PPI3D: a web server for searching, analyzing and modeling protein–protein, protein–peptide and protein–nucleic acid interactions

Justas Dapkūnas [1], Albertas Timinskas[1], Kliment Olechnovič [1,2], Miglė Tomkuvienė[1] and Česlovas Venclovas [1,*]

[1]Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio av. 7, Vilnius LT-10257, Lithuania
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

*To whom correspondence should be addressed. Tel: +370 5 223 4368; Email: ceslovas.venclovas@bti.vu.lt

Abstract

Structure-resolved protein interactions with other proteins, peptides and nucleic acids are key for understanding molecular mechanisms. The PPI3D web server enables researchers to query preprocessed and clustered structural data, analyze the results and make homology-based inferences for protein interactions. PPI3D offers three interaction exploration modes: (i) all interactions for proteins homologous to the query, (ii) interactions between two proteins or their homologs and (iii) interactions within a specific PDB entry. The server allows interactive analysis of the identified interactions in both summarized and detailed manner. This includes protein annotations, structures, the interface residues and the corresponding contact surface areas. In addition, users can make inferences about residues at the interaction interface for the query protein(s) from the sequence alignments and homology models. The weekly updated PPI3D database includes all the interaction interfaces and binding sites from PDB, clustered based on both protein sequence and structural similarity, yielding non-redundant datasets without loss of alternative interaction modes. Consequently, the PPI3D users avoid being flooded with redundant information, a typical situation for intensely studied proteins. Furthermore, PPI3D provides a possibility to download user-defined sets of interaction interfaces and analyze them locally. The PPI3D web server is available at https://bioinformatics.lt/ppi3d.

Graphical abstract
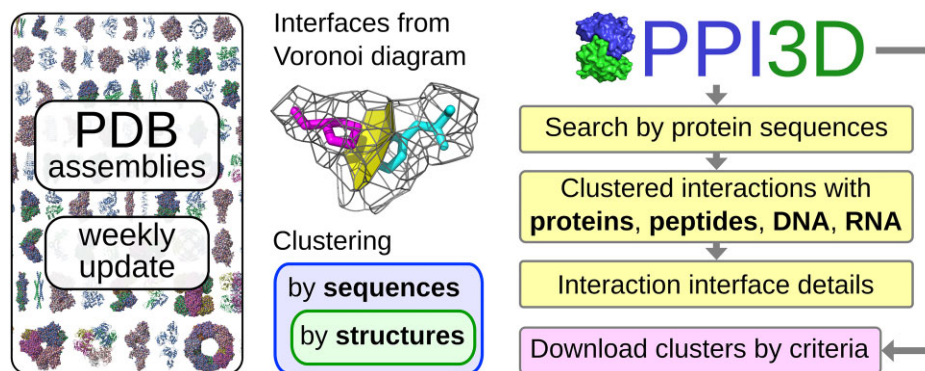
Introduction

Proteins drive most biological processes, but they rarely act alone. Most often, proteins perform molecular functions by forming stable or temporary complexes with other proteins, peptides, nucleic acids and ligands. For comprehensive understanding of biological processes at the molecular level it is essential to know not only protein interaction partners, but also details of these interactions. This information can be obtained directly from the corresponding three-dimensional (3D) structures of protein complexes. These structures can be either determined experimentally or predicted computationally. A recent deep learning-driven breakthrough (1) resulted in accu-

rately predicted structures for millions of individual proteins (2). However, predicting structures for protein–protein and protein–peptide complexes remains challenging (3,4), whereas prediction of protein-DNA or protein-RNA complexes is even harder (5). Therefore, the ability to utilize experimentally determined structures of protein complexes, available in the Protein Data Bank (PDB) (6) is very important for both experimentalists interested in specific proteins and computational biologists aiming at developing methods for modeling protein complexes.

The number of protein complexes in the PDB is already quite large and is growing steadily, providing a rich source

for structure-resolved interaction data. However, it is not always straightforward to extract, analyze and make use of these data. One of the issues with structures determined by X-ray crystallography is how to distinguish biologically relevant interactions from those resulting from crystal packing. Another confounding issue is the redundancy of interaction data. There may be multiple PDB entries for a given protein complex, and even a single PDB entry may contain several instances of this complex. This redundancy cannot be decreased by simple sequence-based filtering, usually sufficient to obtain representative monomeric structures. The interactions within the same or a closely related complex may differ depending on conditions in which the structure was solved, the presence or absence of ligands and/or additional interacting partners. Therefore, obtaining representatives for interaction interfaces necessitates involvement of structure-based comparison. This is not a trivial endeavor in itself, because the results depend on the interface definition and the interface similarity metric.

Over the years, multiple web-based tools have been developed to address these issues and to make use of PDB interaction data for better understanding protein interactions and functions. A large family of tools, exemplified by PISA (7), *de facto* standard in PDB, EPPIC (8), ProtCID (9) and Prot-CAD (10), are aiming at identification of biologically relevant protein–protein interfaces and/or biological assemblies from crystal structures. However, still there is no foolproof method against occasional erroneous assignment of biologically relevant assemblies or interfaces (11). Many other tools are dedicated to annotate and classify PDB interaction data, and use these data to transfer 3D information to homologs or infer new interactions. They all differ greatly in user interface, the range of analyses performed and output data types. Some examples include 3did (12), which links PDB interaction data to Pfam domains, HOMCOS (13), which focuses on applying interaction data for searching and template-based modeling of homologous protein complexes, and DNAproDB (14) specializing in classification and annotation of protein-DNA complexes. There also are web servers that attempt to enrich protein–protein interaction networks with 3D structures including Interactome3D (15), LEVELNET (16) and Proteo3Dnet (17).

