

Proteo3Dnet: a web server for the integration of structural information with interactomics data

Guillaume Postic, Jessica Andreani, Julien Marcoux, Victor Reys, Raphaël Guerois, Julien Rey, Emmanuelle Mouton-Barbosa, Yves Vandenbrouck, Sarah Cianferani, Odile Burlet-Schiltz, et al.

▶ **To cite this version:**

Guillaume Postic, Jessica Andreani, Julien Marcoux, Victor Reys, Raphaël Guerois, et al.. Proteo3Dnet: a web server for the integration of structural information with interactomics data. Nucleic Acids Research, 2021, 49 (W1), pp.W567-W572. 10.1093/nar/gkab332 . hal-03374264

HAL Id: hal-03374264
https://hal.science/hal-03374264

Submitted on 12 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published online 8 May 2021 *Nucleic Acids Research, 2021, Vol. 49, Web Server issue* ***W567–W572***
https://doi.org/10.1093/nar/gkab332

Proteo3Dnet: a web server for the integration of structural information with interactomics data

Guillaume Postic[1,2], Jessica Andreani[3], Julien Marcoux[4], Victor Reys[5], Raphaël Guerois[3], Julien Rey[1], Emmanuelle Mouton-Barbosa[4], Yves Vandenbrouck[6], Sarah Cianferani[7], Odile Burlet-Schiltz[4], Gilles Labesse[5,*] and Pierre Tufféry[1,*]

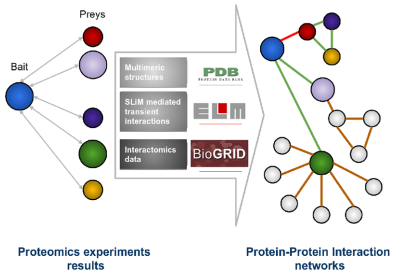
[1]Université de Paris, CNRS UMR 8251, INSERM U1133, RPBS, Paris 75205, France, [2]Institut Français de Bioinformatique (IFB), UMS 3601-CNRS, Université Paris-Saclay, Orsay, France, [3]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France, [4]Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, UPS, Toulouse, France, [5]Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France, [6]Univ. Grenoble Alpes, INSERM, CEA, UMR BioSanté U1292, CNRS, CEA, FR2048, 38000 Grenoble, France and [7]Laboratoire de Spectrométrie de Masse BioOrganique, Université de Strasbourg, CNRS, IPHC UMR 7178, 67000 Strasbourg, France

Received March 16, 2021; Revised April 09, 2021; Editorial Decision April 17, 2021; Accepted April 19, 2021

ABSTRACT

Proteo3Dnet is a web server dedicated to the analysis of mass spectrometry interactomics experiments. Given a flat list of proteins, its aim is to organize it in terms of structural interactions to provide a clearer overview of the data. This is achieved using three means: (i) the search for interologs with resolved structure available in the protein data bank, including cross-species remote homology search, (ii) the search for possibly weaker interactions mediated through Short Linear Motifs as predicted by ELM—a unique feature of Proteo3Dnet, (iii) the search for protein–protein interactions physically validated in the BioGRID database. The server then compiles this information and returns a graph of the identified interactions and details about the different searches. The graph can be interactively explored to understand the way the core complexes identified could interact. It can also suggest undetected partners to the experimentalists, or specific cases of conditionally exclusive binding. The interest of Proteo3Dnet, previously demonstrated for the difficult cases of the proteasome and pragmin complexes data is, here, illustrated in the context of yeast precursors to the small ribosomal subunits and the smaller interactome of 14–3–3zeta frequent interactors. The Proteo3Dnet web server is accessible at http://bioserv.rpbs.univ-paris-diderot.fr/services/Proteo3Dnet/.

GRAPHICAL ABSTRACT



INTRODUCTION

MS-based proteomics techniques such as combination of affinity purification and mass spectrometry (AP-MS) (1,2), aim to identify sets of proteins that interact to fulfill cellular functions. In such an experiment, a protein of interest (the 'bait') is co-precipitated along with its bound partners (the 'preys'). After elution, those interacting proteins, as well as their relative amounts, are determined by liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS). The resulting list of identified proteins thus defines the interactome of the bait. In this process, the main drawback is the loss of information about the network organization of protein–protein interactions (PPIs). Thus, proteomics experiments generate 'flat' lists of candidate partners as an output. To a lesser extent, information loss may also take the form of false negatives, i.e. proteins of the targeted interactome that are found missing in the experimental results. Misleading hits may also occur, for example when two mutually exclusive subunits of a multiprotein complex are both included in the output list.

