromeric) or even with another copy of its own (homomeric) has been studied. Likewise, transient protein–protein interactions are known to have relatively weak interactions and noted that they, upon co-operative binding to a macromolecule, become effectively permanent (eg, transient DNA-binding homodimers stabilize upon binding to their target DNA). The residue contact specificities and the residue preference at the interfaces have been well studied and are reported to be playing a crucial role in determining the type of interaction (obligatory or non-obligatory).<sup>8,9</sup> Many proteins that are known to be involved in transient interactions are also found to be retaining dynamic participation in more than one PPI. The optimization of function during the evolution is expected to be the key determinant of the observed character for each complex as the functional rationale for many of these complexes is unknown.<sup>10</sup>



Despite such general analyses and inferences, in-depth analysis of macromolecular assemblies is still lacking. PPI networks can provide a complementary view and conceptual abstraction of the biological pathways that enclose the corresponding proteins.<sup>1</sup> Earlier, we had developed PPCheck<sup>11</sup> to measure the strength of interactions using pseudoenergies. These energy values were employed to present protein–protein interactions within a macromolecular assembly using conceptualization and abstractions as networks, using our in-house Protein Interactions in Macromolecular Assemblies (PIMA) tool.<sup>12</sup> We had demonstrated that such a tool can conveniently enable the easy analysis of macromolecular assemblies, as large as 21 chains. We report a database of PPI networks of protein assemblies, as recorded in structural databanks, for 60,555 entries.

## Methodology

**PIMA tool.** Our previous method of quantifying protein–protein interactions using PPCHECK and PIMA (for assemblies) has been adopted to identify interactions in the macromolecular assemblies. Pseudoenergies are calculated, depending on the nature of the interactions, for residues that are identified within a C $\beta$ –C $\beta$  distance threshold of 10Å.<sup>11,12</sup>

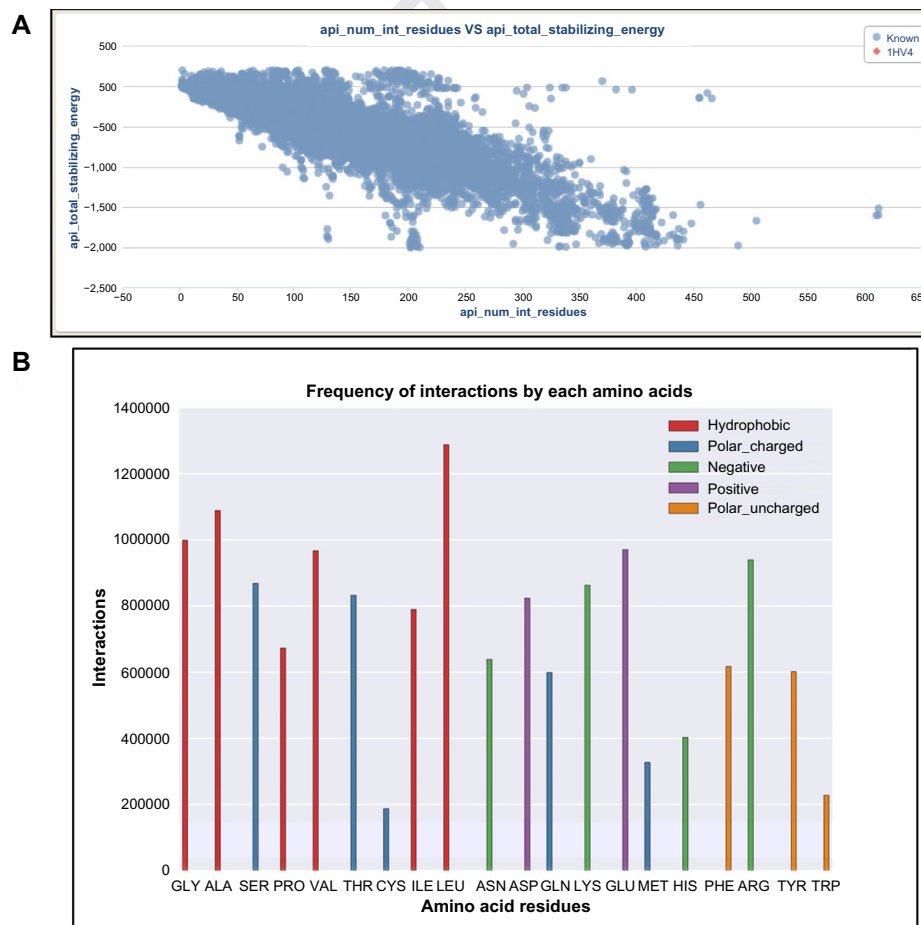
Parallel processing version of PPCheck is enabled at the background of PIMA for the energy calculations.