Here, we present the updated PPI3D server, which provides a possibility to search through a non-redundant set of pairwise interactions derived from an up-to-date set of PDB biological assemblies and to analyze the obtained results in detail. In addition, PPI3D users can make homology-based inferences regarding interaction sites of their query proteins and construct template-based models. PPI3D stands out among other similar tools by the use Voronoi tessellation to derive and analyze interaction interfaces. One of the strengths of this approach is that it unambiguously defines the contribution of each residue-residue contact to the interaction interface. Furthermore, representation of contacts via contact surface areas in PPI3D enables robust structure-based clustering of interfaces and binding sites. This step is important in detecting alternative interactions, that would be lost in clustering based only on sequence similarity. Compared with the initial version of PPI3D (18), we introduced two major improvements: (i) extended the PPI3D functionality into protein–nucleic acid interactions and (ii) provided a possibility to download customizable sets of interaction interfaces. These improvements in PPI3D open up new opportunities to study interactions for
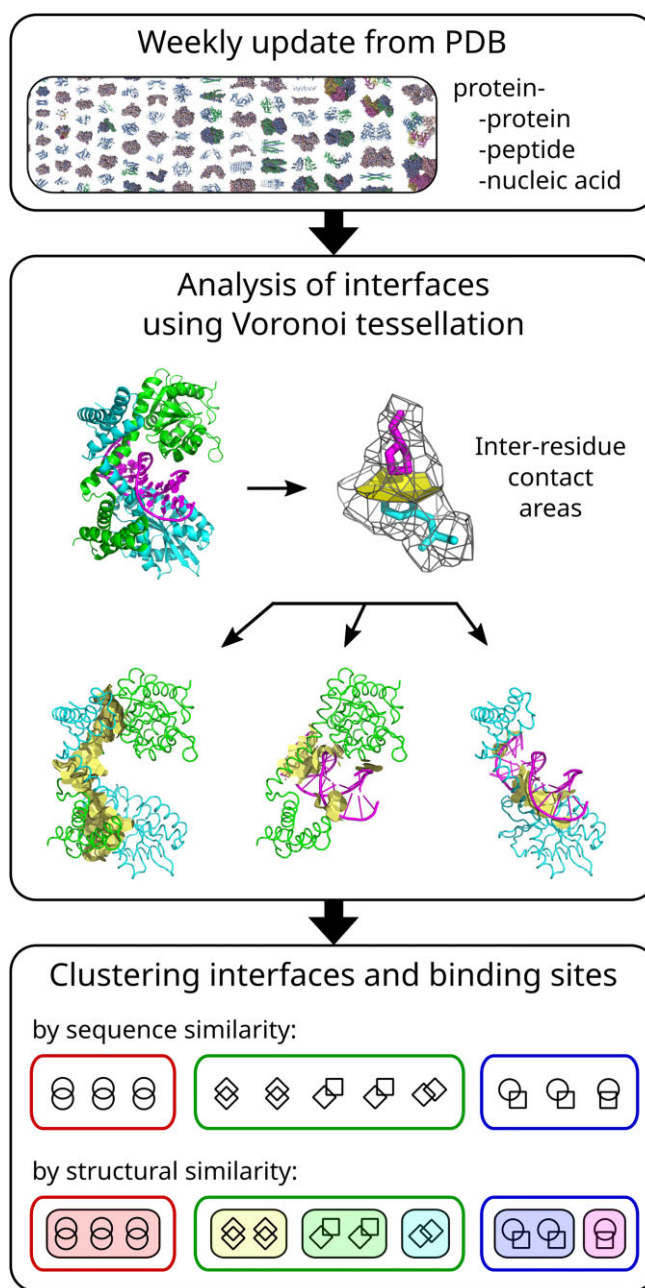
Figure 1. Structural data pre-processing in the PPI3D web server.

protein(s) of interest and to analyze interaction interfaces in bulk.

Materials and methods

Analysis of structural data and definition of interaction interfaces

Main steps in data pre-processing by PPI3D are shown in Figure 1. The Biological Assemblies for all non-NMR structures having resolution better than 4 Å are downloaded from the PDB (6). Polypeptide chains are classified into proteins and peptides. Peptides are defined as polypeptide chains with <20 structurally-resolved residues or <40 structurally-resolved residues if more than half represent non-standard amino acids. Biological assemblies containing nucleic acids are analyzed