*To whom correspondence should be addressed. Tel: +33 1 57 27 83 74; Fax: +33 1 1 57 27 83 72; Email: pierre.tuffery@univ-paris-diderot.fr
Correspondence may also be addressed to Gilles Labesse. Tel: +33 4 67 41 77 12; Fax: +33 4 67 41 79 13; Email: labesse@cbs.cnrs.fr

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021

Large-scale methods do not inform either about the binding strength in PPIs, as permanent and transient complexes are indistinct in the flat lists. Knowing the types of the amino acid residues located at the protein–protein interface would also be particularly valuable, regarding applications such as thorough validation, protein engineering or drug design. Unfortunately, proteomics data do not provide atomic-level knowledge about the molecular mechanisms through which proteins interact. Finally, besides proteomics, it should be noted here that such flat lists may be produced by computational studies, e.g. when considering proteins whose genes are located within the same genomic region as candidate partners of a multimeric complex.

Establishing connectivity between the different proteins of a proteomics-derived list can hardly be carried out manually, as it requires processing large amounts of data from various sources. With this regard, three web servers are currently available to assist biologists in this task. The STRING database (https://string-db.org/) (3), one of the most cited resources for building PPI networks, is a precomputed global resource including both physical interactions as well as functional associations. It uses multiple types of information, such as genetic interactions, text mining, or experimentally and predicted PPIs from other databases. The GeneMANIA web server (http://genemania.org) (4) uses lists of gene ID as input to perform functional annotation and prediction, by integrating data from PPI databases such as BioGRID (https://thebiogrid.org/) (5) or PathwayCommons (www.pathwaycommons.org) (6). Lastly, Interactome3D (https://interactome3d.irbbarcelona.org/) (7), introduced the use of protein three-dimensional (3D) structures for modeling the protein complexes of a given organism and, thus, may shed light on the protein residues involved in a given interaction. It also includes information retrieved from external resources, such as IntAct (https://www.ebi.ac.uk/intact/) (8). These three web resources are capable of finding interologs, i.e. complex-forming partners predicted on the basis of the interaction between their respective homologous proteins, providing that this interaction has been conserved throughout evolution. This task is performed either through identical gene names across different species or, in the case of Interactome3D, by homology search using BLAST (9). Finally, other useful tools for studying protein complexes can be cited, such as PrePPI (10), ProtCID (11), ComplexPortal (12), CORUM (13), and the XlinkDB database and software tools, specifically dedicated to cross-linking data analysis (14). However, these cannot be compared to the aforementioned servers, as most fall in the database category and focus on different aspects of PPIs. PPI3D (15) is another tool to search for interologs with a 3D interface modeling perspective, but is presently limited to only two sequences.

Recently, we have proposed Proteo3Dnet (16), a new integrative pipeline which stands out from the state of the art mainly by its detection of distant homologies and combination of domain-domain interactions as well as motif-domain contacts. We have shown that our method can identify some direct interaction partners to which proteomics techniques and other algorithms are blind. A unique feature of Proteo3Dnet is the integration of information about transient interactions involving Pfam (17) domains and short linear motifs (SLiMs) located in intrinsically disordered regions. Here, we present a web implementation of Proteo3Dnet, which allows for rapid detection of PPI networks from an input list of proteins, and their online exploration.

Data integration

Given a set of protein sequences (bait and preys), Proteo3Dnet provides end-users with an integrative protocol aimed at validating and connecting candidate protein partners detected by interactomics experiments; the information is organized and structured using three means:

Search for complexes of resolved structure. The Proteo3Dnet core processing consists of 4 steps.