## Construction and Content

**Data collection and integration.** The number of protein–protein assemblies that have been collected from the Protein Data Bank (PDB)<sup>13</sup> and analyzed using the PIMA tool are 40,049. The results are recorded and organized in PIMADb. Dimers comprise the largest of assemblies in this database. Detailed statistics is provided in Supplementary Table 1.

Presently, the database contains 189,532 interactions identified from all the assemblies. The number of interface residues range from 2 to 1,844 (Fig. 1A). Figure 1B explains the frequency of each amino acid participating in an interchain interaction and providing hints about the importance of hydrophobicity over the residual mass of each amino acid.

**Database implementation.** The data have been organized as a database using the MySQL database management system and the backend logic has been implemented using Python, Python-CGI, and BioPython. The user-friendly web interface has been implemented using various technologies such as HTML5, CSS, JavaScript, Ajax, and jQuery.



**Figure 1.** (A) The correlation between the total stabilizing energy and the number of interface residues. (B) Frequency of interaction made by each amino acid that is ordered according to their residue mass.

## Utility, Features, and Discussion

**User interactive interaction network and three-dimensional visualization.** The major component of the PIMADb result page is the user interactive interaction network and the three-dimensional (3D) molecular visualization of the assembly. The network view is re-rendered in correlation with the user interactions (clicks and selections) (Fig. 2).

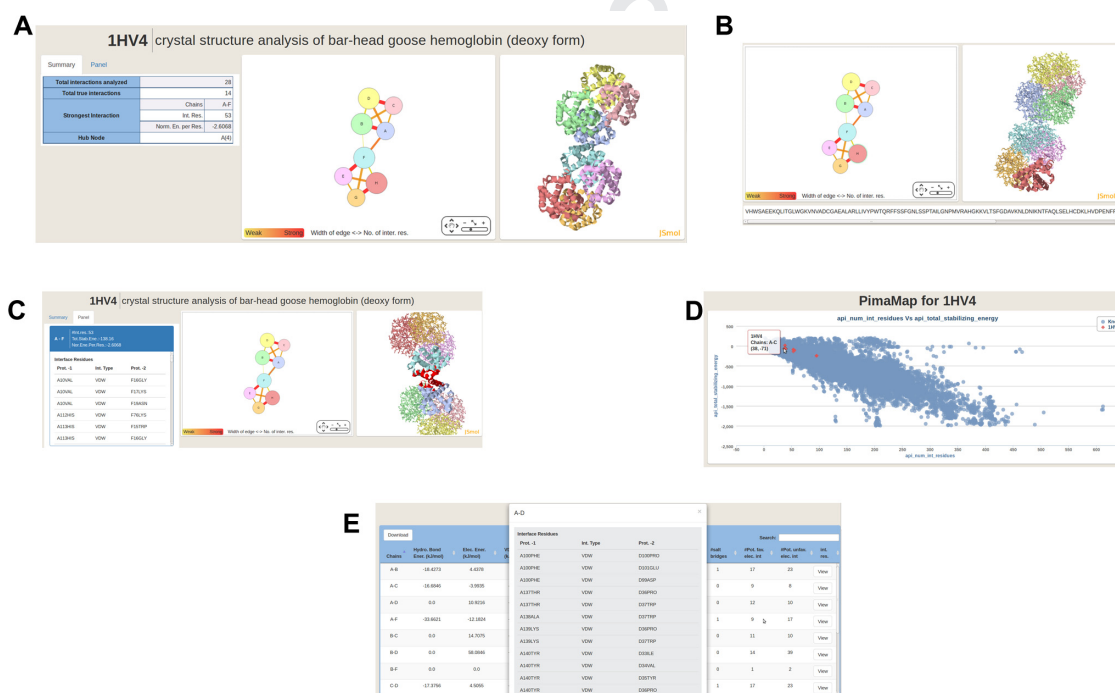
The interaction network visualizes the true interactions, identified by the underlying algorithm, PIMA<sup>12</sup> in the macromolecular assembly in the form of a graph with nodes and edges. Each protein in the assembly is represented as a node (circle) colored in the same manner as in the 3D visualization component that enables the user to establish the visual correlation between these two different visualizations. The nodes retain a label carrying the chain id of the protein. The size of the node demonstrates the size of the protein, that is, the number of atoms in it. Each edge (line) in the graph symbolizes a true interaction between the two proteins that are connected with the edge. The width of the edge demonstrates the elaboration of the interface (interface area), while the color demonstrates the strength from red (strong) to yellow (weak) as determined by PPCheck normalized energy values. The interaction network view is implemented using Cytoscape-Web.<sup>14</sup>