using DSSR ([19]), and chains that form double-stranded DNA or RNA helices are joined into a single nucleic acid entity. Next, binary protein–protein, protein–peptide and protein–nucleic acid interactions are identified and analyzed by means of Voronoi tessellation, implemented in Voronota ([20]). Voronoi tessellation is a space-partitioning method. When applied to molecular structures, it assigns every atom a region of space, called Voronoi cell, which encompasses all the space points that are closer to that atom than to any other atom. Adjacent Voronoi cells share a surface, called Voronoi face, which can be viewed as geometric representation of a contact between two atoms. Contacts between atoms can be aggregated into contacts between residues (see Supplementary Data and Supplementary Figure S1 for details). The interaction interface is defined as the set of contacts between residues from different chains. The binding site is defined as the set of protein residues involved in the interaction with another entity (protein, peptide or nucleic acid). Only interfaces with the surface area over 100 Å^2 and only unique interfaces within each PDB entry are retained for further analysis. Next, hydrogen bonds ([21]), disulfide bonds and salt bridges are assigned for each interface. The distributions of interface areas, numbers of inter-chain contacts and inter-chain hydrogen bonds for the interfaces in the PPI3D database are provided in Supplementary Figures S2-S4.

Clustering of interaction interfaces and binding sites

Protein-protein interaction interfaces and protein–protein, protein–peptide, protein–nucleic acid binding sites are clustered based on protein sequence and interaction interface/binding site similarity. Initially, the protein sequences are clustered with CD-HIT ([22]) and for every cluster a multiple sequence alignment (MSA) is generated using L-INS-i, an accuracy-oriented MAFFT algorithm ([23]). MSAs are needed at a later stage to assign corresponding residues in different proteins for structure-based clustering.

Next, the protein–protein interaction interfaces are clustered by grouping interfaces where both proteins are in the same respective CD-HIT clusters. Protein binding sites are grouped by assigning the proteins of the same CD-HIT cluster to a single cluster of binding sites. These sequence-based clusters are further analyzed at the structural level by considering the interaction interface or binding site similarity, defined by the variants of CAD-score ([24],[25]). Definition of CAD-score variants is provided in Supplementary information. To compute CAD-score values, the one-to-one correspondence between the residues in different proteins is required. Here, residues are considered to be equivalent if they are in the same column of the MSA representing a CD-HIT cluster. In case CAD-score indicates that the sequence-based clusters display structural heterogeneity, they are further split into structurally similar clusters. This sequence- and structure-based clustering procedure ensures identification of different binding sites for a given protein and/or alternative protein–protein interaction modes. Clustering of structurally similar interfaces and binding sites is done using the Taylor-Butina algorithm ([26]). The algorithm uses a matrix of all pairwise similarities between objects to group those objects into clusters using a provided similarity threshold (see Supplementary data for details). One of the advantages, offered by this algorithm, is a possibility to easily update clusters.

PPI3D offers the following interface clusters: (i) identical or nearly identical interface, typically representing multiple instances of the same interacting proteins or their point mutants (sequence similarity > 95%, similarity of interface contacts > 50%), (ii) highly similar interfaces usually derived from homologous protein complexes (sequence similarity > 40%, similarity of interface contacts > 50%), and (iii) similar interfaces that correspond to similar surface patches, but tolerate rearrangement of residue-residue contacts across the interface (sequence similarity > 40%, similarity of interface areas > 50%). Likewise, protein binding sites are grouped into: (i) identical or nearly identical binding sites (sequence similarity > 95%, similarity of binding site residue areas > 50%), (ii) highly similar binding sites (sequence similarity > 40%, similarity of binding site residue areas > 50%) and (iii) similar binding sites (sequence similarity > 40%, similarity of binding site areas > 50%). Detailed statistics on data reduction upon clustering is provided in Supplementary Table S1 and Supplementary Figure S5.

Web server implementation

All the pre-processed and clustered structural data on diverse protein interactions are saved in a MySQL database. Newly released PDB entries are analyzed using the same pipeline and added to the database every week to keep in sync with the newest experimental data.

The PPI3D web server was developed using the CodeIgniter framework (https://www.codeigniter.com/). The interactive features were implemented using jQuery (https://jquery.com/). Structures are visualized interactively in a web browser using JSmol (http://jsmol.sourceforge.net/). For offline visualization, PyMOL scripts are provided. BLAST applications are used from the BLAST + package ([27]). Homology modeling is done by MODELLER ([28]). Structure alignments are generated using TM-align ([29]).

PPI3D web server description

Web server workflow

Input

A typical workflow of the PPI3D web server is illustrated in Figure [2]. The input into the server are protein sequences, UniProt accession codes, or PDB IDs. In the latter case, the server retrieves and displays all the binary interactions in a single PDB entry. Protein sequences (or corresponding UniProt codes) are used to find structural data on interactions of the query proteins and/or their homologs. Sequence-based search has two modes: (i) 'single-sequence' search to query interactions for a given protein with any proteins, peptides and nucleic acids and (ii) 'two-sequences' search to identify only protein–protein interactions between the first and the second proteins or their homologs.