1. Search for homologs in the Protein Data Bank using HHsearch. We use the precalculated banks available from the HHsuite repository, in which a non-redundant subset of PDB entries at 70% sequence identity is maintained. To reduce the chance to identify irrelevant homologs, only hits associated with a probability of >95%, a cutoff, that in our experience fits well the objective of lowering the risk to miss some remote templates while identifying the relevant ones.
2. Proteo3Dnet then enlarges the collection of hits with the protein chains belonging to the same cluster. Because the initial search is performed on a subset of the PDB and because it could occur that non-representatives of the clusters correspond to PDB entries containing the structures of complexes of interest, this step extends the search to all homologs of the PDB.
3. Complex identification is based on the search for PDB entries for which several—if not all—chains of the input are homologous. To avoid incorrect complex assignment that could occur in the unit cell of the crystals, the search is performed considering the biological unit specified by the authors, or ranked one otherwise, as provided by the PDB entries in the mmCIF format. Information about homo-oligomers is also collected from these entries.
4. Complexes are then ranked according to the number of chains of the input they encompass, and conditional chain binding stable core complexes are identified.

We emphasize that this search procedure covers the entirety of the PDB and, thus, is not *per se* biased by the number of structures of complexes that are evolutionarily conserved, neither is it regarding species specificity. It simply collects and organizes all the structural information available.

Note that in the web server implementation, it is possible to bypass this step by directly querying the Swiss Model Repository (SMR) (18). However, since the SMR is limited to high enough confidence models, this facility usually leads to fewer hits than the Proteo3Dnet standard protocol. It can nevertheless provide a quick overview of the organization of data.

Search for weak/transient interactions. Previous processing corresponds to the identification of complexes at rather

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021

high affinity. It is however well known that weaker and more transient functional interactions can occur. Particularly in eukaryotes, these can be mediated by Short Linear Motifs (SLiMs) (19). These are usually located in disordered regions of proteins, often in missing parts of resolved structures. To overcome this limitation, Proteo3Dnet relies on the ELM database that links the motifs with interacting Pfam domains. Occurrences of pairs involving SLiMs-Pfam domains correspond to potential PPIs. The ELM database currently gathers 291 linear motifs ('ELM classes') which are reported to bind one or several Pfam domains, according to published experiments ('ELM instances'). To identify transient interactions, Proteo3Dnet searches for these motifs and their corresponding Pfam domains in the proteins of the input list. The resulting connections are then filtered, based on the location of the motifs within intrinsically disordered regions, as predicted by the ANCHOR2 score of IUPred2A (20), computed and averaged for all positions of the motif. The default value of 0.95 can be raised (up to 1.0) in the 'Advanced options', to increase the specificity of these predicted transient PPIs, thus reducing their number.

Additional sources of information. For cases where no 3D structure can be found, the pipeline integrates interaction data from the Biological General Repository for Interaction Data sets (BioGRID) (5). Only interactions characterized by physical (not genetic) interactions between proteins are retrieved, though we discarded those associated with the 'Far Western', 'Co-fractionation', 'Co-localization', 'Biochemical Activity' and 'High Through-put' experimental systems.

Implementation

Server input. The Proteo3Dnet input consists of a specification of the sequences identified during an interactome experiment. These can be specified either in the form of the UniProt identifiers or full UniProt sequences in the FASTA format. Note that presently, it is not possible to enter data without specifying the UniProt identifier.

Output. The outputs consist of three main sections, that encompass interactive graph exploration, tables reporting information about the hetero- and homo-multimers identified, along with raw information on the homologs identified, and for each input protein, a list of experimentally verified interactions from the IntAct database (8) —which may include additional protein partners—presented as an interactive PPIs matrix (https://github.com/ebi-uniprot/interaction-viewer/).

We focus here on the main output that corresponds to the interactive exploration of the groups of sequences identified as in interaction (Figure 1A). This facility is based on cytoscape.js, a JavaScript library for interactive network visualization (https://js.cytoscape.org/) (21).

The graph consists of nodes that correspond to proteins and edges that represent the interactions between them. Three types of nodes are distinguished in the graph and correspond to: (i) input proteins, (ii) the proteins not present in the input but in interaction with those in at least one complex structure, (iii) proteins with interaction suggested by BioGRID (see upper). In terms of AP-MS, the second type can correspond to either proteins that could be detected during the experiment but were for instance below the cut-off, or to proteins that interact conditionally with a 'core' complex.