The 3D molecular visualization component on the result page exhibits the tertiary structural components of the molecular assembly in 3D view using Jmol.<sup>15</sup> Each protein in the assembly is displayed in the cartoon representation

and follows the standard Jmol/JSmol color coding. This view enables the user to visualize the secondary structures in each of the proteins and examine the assembly.

This interlinked visualization of the interaction network graph and the 3D display enable the user to click on the nodes or edges and fetch the intrinsic details of the proteins and interactions. These details include the number of atoms in the protein, number of interface residues, pairs of interacting residues along with the type of interaction, total stabilizing energy, and normalized energy per residue. A click on the node, which represents a protein, highlights the same protein in the 3D view with a distinguishable appearance and displays its amino acid sequence on a different panel in addition to the other details like the number of residues and atoms in it. A click on the edge, which symbolizes an interaction, highlights the details of the partner proteins involved in the particular interaction as well as the interface residues in the highly distinguishable appearance. The display of the details include the normalized energy per residue and the list of pairs of residues that are interacting distance at the interface and the nature of interactions (hydrophobic/electrostatics) as identified by PPCheck.

**PIMAMAP.** PIMAMAP is a scatter plot that displays the correlation between the number of interface residues and the normalized energy per residue (Fig. 2d). The values corresponding to the existing interactions from all the PDB complexes are marked in blue and the interactions from the



**Figure 2.** Web interface of PIMADb results. (A) Main result component that exhibits the network diagram of the interactions and the three dimensional view of the complex. (B) the protein of query is highlighted (cartoon) in the three dimensional view in response to the click on the same (node) in the network (C) the interacting proteins (cartoon) and the interface region (coloured red) is highlighted in the three dimensional view in response to the click on an edge in the network (D) the PIMAMap displays the interactions from the complex of query along with the known interactions from other complexes. (E) the list of interactions and the details of the residue pair interactions.