Query processing

When the user inputs the query sequence(s), the PPI3D server searches in its database of protein sequences associated with structural interaction data using either BLAST or PSI-BLAST ([27],[30]). The PPI3D job with the BLAST option runs very fast, because the search is performed directly in the PPI3D database of protein sequences, but detects only close homologs. If the PSI-BLAST option is chosen, the server first generates a Position-Specific Scoring Matrix (PSSM) for the
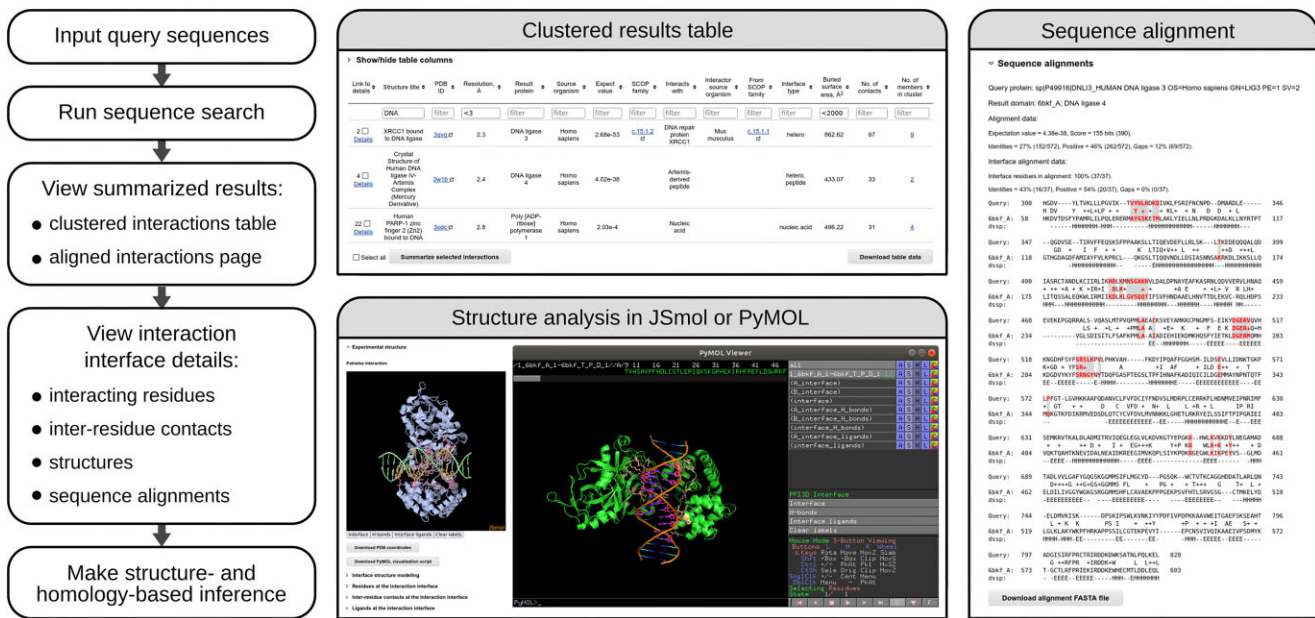
Figure 2. PPI3D search and analysis workflow with examples of output features.

query sequence by searching iteratively in a local copy of the NCBI non-redundant sequence database (https://ftp.ncbi.nlm.nih.gov/blast/db/), and then uses the resulting PSSM to search the PPI3D database. In this case a search takes more time to complete, but more distant homologs can be detected.

Output

The results on the interaction data are provided in a hierarchical manner. Initially, only the summarized information is shown, simply displaying how many pairwise interactions of different types (protein–protein, protein–peptide, protein–nucleic acid) are found for each of the query proteins (or protein pairs). By default, the output represents the most stringent clustering option, but the clustering stringency can be interactively adjusted. Next, the user might choose to analyze the lists of identified interactions.

The lists of identified interaction interfaces or binding sites are displayed in a table, containing protein annotations, calculated interface properties, BLAST *E*-values (if search by protein sequences was used), and cluster sizes. Sometimes the redundancy reduction by clustering might be insufficient, therefore, PPI3D allows selecting multiple results and summarizing them. This is done by displaying the alignment of selected results to the query sequence. The interacting residues are highlighted in the resulting multiple sequence alignment, making it easy to see whether there is a similar pattern of interacting residues at least in some of the interfaces or binding sites. To further inspect the similarities or differences between the search results, one can align selected structures with TM-align (29) and then visualize them in JSmol.

For even more in-depth analysis, PPI3D provides a very detailed page for every search result, showing the interaction properties, the structures of the binary interaction alone and in the context of the entire biological assembly, and tables listing interface residues and inter-residue contacts. The interacting residues for the user's query proteins can be inferred from the displayed highlighted sequence alignments or from generated homology models (31). The structures can be inspected using JSmol directly in the browser or could be downloaded as scripts for visualization in PyMOL. The users, interested in the analysis of interactions not only at the residue, but also at the atomic level, may choose to automatically transfer the structures to the VoroContacts server (32). VoroContacts makes it possible to analyze not only the entire interface, but also user-defined subsets of interface contacts that may be further filtered by various attributes.

Use case examples

In this section we provide several examples, illustrating how PPI3D could be used to search and analyze diverse protein interactions.