Different levels of interaction are proposed:

1. The first one concerns the ability to simplify graph visualization and facilitate its exploration. Different mechanisms for node and edges selection and visualization are proposed. Visualization relies on three levels of visibility (hidden, background, foreground), that, coupled with selection makes it possible to simplify the visualization to obtain a simplified presentation of the information of interest. Selection encompasses interactive selection using the mouse, and higher level and more complex selections using text boxes to specify proteins (nodes), identified complexes or PDBs. Higher levels of selection include selection by edge type (structurally stable, weak and available in BioGRID), expansion to the neighbors of the visible nodes, or the identification of paths linking distant nodes in the graph. Selections can be assigned to each of the different levels of visibility. Additional mechanisms provide means to handle node/edge display and position in the display, e.g. using the mouse.
2. A second one concerns backpropagation to the information available concerning the interactions. Nodes and edges are right-clickable to open pop-up menus. From these menus, it is possible to rapidly get information about the sequences (UniProt), structures (PDB, MolArt (22)) (Figure 1B) and the raw information about the complexes detected (Figure 1C). Similarly, right click on the edges provides access to the PDB complexes in which the nodes are seen together or to ELM motif(s) connecting the two nodes.

Runtime. Runtime depends on the number of sequences and the complexity of the graph generated. It typically ranges from several minutes for small datasets of less than 20 sequences to up to several tens of minutes when the sequence number exceeds one hundred. For example, the 62 proteins of the Pragmin dataset (23) were processed in 15 minutes. For the larger proteasome 20S dataset ($n = 192$) (24), the calculations were completed in only 20 min. These times were obtained using the 'Normal' (i.e. slowest) mode of the server. To avoid very long runs, the maximum number of input proteins currently accepted is 400.

USE-CASES

Advantages of our protocol over the other approaches have been recently discussed through two case studies from published AP-MS data (16). Here we introduce two new case studies showing how Proteo3Dnet can assist users to analyse their data.

Use case of yeast precursors to the small ribosomal subunits

We have queried Proteo3Dnet with a list of interactants obtained by tandem-affinity purification of *Saccaromyces*