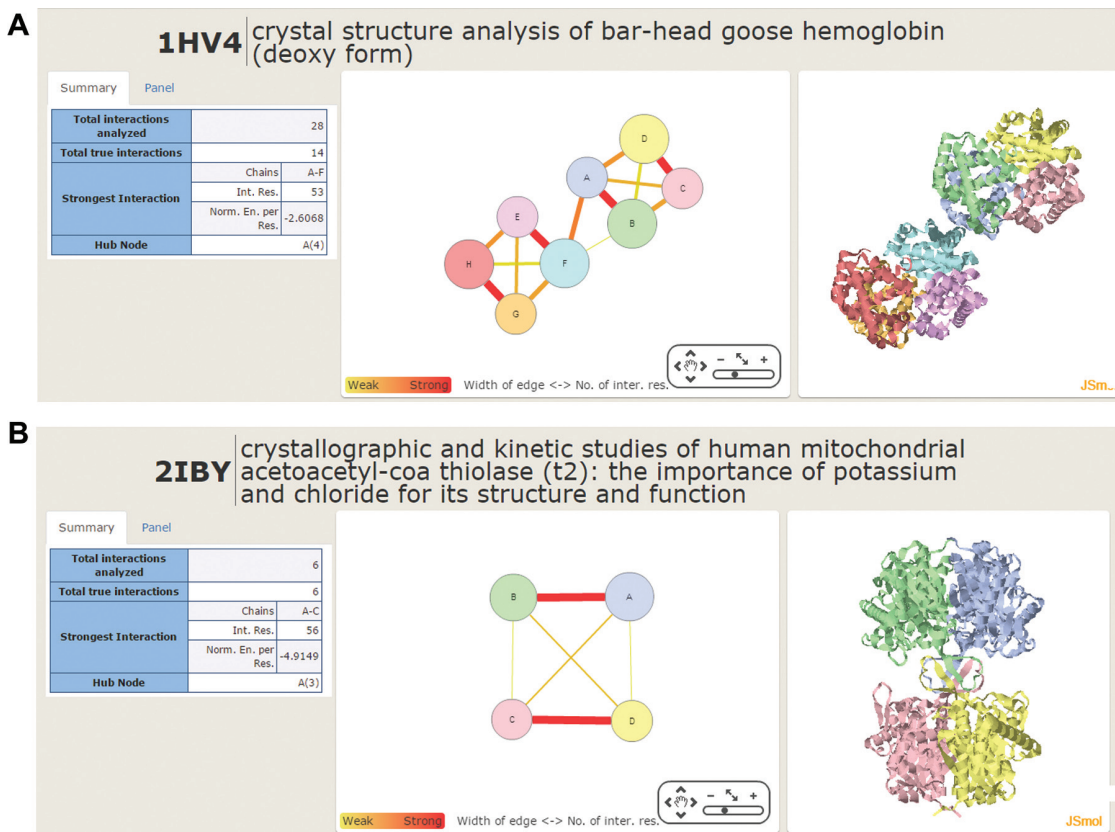


Figure 3. Symmetry of interaction pattern.

particular complex of query are marked in red. PIMAMAP is an interactive plot that enables the user to zoom in/out into a particular range and extract the details.

**Interactive list of interactions.** The result page also contains a list of all the interactions in the form of a highly interactive table that has many columns carrying different parameters such as total stabilizing energy, Van der Waals energy, and electrostatic energy of the interaction (Fig. 2E). Users can sort the rows of the table by any column in it. This table also has a feature to facilitate search and filter interactions.

**Browse through the PDBs.** The web interface of PIMADb allows the user to browse through the data in various ways, which include the number of chains and PDB Id. The list of PDBs is displayed as a table in which the rows can be filtered based on users' input. This provides an easy access to the specific complex of interest. The complete database is available for download.

**Keyword searching.** The search and find module of PIMADb allows the user to retrieve the relevant data quickly and accurately. The search can be performed through many different ways such as PDB Id and compound name.

**Inferring partner through homology.** This module of PIMADb that enables to identify the partner through homology is a great tool to obtain clues about the putative interaction partners for the protein under study. This module accepts an amino acid sequence from the user and identifies the possible partners from the known interactions through homology.

**Interaction pair searches.** This module, available at a tab on the home page, accepts a pair of amino acids as inputs and lists out the PDB complexes and the corresponding interacting chains in which this pair of residues occurs as interface residues.

### Applications of PIMADb

**Easy visualization of interchain interactions in huge complexes.** One of the main applications of PIMADb is the easy and objective analysis of large PDB complexes and the visual examination and abstraction of the interchain interactions in huge protein complexes. For example, PDB-1GAV contains a protein complex with 48 chains and the manual

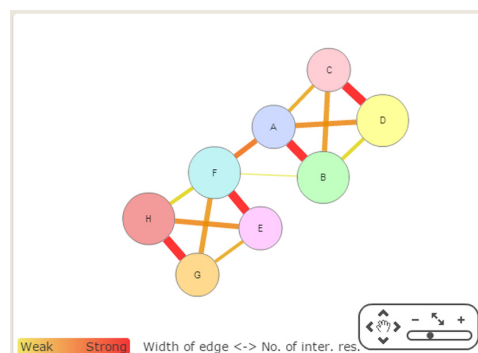


Figure 4. Symmetry of interaction network in a homo-dimer of a hetero tetramer protein complex.

analysis of the interactions is difficult or even impossible. Figure 3A demonstrates the interactions within the complex. Figure 3B exhibits the 3D view of the complex. These two components on the result page of PIMADb enable the easy retrieval and analysis of each interaction and its intrinsic details such as the strength and the elaboration.