Exploring bacterial DNA sliding clamp interactions

DNA sliding clamp functions by encircling the DNA helix and serving as a mobile platform, to which various proteins involved in DNA transactions can bind. To identify interactions that the clamp participates in, we used the PPI3D 'single-sequence' search mode. A BLAST search with *E. coli* DNA sliding clamp (Uniprot AC: P0A988) revealed a large number of interactions, involving proteins, peptides and DNA (Supplementary Table S2). Most homomeric protein–protein interactions fall into two large clusters representing *E. coli* and *M. tuberculosis* proteins. Summarizing clamp binding sites both at the sequence level and by superimposing cluster representatives revealed that they all bind another subunit to form a closed ring (Figure 3A). An exception is the DNA sliding clamp from *Elizabethkingia anopheles* (PDB: 8DT6), which has alternative interfaces in addition to the consensus interface (Figure 3B). The PPI3D data indicate that the alternative interfaces, resulting from two stacked rings, are outliers. Surprisingly, both PDBePISA (7) and EPPIC (8) consider the two stacked rings to represent a biological assembly. However, in such assembly the central cavity of each ring is blocked and the DNA cannot be threaded through the sliding clamp suggesting that this type of arrangement is the result of crystal packing rather than a biologically relevant structure.
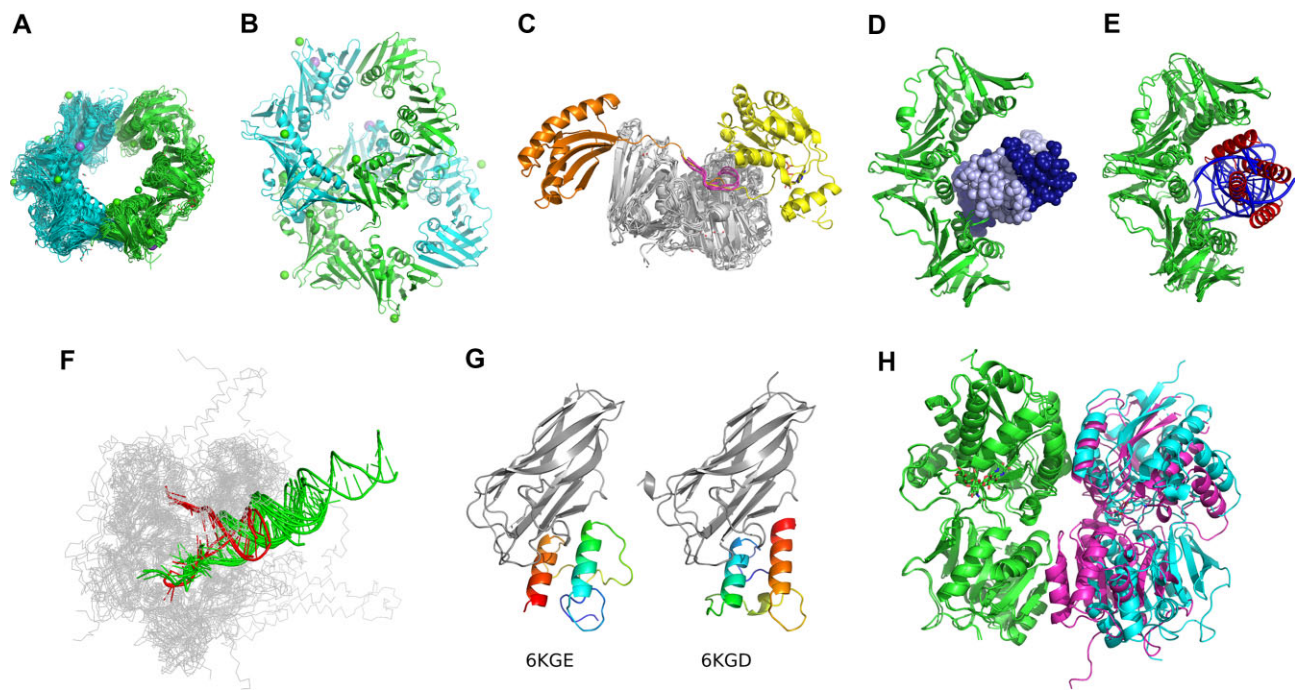
Figure 3. Examples of analyses using PPI3D. (**A**) Homodimers of DNA sliding clamp homologs listed in Supplementary Table S2; (**B**) two stacked rings of sliding clamps in PDB entry 8DT6; (**C**) heteromeric interactions of DNA sliding clamps with protein linear motifs or peptides (sliding clamp, gray; DNA polymerase IV (PDB: 1UNN), orange; DnaA regulatory inactivator Hda (PDB: 5X06), yellow; 7 peptides, magenta); (**D**) DNA (dark and light blue) binds differently to sliding clamp subunits (green) (PDB: 3BEP); (**E**) viral replication inhibitors (red, PDB: 7EVP) bind to the same site of sliding clamp as DNA (blue, PDB: 3BEP); (**F**) DNA polymerases (grey) show distinct DNA binding modes in the polymerization (green) and proofreading (red) modes; (**G**) dual binding modes of cohesin-dockerin interaction; dockerin (colored) bound to cohesin (gray) is flipped by 180 degrees in different modes; (**H**) extracellular domains of human GABA$_B$ receptor heterodimer in active (magenta and green, PDB: 4MS3) and inactive (cyan and green, PDB: 4MR8) forms; heterodimers are superimposed by aligning subunit 2 (green).