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021

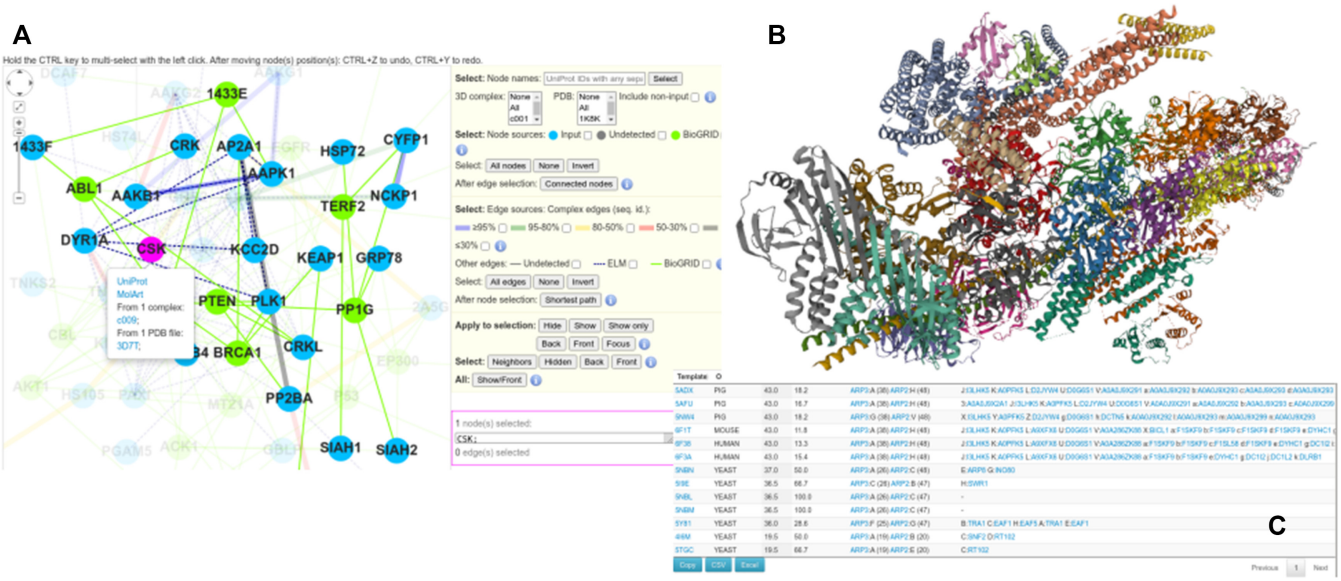


Figure 1. Proteo3Dnet results presentation. (**A**) The interaction graph can be explored interactively using a javascript viewer adapted from cytoscape.js. The top panel provides access to node identification, selection and display modes. Right clicking the nodes or the edges provides access to further information about the sequences, the structures of the complexes identified (**B**) and the details of the homology search leading to complex identification (**C**).

cerevisiae precursors to the small ribosomal subunits (hereafter termed pre-40S particles; (25). These 62 proteins include highly abundant 39 Ribosomal Proteins of the Small ribosomal subunit (RPS) and 8 ribosome biogenesis factors (RBFs) together with 14 much less abundant Ribosomal Proteins of the Large ribosomal subunit (RPL), and the Lrg1 protein.

Proteo3Dnet, identified templates for all the 62 proteins (Figure 2A). 56 of them were identified in *Saccharomyces cerevisiae*. There is currently no structure of yeast Dim1 and Nob1 RBFs and these were matched to their human (PDB: 1ZQ9) and archaeal (PDB: 2LCQ) counterparts. A human model (PDB: 3EAP) was also proposed for Lrg1. It is important to note that the presence of RNA molecules within ribosomal particles (or any other type of structuring molecule) is not an issue for our pipeline, since it only considers proteins.

Of the 145 complexes retrieved, ∼80 correspond to the full ribosome (either from yeast, as c006, or from other related species like c130 (human), c144 (*M. tuberculosis*) or c065 (pig). They do not correspond but rather include the pre-40S particles, with low maximal completeness—from 7% (c107) to 60% (c001). On the contrary, ∼60 complexes with maximal completeness ranging from 30% (c016) to 100% (c009) and corresponding to small ribosomal subunits were identified, either from yeast (as c009) or other species. The c009 actually corresponds to the *S. cerevisiae* pre-40S particle (PDB: 6FAI) and served as a positive control here (Figure 2B). It contains 35 (out of 62) entries with 100% sequence identity. The c027 is a subcomplex corresponding to the dimeric Enp1/Ltv1 RBFs (PDB: 5WWO). Finally, a couple of complexes (c095, c098…), matched large ribosomal subunits that were co-purified traces rather than true partners of the pre-40S particles.

Very similar results were obtained on this example using interactome3D although some discrepancies exist. For example, only one interaction is found for Rps6 (with Rps4A) in interactome3D, whereas Proteo3Dnet reports interactions with all the other 38 RPS. In interactome3D, the model of Rio2 is based on a distant similarity with an archeal protein (PDB: 4GYI) while the 3D structure from *S. cerevisiae* is known (PDB: 6RBD) (26). In addition, this server provides no model for Nob1 while Proteo3Dnet does, using the NMR structure of an archaeal Nob1 (PDB: 2LCQ). Hence, the two servers appeared rather complementary on this example due to small variations, although they both clearly matched the macromolecular assemblage of the yeast small ribosomal subunit.

Case study of 14-3–3zeta frequent interactors

BioGRID lists three partners (Raf1, Bad and Tau) that are detected >12 times in interaction with the human 14–3-3zeta protein. Complete or partial structures are available for all of them and we interrogated Interactome3D and Proteo3Dnet to highlight their likely interfaces. Domain–domain interactions are suggested by the two servers to derive complexes for 14–3–3zeta and Raf1 or 14–3–3zeta and Tau based on known crystal structures (at 100% and 87% of sequence identity, respectively). While Interactome3D pinpoints experimental interactions listed in Intact for 14–3–3zeta and Bad, no structural clue for those interactions is provided. Thanks to an ELM motif for 14–3–3 binding, a structural connection is provided by Proteo3Dnet (Figure 2C). Noteworthy, Bad contains up to 6 canonical 14–3-3 recognition motifs (ELM *E*-value ∼ 0.0045) among which four corresponds to serine phosphorylation sites according to PhosphoSitePlus (https://www.phosphosite.org/). In conclusion, all the detected 3D interactions with Raf1, Tau and Bad, correspond to recognition of similar phosphosites by the 14–3-3zeta proteins and this suggests competition for the same binding groove.