**Recognize the symmetry in protein complexes.** Another application of PIMADb is the ease of visualizing and examining the structural symmetry (through the 3D view) and the symmetry counterparts in the interaction patterns (through the network view) of any protein complex in PDB. PDB-1HV4 is an example of a protein that is symmetrical in structure and follows the interaction pattern symmetry (Fig. 4). Graphical networks provide a rapid conceptualization of interacting chains, their interfaces and overall oligomerization patterns (such as dimer or tetramers).

**Hotspot analysis.** The fundamental details, such as the normalized energy per residue and the pairs of interface residues of each interaction, make it easy to understand the significance of each residue in a particular interaction. This enables the biologists to design the mutation studies and understand or to perturb the complex formation better.

## Conclusion

Protein–protein interactions are precise and large in the cellular environment and better understanding of the intricate details of interchain interactions in protein complexes is very important. The knowledge about the interacting partners and their fundamental details will help in understanding the function and behavior of the protein complexes. This knowledge can even provide appropriate clues to the ways of controlling the actions carried out by them using externally administered small molecules leading to drug discovery. PIMADb can provide intrinsic details, such as the symmetry of interchain interactions and the strength of the interactions and the residue pairs at the interface. PIMADb would be a great resource for analyzing and visually examining all the interacting partners and interactions in all complexes in PDB and to design experiments for biochemists and biologists.

## Acknowledgments

OKM would like to thank the Vice-Chancellor of the SAstra University for supporting his Ph.D. The work is funded by the Centre of Excellence project (BT/01/COE/09/01), Department of Biotechnology,

India. The authors also thank NCBS (TIFR) for providing adequate infrastructural support.

## Author Contributions

Conceived and designed the experiments: RS. Analysed the data: OKM. Wrote the first draft of the manuscript: OKM. Contributed to the writing of the manuscript: RS. Agree with manuscript results and conclusions: OKM, RS. Jointly developed the structure and arguments for the paper: OKM, RS. Made critical revisions and approved final version: RS. Both authors reviewed and approved of the final manuscript.

## Supplementary Material

**Supplementary Table 1.** Number of assemblies that are classified according to the number of protein chains.

## REFERENCES

1. Rivas JDL, Fontanillo C. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol*. 2010;6(6):e1000807.
2. von Mering C, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*. 2002;417(6887):399–403.
3. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci*. 1997;6(1):53–64.
4. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein–protein interactions. *Curr Opin Struct Biol*. 2000;10(2):153–9.
5. Conte LL, Chothia C, Janin J. The atomic structure of protein–protein recognition sites 11 Edited by A. R. Fersht. *J Mol Biol*. 1999;285(5):2177–98.
6. Nooren IM, Thornton JM. Structural characterisation and functional significance of transient protein–protein interactions. *J Mol Biol*. 2003;325(5):991–1018.
7. Jones S, Thornton JM. Review principles of protein–protein interactions. *Proc Natl Acad Sci U S A*. 1996;93(January):13–20.
8. De S, Krishnadev O, Srinivasan N, Rekha N. Interaction preferences across protein–protein interfaces of obligatory and non-obligatory components are different. *BMC Struct Biol*. 2005;5(1):15.
9. Swapna LS, Bhaskara RM, Sharma J, Srinivasan N. Roles of residues in the interface of transient protein–protein complexes before complexation. *Sci Rep*. 2012;2:1–9.
10. Nooren IM. NEW EMBO MEMBER'S REVIEW: diversity of protein–protein interactions. *EMBO J*. 2003;22(14):3486–92.
11. Sukhwal A, Sowdhamini R. PPCheck: a webserver for the quantitative analysis of protein–protein interfaces and prediction of residue hotspots. *Bioinform Biol Insights*. 2015;9:141–51.
12. Mathew OK, Sowdhamini R. PIMA: protein–protein interactions in macromolecular assembly – a web server for its analysis and visualization. *Bioinformatics*. 2016;12(1):9–11.
13. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–42.
14. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. Cytoscape web: an interactive web-based network browser. *Bioinformatics*. 2010;26(18):2347–8.
15. Hanson RM. Jmol – a paradigm shift in crystallographic visualization. *J Appl Crystallogr*. 2010;43(5 pt 2):1250–60.