Over 100 detected protein–peptide binding sites within sliding clamps can be grouped into just 7 clusters. Structure superposition of the DNA sliding clamp subunits further demonstrates that all these binding sites correspond to the same structural region. Proteins also tend to bind to the same region using linear motifs (Figure 3C). This is indeed expected as it is known that different proteins bind to the DNA sliding clamp within a structurally conserved pocket (33,34).

There is only one PDB entry that has a clamp bound to DNA (PDB: 3BEP) (35). PPI3D data show that DNA is bound asymmetrically to the clamp (Figure 3D). The detailed binding site data reveal that R24, one of the two residues important for the clamp function (35), binds DNA in both subunits, whereas Q149 only binds DNA in one of the subunits. Interestingly, viral protein Gp168 that inhibits bacterial DNA replication also binds to the same site of the sliding clamp (Figure 3E) (36).

Detecting different DNA binding modes in DNA polymerases
DNA polymerases are essential enzymes that catalyze DNA synthesis during replication and reparation. The most widespread B-family DNA polymerases have DNA polymerase and 3′–5′ exonuclease activities that are located in different domains. To investigate how these enzymes interact with DNA, we queried PPI3D with the sequence of an unexplored DNA polymerase from *Halorubrum halophilum* (RefSeq: WP_050032690.1) using PSI-BLAST. The server found 271 protein–nucleic acid binding sites, grouped into 33 clusters, in polymerases from eukaryotes, archaea, bacteria, and viruses. We selected 11 representative binding sites with

largest surface areas (>1800 Å^2) (Supplementary Table S3). Using the 'Summarize selected interactions' feature we aligned their structures on the *Thermococcus sp.* DNA polymerase solved in the replicative state (PDB: 5OMV) (37). Superposition revealed that polymerases bind DNA in two different modes, corresponding to DNA synthesis and proofreading (Figure 3F) (38,39).

Detecting alternative protein–protein interactions
Among the vast diversity of protein–protein interactions, alternative binding modes are occasionally observed (40). Therefore, searches for interactions in databases clustered only by protein sequences might miss some of the interfaces. Since PPI3D clusters the interaction interfaces not only by sequence, but also by structure, it allows identification of the alternative binding modes.

One of the well-known dual binding protein pairs is cohesion and dockerin, domains found in a cellulosome, an enzymatic complex of anaerobic cellulolytic microorganisms (Figure 3G). The biological significance of this dual binding mode is still unknown (41), but it was discovered that it can be regulated by pH (42). After a PSI-BLAST search in PPI3D using the sequences of PDB entry 6KGE, we found 32 structures that can be clustered differently. A more stringent clustering (sequence identity 40%, similarity of interface residue contacts $> 50\%$) produced 17 clusters, whereas a more lenient clustering that disregards specific residue-residue contacts (sequence identity 40%, similarity of interface areas $> 50\%$) produced only 12 clusters (Supplementary Table S4).

Changes of protein binding modes can also occur upon ligand binding. For example, the agonist binding causes large conformational changes in heterodimeric human GABA$_B$ receptor, inducing formation of additional inter-subunit contacts and doubling the total interface area (from ∼700 to ∼1400 Å^2) (Figure 3H) (43). PPI3D clustering recognized these two interfaces as distinct clusters.

Downloading of the PPI3D data

The PPI3D user interface offers interactive analysis of diverse protein interactions. All the tables can be sorted by different properties and filtered using text, regular expressions or numerical values. The structures can be visualized in JSmol. Yet, in some cases it may be more convenient to analyze the data offline. Therefore, the data displayed in the PPI3D web server including tables, structures, and sequence alignments can be downloaded for local use.

In addition to the analysis of interaction data for specific protein(s), PPI3D also provides a possibility to download other user-defined subsets of the clustered structural protein interaction data in bulk. The users can select the PPI3D data subsets according to different criteria and download the data in tabular format as well as coordinate files. These data sets might be useful for detailed investigation of protein interactions at scale or for training machine learning models.

Discussion

PPI3D web server offers a user-friendly environment for searching and analyzing structure-resolved protein-centered interactions. PPI3D may be especially helpful if no interaction data are available for the protein(s) of interest. Sequence searches in the 'single-sequence' mode may help to infer putative interaction partners based on the identified structure-resolved homologs bound to other proteins, peptides or nucleic acids. Likewise, the identified interactions in the 'two-sequences' search mode may suggest that two query proteins interact. In both cases, these initial hypotheses can be further explored at the residue level using both sequence alignments with the detected structural homologs and template-based models.

The server has already proved useful in both experimental (44,45) and computational studies (46,47). PPI3D also helped our group to achieve top results in the protein assemblies modeling category in recent CASP and CAPRI experiments (48,49). In the AlphaFold era template-based modeling is becoming less important (50), but the ability to quickly test hypotheses on whether specific residues might be involved in binding with the help of homology models and to survey the broader structural context for the query protein(s) remains very useful. In contrast to protein–protein complexes, the accurate structure prediction of protein–nucleic acid complexes is still largely refractory. Therefore, the ability of PPI3D to provide homology-based inferences related to protein-DNA or protein-RNA interactions is highly relevant.