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021

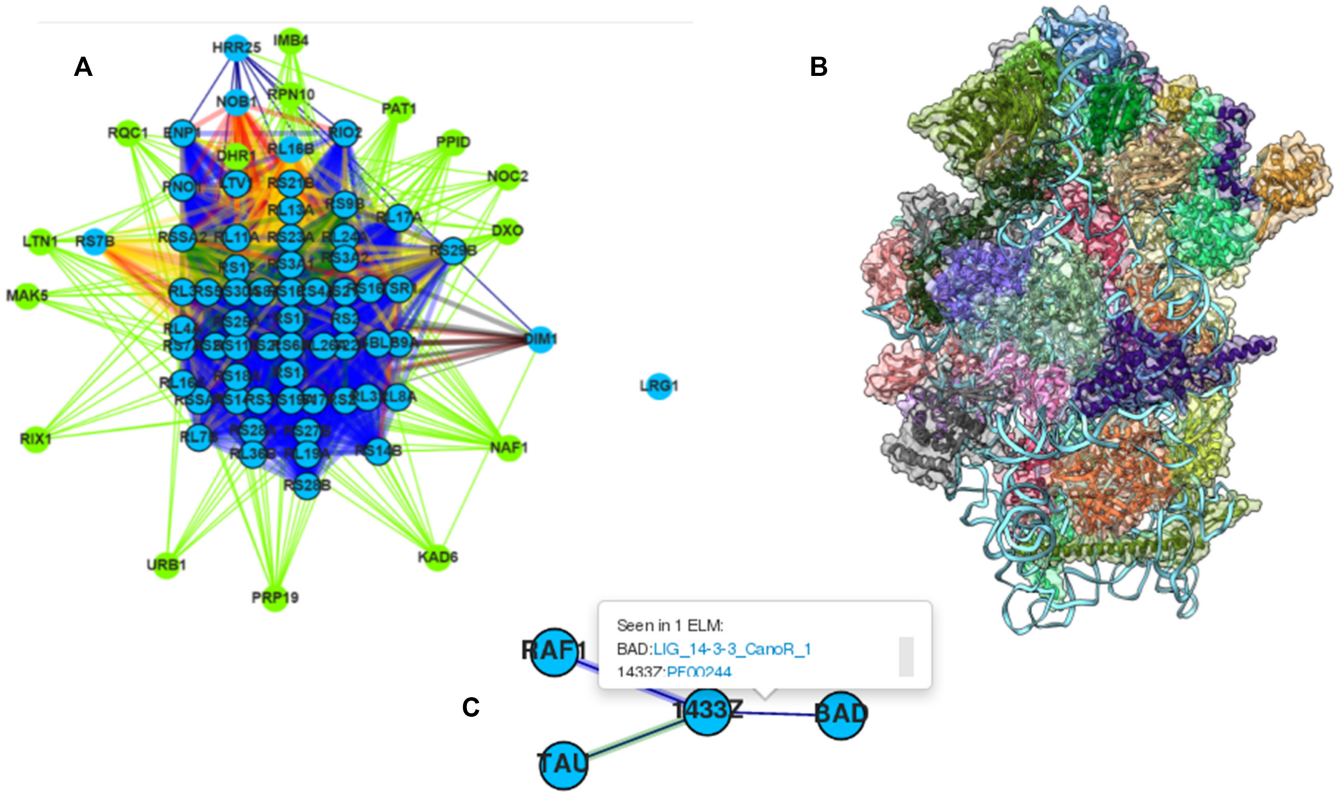


Figure 2. Use cases. (A, B) Use case of *S. cerevisiae* precursors to the small ribosomal subunits. (**A**) Snapshot of Proteo3Dnet showing the 62 proteins identified by AP-MS of pre-40S particles (blue), together with potential interactors from BioGrid analysis (green circles). (**B**) Structural representation of a eukaryotic cytoplasmic pre-40S ribosomal subunit (PDB: 6FAI) (28). Figure generated in UCSF ChimeraX v.0.9 (29). (**C**) Use case of 14–3–3zeta frequent interactors. The small subset of interactors of 14–3–3zeta is shown within the cytoscape.js viewer of Proteo3Dnet. 3D based interfaces are highlighted by thicker blue edges and the ELM connection by a thinner blue edge. For the latter, a popup window shows the type of motif identified.

CONCLUSIONS AND PERSPECTIVES

Protein−protein interactions play a major role in the molecular machinery of life, and their identification through techniques such as AP-MS is highly beneficial. Proteo3Dnet is an on-line facility to assist their analysis. Performing on-the-fly analysis, being not limited to a per species perspective, it proposes a complementary resource to well established tools such as Interactome3D, GeneMania or STRING. It also embeds unique features such as analysis in terms of transient interactions that are out of the scope of former tools. In the context of the rapidly evolving field of PPI identification and analysis, several evolutions of Proteo3Dnet can be foreseen. Firstly, its detection of remote interologs clearly opens the door to the development of the comparative modeling of protein complexes that could contribute to servers such as InterEvDock3 (27). Secondly, Proteo3Dnet is not *per se* limited to AP-MS data, though several adaptations could be required to make the current implementation independent of the UniProt identifiers.

ACKNOWLEDGEMENTS

The authors would like to thank Célia Plisson-Chastang, for careful proof-reading of the pre-40S data analysis. The authors are also grateful to the referees for their constructive feedback.

FUNDING

Investissement d'Avenir Infrastructures Nationales en Biologie et Santé [ANR-10-INBS-05-01 (French Infrastructure for Integrated Structural Biology - FRISBI), ANR-10-INBS-08 (Proteomics French Infrastructure - ProFI), ANR-11-INBS-0013 (French Institute of Bioinformatics - IFB), ANR-18-IDEX-0001 (IdEx Université de Paris)]. Funding for open access charge: INSERM [U1133].
Conflict of interest statement. None declared.

REFERENCES

1. Bauer,A. and Kuster,B. (2003) Affinity purification-mass spectrometry. powerful tools for the characterization of protein complexes. *Eur. J. Biochem.*, **270**, 570–578.
2. Dunham,W.H., Mullin,M. and Gingras,A.-C. (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics*, **12**, 1576–1590.
3. Szklarczyk,D., Gable,A.L., Lyon,D., Junge,A., Wyder,S., Huerta-Cepas,J., Simonovic,M., Doncheva,N.T., Morris,J.H., Bork,P. *et al.* (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.
4. Franz,M., Rodriguez,H., Lopes,C., Zuberi,K., Montojo,J., Bader,G.D. and Morris,Q. (2018) GeneMANIA update 2018. *Nucleic Acids Res.*, **46**, W60–W64.
5. Oughtred,R., Stark,C., Breitkreutz,B.-J., Rust,J., Boucher,L., Chang,C., Kolas,N., O'Donnell,L., Leung,G., McAdam,R. *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021