Within the ecosystem of tools dedicated to the analysis of structure-resolved data on biomolecular interactions (7–10,12–17), PPI3D features a unique set of capabilities. PPI3D uses precomputed non-redundant structural data that are updated weekly to keep in sync with PDB. In contrast, most other servers, except for those directly associated with PDB (7,13), are usually based on PDB data that are several months or even

several years old. PPI3D offers sequence-based searches that can be tuned to detect either only close or also distant interacting homologs. The user interface allows interactive analysis of diverse interactions for the proteins of interest within the common framework both at the sequence and structure levels. The analysis may range from the most general data regarding the identified interfaces/binding sites down to the properties of individual residue-residue contacts. To the best of our knowledge, PPI3D is the only server that uses rigorous Voronoi tessellation-based methodology for the clustering and analysis of interactions. A newly introduced option to download all the data on interaction interfaces might be useful for large-scale analyses. It can also be beneficial for providing well-defined up-to-date datasets for training and testing machine learning methods for predicting structures or properties of macromolecular complexes. The information about how interaction interfaces are clustered on both sequence and structure levels may be especially important for defining training/validation/testing data splits. To conclude, the PPI3D web server might be useful for both experimental and computational research involving protein interactions.

Data availability

The PPI3D server and downloadable data on protein interaction interfaces are freely available at https://bioinformatics.lt/ppi3d.

Supplementary data

Supplementary Data are available at NAR Online.

Conflict of interest statement

None declared.

References

1. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
2. Varadi,M., Bertoni,D., Magana,P., Paramval,U., Pidruchna,I., Radhakrishnan,M., Tsenkov,M., Nair,S., Mirdita,M., Yeo,J., *et al.* (2024) AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.*, **52**, D368–D375.
3. Kryshtafovych,A., Schwede,T., Topf,M., Fidelis,K. and Moult,J. (2023) Critical assessment of methods of protein structure prediction (CASP)-Round XV. *Proteins*, **91**, 1539–1549.
4. Ozden,B., Kryshtafovych,A. and Karaca,E. (2023) The impact of AI-based modeling on the accuracy of protein assembly prediction: insights from CASP15. *Proteins*, **91**, 1636–1657.
5. Baek,M., McHugh,R., Anishchenko,I., Jiang,H., Baker,D. and DiMaio,F. (2024) Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods*, **21**, 117–121.
6. Burley,S.K., Berman,H.M., Kleywegt,G.J., Markley,J.L., Nakamura,H. and Velankar,S. (2017) Protein Data Bank (PDB):

the single global macromolecular structure archive. *Methods Mol. Biol.*, **1607**, 627–641.

7. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.

8. Bliven,S., Lafita,A., Parker,A., Capitani,G. and Duarte,J.M. (2018) Automated evaluation of quaternary structures from protein crystals. *PLoS Comput. Biol.*, **14**, e1006104.

9. Xu,Q. and Dunbrack,R.L. Jr (2020) ProtCID: a data resource for structural information on protein interactions. *Nat. Commun.*, **11**, 711.

10. Xu,Q. and Dunbrack,R.L. (2023) The protein common assembly database (ProtCAD) - a comprehensive structural resource of protein complexes. *Nucleic Acids Res.*, **51**, D466–D478.

11. Xu,Q. and Dunbrack,R.L. Jr (2019) Principles and characteristics of biological assemblies in experimentally determined protein structures. *Curr. Opin. Struct. Biol.*, **55**, 34–49.

12. Mosca,R., Ceol,A., Stein,A., Olivella,R. and Aloy,P. (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **42**, D374–D379.

13. Kawabata,T. (2016) HOMCOS: an updated server to search and model complex 3D structures. *J. Struct. Funct. Genomics*, **17**, 83–99.

14. Sagendorf,J.M., Markarian,N., Berman,H.M. and Rohs,R. (2020) DNAproDB: an expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.*, **48**, D277–D287.

15. Mosca,R., Ceol,A. and Aloy,P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.

16. Mohseni Behbahani,Y., Saighi,P., Corsi,F., Laine,E. and Carbone,A. (2023) LEVELNET to visualize, explore, and compare protein–protein interaction networks. *Proteomics*, **23**, e2200159.

17. Postic,G., Andreani,J., Marcoux,J., Reys,V., Guerois,R., Rey,J., Mouton-Barbosa,E., Vandenbrouck,Y., Cianferani,S., Burlet-Schiltz,O., *et al.* (2021) Proteo3Dnet: a web server for the integration of structural information with interactomics data. *Nucleic Acids Res.*, **49**, W567–W572.

18. Dapkūnas,J., Timinskas,A., Olechnovič,K., Margelevičius,M., Dičiūnas,R. and Venclovas,Č. (2017) The PPI3D web server for searching, analyzing and modeling protein–protein interactions in the context of 3D structures. *Bioinformatics*, **33**, 935–937.

19. Lu,X.J., Bussemaker,H.J. and Olson,W.K. (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.*, **43**, e142.

20. 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.

21. McDonald,I.K. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.

22. 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.

23. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.

24. Olechnovič,K., Kulberkytė,E. and Venclovas,Č. (2013) CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins*, **81**, 149–162.