6. Rodchenkov,I., Babur,O., Luna,A., Aksoy,B.A., Wong,J.V., Fong,D., Franz,M., Siper,M.C., Cheung,M., Wrana,M. *et al.* (2020) Pathway commons 2019 update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.*, **48**, D489–D497.
7. Mosca,R., Céol,A. and Aloy,P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.
8. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
9. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
10. Zhang,Q.C., Petrey,D., Garzón,J.I., Deng,L. and Honig,B. (2013) PrePPI: a structure-informed database of protein–protein interactions. *Nucleic Acids Res.*, **41**, D828–D833.
11. Xu,Q. and Dunbrack,R.L. (2020) ProtCID: a data resource for structural information on protein interactions. *Nat. Commun.*, **11**, 711.
12. Meldal,B.H.M., Forner-Martinez,O., Costanzo,M.C., Dana,J., Demeter,J., Dumousseau,M., Dwight,S.S., Gaulton,A., Licata,L., Melidoni,A.N. *et al.* (2015) The complex portal - an encyclopaedia of macromolecular complexes. *Nucleic Acids Res.*, **43**, D479–D484.
13. Giurgiu,M., Reinhard,J., Brauner,B., Dunger-Kaltenbach,I., Fobo,G., Frishman,G., Montrone,C. and Ruepp,A. (2019) CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.*, **47**, D559–D563.
14. Schweppe,D.K., Zheng,C., Chavez,J.D., Navare,A.T., Wu,X., Eng,J.K. and Bruce,J.E. (2016) XLinkDB 2.0: integrated, large-scale structural analysis of protein crosslinking data. *Bioinformatics*, **32**, 2716–2718.
15. Dapkunas,J., Timinskas,A., Olechnovic,K., Margelevicius,M., Diciunas,R. and Venclovas,C. (2017) The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinforma. Oxf. Engl.*, **33**, 935–937.
16. Postic,G., Marcoux,J., Reys,V., Andreani,J., Vandenbrouck,Y., Bousquet,M.-P., Mouton-Barbosa,E., Cianférani,S., Burlet-Schiltz,O., Guerois,R. *et al.* (2020) Probing protein interaction networks by combining MS-based proteomics and structural data integration. *J. Proteome Res.*, **19**, 2807–2820.
17. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2020) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
18. Bienert,S., Waterhouse,A., de Beer,T.A.P., Tauriello,G., Studer,G., Bordoli,L. and Schwede,T. (2017) The SWISS-MODEL repository—new features and functionality. *Nucleic Acids Res.*, **45**, D313–D319.
19. Zanzoni,A., Ribeiro,D.M. and Brun,C. (2019) Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cell. Mol. Life Sci. CMLS*, **76**, 4407–4412.
20. Mészáros,B., Erdős,G. and Dosztányi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
21. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.
22. Hoksza,D., Gawron,P., Ostaszewski,M. and Schneider,R. (2018) MolArt: a molecular structure annotation and visualization tool. *Bioinformatics*, **34**, 4127–4128.
23. Lecointre,C., Simon,V., Kerneur,C., Allemand,F., Fournet,A., Montarras,I., Pons,J.-L., Gelin,M., Brignatz,C., Urbach,S. *et al.* (2018) Dimerization of the pragmin pseudo-kinase regulates protein tyrosine phosphorylation. *Struct. Lond. Engl. 1993*, **26**, 545–554.
24. Fabre,B., Lambour,T., Delobel,J., Amalric,F., Monsarrat,B., Burlet-Schiltz,O. and Bousquet-Dubouch,M.-P. (2013) Subcellular distribution and dynamics of active proteasome complexes unraveled by a workflow combining in vivo complex cross-linking and quantitative proteomics. *Mol. Cell. Proteomics MCP*, **12**, 687–699.
25. Shayan,R., Rinaldi,D., Larburu,N., Plassart,L., Balor,S., Bouyssié,D., Lebaron,S., Marcoux,J., Gleizes,P.-E. and Plisson-Chastang,C. (2020) Good vibrations: structural remodeling of maturing yeast Pre-40S ribosomal particles followed by cryo-electron microscopy. *Molecules*, **25**, 1125.
26. Mitterer,V., Shayan,R., Ferreira-Cerca,S., Murat,G., Enne,T., Rinaldi,D., Weigl,S., Omanic,H., Gleizes,P.-E., Kressler,D. *et al.* (2019) Conformational proofreading of distant 40S ribosomal subunit maturation events by a long-range communication mechanism. *Nat. Commun.*, **10**, 2754.
27. Quignot,C., Postic,G., Bret,H., Rey,J., Granger,P., Murail,S., Chacón,P., Andreani,J., Tufféry,P. and Guerois,R. (2021) InterEvDock3: A combined template-based and free docking server with increased performance through explicit modeling of complex homologs and integration of covariation-based contact maps. *Nucleic Acids Res.*, doi:10.1093/nar/gkab358.
28. Scaiola,A., Peña,C., Weisser,M., Böhringer,D., Leibundgut,M., Klingauf-Nerurkar,P., Gerhardy,S., Panse,V.G. and Ban,N. (2018) Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. *EMBO J.*, **37**, e98499.
29. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci. Publ. Protein Soc.*, **27**, 14–25.

Downloaded from https://academic.oup.com/nar/article/49/W1/W567/6272410 by Universite de Strasbourg France user on 11 October 2021