25. Olechnovič,K. and Venclovas,Č. (2020) Contact area-based structural analysis of proteins and their complexes using CAD-score. *Methods Mol. Biol.*, **2112**, 75–90.

26. Butina,D. (1999) Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comp. Sci.*, **39**, 747–750.

27. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.

28. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

29. Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.

30. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

31. Dapkūnas,J. and Venclovas,Č. (2020) Template-based modeling of protein complexes using the PPI3D web server. *Methods Mol. Biol.*, **2165**, 139–155.

32. Olechnovič,K. and Venclovas,Č. (2021) VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures. *Bioinformatics*, **37**, 4873–4875.

33. Dalrymple,B.P., Kongsuwan,K., Wijffels,G., Dixon,N.E. and Jennings,P.A. (2001) A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 11627–11632.

34. Wijffels,G., Johnson,W.M., Oakley,A.J., Turner,K., Epa,V.C., Briscoe,S.J., Polley,M., Liepa,A.J., Hofmann,A., Buchardt,J., *et al.* (2011) Binding inhibitors of the bacterial sliding clamp by design. *J. Med. Chem.*, **54**, 4831–4838.

35. Georgescu,R.E., Kim,S.S., Yurieva,O., Kuriyan,J., Kong,X.P. and O'Donnell,M. (2008) Structure of a sliding clamp on DNA. *Cell*, **132**, 43–54.

36. Liu,B., Li,S., Liu,Y., Chen,H., Hu,Z., Wang,Z., Zhao,Y., Zhang,L., Ma,B., Wang,H., *et al.* (2021) Bacteriophage twort protein Gp168 is a beta-clamp inhibitor by occupying the DNA sliding channel. *Nucleic Acids Res.*, **49**, 11367–11378.

37. Kropp,H.M., Betz,K., Wirth,J., Diederichs,K. and Marx,A. (2017) Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *PLoS One*, **12**, e0188005.

38. Killelea,T., Ghosh,S., Tan,S.S., Heslop,P., Firbank,S.J., Kool,E.T. and Connolly,B.A. (2010) Probing the interaction of archaeal DNA polymerases with deaminated bases using X-ray crystallography and non-hydrogen bonding isosteric base analogues. *Biochemistry*, **49**, 5772–5781.

39. Gouge,J., Ralec,C., Henneke,G. and Delarue,M. (2012) Molecular recognition of canonical and deaminated bases by P. abyssi family B DNA polymerase. *J. Mol. Biol.*, **423**, 315–336.

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

41. Bule,P., Pires,V.M., Fontes,C.M. and Alves,V.D. (2018) Cellulosome assembly: paradigms are meant to be broken! *Curr. Opin. Struct. Biol.*, **49**, 154–161.

42. Yao,X., Chen,C., Wang,Y., Dong,S., Liu,Y.J., Li,Y., Cui,Z., Gong,W., Perrett,S., Yao,L., *et al.* (2020) Discovery and mechanism of a pH-dependent dual-binding-site switch in the interaction of a pair of protein modules. *Sci. Adv.*, **6**, eabd7182.

43. Geng,Y., Bush,M., Mosyak,L., Wang,F. and Fan,Q.R. (2013) Structural mechanism of ligand activation in human GABA(B) receptor. *Nature*, **504**, 254–259.

44. Manoharan,J., Rana,R., Kuenze,G., Gupta,D., Elwakiel,A., Ambreen,S., Wang,H., Banerjee,K., Zimmermann,S., Singh,K., *et al.* (2024) Tissue factor binds to and inhibits interferon-alpha receptor 1 signaling. *Immunity*, **57**, 68–85.

45. Laan,M., Kasak,L., Timinskas,K., Grigorova,M., Venclovas,Č., Renaux,A., Lenaerts,T. and Punab,M. (2021) NR5A1 c.991-1G >C splice-site variant causes familial 46,XY partial gonadal dysgenesis with incomplete penetrance. *Clin. Endocrinol. (Oxf)*, **94**, 656–666.

46. Bret,H., Gao,J., Zea,D.J., Andreani,J. and Guerois,R. (2024) From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nat. Commun.*, **15**, 597.

47. Farajzadeh-Dehkordi,M., Mafakher,L., Harifi,A., Haghdoost-Yazdi,H., Piri,H. and Rahmani,B. (2024) Unraveling the function and structure impact of deleterious missense SNPs in the human OX1R receptor by computational analysis. *Sci. Rep.*, **14**, 833.

48. Dapkūnas,J., Olechnovič,K. and Venclovas,Č. (2021) Modeling of protein complexes in CASP14 with emphasis on the interaction interface prediction. *Proteins*, **89**, 1834–1843.

49. Lensink,M.F., Brysbaert,G., Mauri,T., Nadzirin,N., Velankar,S., Chaleil,R.A.G., Clarence,T., Bates,P.A., Kong,R., Liu,B., *et al.* (2021) Prediction of protein assemblies, the next frontier: the CASP14-CAPRI experiment. *Proteins*, **89**, 1800–1823.

50. Olechnovič,K., Valančauskas,L., Dapkūnas,J. and Venclovas,Č. (2023) Prediction of protein assemblies by structure sampling followed by interface-focused scoring. *Proteins*, **91**, 1724–1733.