



Article
scientifique

Revue de la
littérature

2022

Published
version

Open
Access

This is the published version of the publication, made available in accordance with the publisher's policy.

QAlignWeb: A Server to Predict and Analyze Protein Quaternary Structure

Dey, Sucharita; Prilusky, Jaime; Levy, Emmanuel

How to cite

DEY, Sucharita, PRILUSKY, Jaime, LEVY, Emmanuel. QAlignWeb: A Server to Predict and Analyze Protein Quaternary Structure. In: *Frontiers in molecular biosciences*, 2022, vol. 8, p. 787510. doi: 10.3389/fmolb.2021.787510

This publication URL: <https://archive-ouverte.unige.ch/unige:176894>

Publication DOI: [10.3389/fmolb.2021.787510](https://doi.org/10.3389/fmolb.2021.787510)



QSalgnWeb: A Server to Predict and Analyze Protein Quaternary Structure

Sucharita Dey^{1†}, Jaime Prilusky² and Emmanuel D. Levy^{1*}

¹Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel, ²Department of Life Sciences and Core Facilities, Weizmann Institute of Science, Rehovot, Israel

OPEN ACCESS

Edited by:

Jessica Andreani,
UMR9198 Institut de Biologie
Intégrative de la Cellule (I2BC), France

Reviewed by:

Ezgi Karaca,
Dokuz Eylul University, Turkey
Elodie Laine,
Sorbonne Université, France

*Correspondence:

Emmanuel D. Levy
emmanuel.levy@weizmann.ac.il

†Present Address:

Sucharita Dey,
Department of Bioscience and
Bioengineering, Indian Institute of
Technology, Jodhpur, India

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 30 September 2021

Accepted: 02 December 2021

Published: 05 January 2022

Citation:

Dey S, Prilusky J and Levy ED (2022)
QSalgnWeb: A Server to Predict and
Analyze Protein Quaternary Structure.
Front. Mol. Biosci. 8:787510.
doi: 10.3389/fmolb.2021.787510

The identification of physiologically relevant quaternary structures (QSs) in crystal lattices is challenging. To predict the physiological relevance of a particular QS, QSalgn searches for homologous structures in which subunits interact in the same geometry. This approach proved accurate but was limited to structures already present in the Protein Data Bank (PDB). Here, we introduce a webserver (www.QSalgn.org) allowing users to submit homo-oligomeric structures of their choice to the QSalgn pipeline. Given a user-uploaded structure, the sequence is extracted and used to search homologs based on sequence similarity and PFAM domain architecture. If structural conservation is detected between a homolog and the user-uploaded QS, physiological relevance is inferred. The web server also generates alternative QSs with PISA and processes them the same way as the query submitted to widen the predictions. The result page also shows representative QSs in the protein family of the query, which is informative if no QS conservation was detected or if the protein appears monomeric. These representative QSs can also serve as a starting point for homology modeling.

Keywords: web server, protein evolution, protein quaternary structure, protein structure alignment, physiological interface, crystal contact, protein superposition, protein interactions

INTRODUCTION

Protein self-interactions are prevalent, and they drive the formation of homo-oligomeric structures (Goodsell and Olson, 2000; Levy et al., 2005; Levy and Teichmann, 2013; Marsh and Teichmann, 2015). The spatial organization of the subunits within a protein homo-oligomer defines its quaternary structure (QS). Knowledge of the physiological QS for a protein is not only key to understand its function (Goodsell and Olson, 2000; Marianayagam et al., 2004; Amoutzias et al., 2008), but also to analyze its evolution (Franzosa and Xia, 2009; Garcia-Seisdedos et al., 2017) or to predict the impact of polymorphisms in human diseases (Yates and Sternberg, 2013).

With currently over 150,000 crystallographic structures of proteins in the Protein Data Bank (PDB) (Rose et al., 2017; Armstrong et al., 2019), much of our knowledge on protein QS comes from X-ray crystallography (Perutz et al., 1960). However, one caveat of X-ray crystallography is its requirement for protein molecules to be arranged in a regular array to form a crystal lattice. In this lattice, some protein-protein contacts may be part of a protein's quaternary structure, whereas others only result from the crystal formation and are called crystal contacts.

Much work has been dedicated to distinguishing physiological interfaces from fortuitous crystal contacts (Capitani et al., 2016; Dey and Levy, 2018; Xu and Dunbrack, 2019; Elez et al., 2020). Several physicochemical, geometric, and evolutionary properties of the protein-protein binding surface such as amino acid composition (Ponstingl et al., 2000; Bahadur et al., 2004; Zhu et al., 2006), interface size (Janin, 1997; Henrick and Thornton, 1998; Krissinel and Henrick, 2007), shape (Tsuchiya et al.,

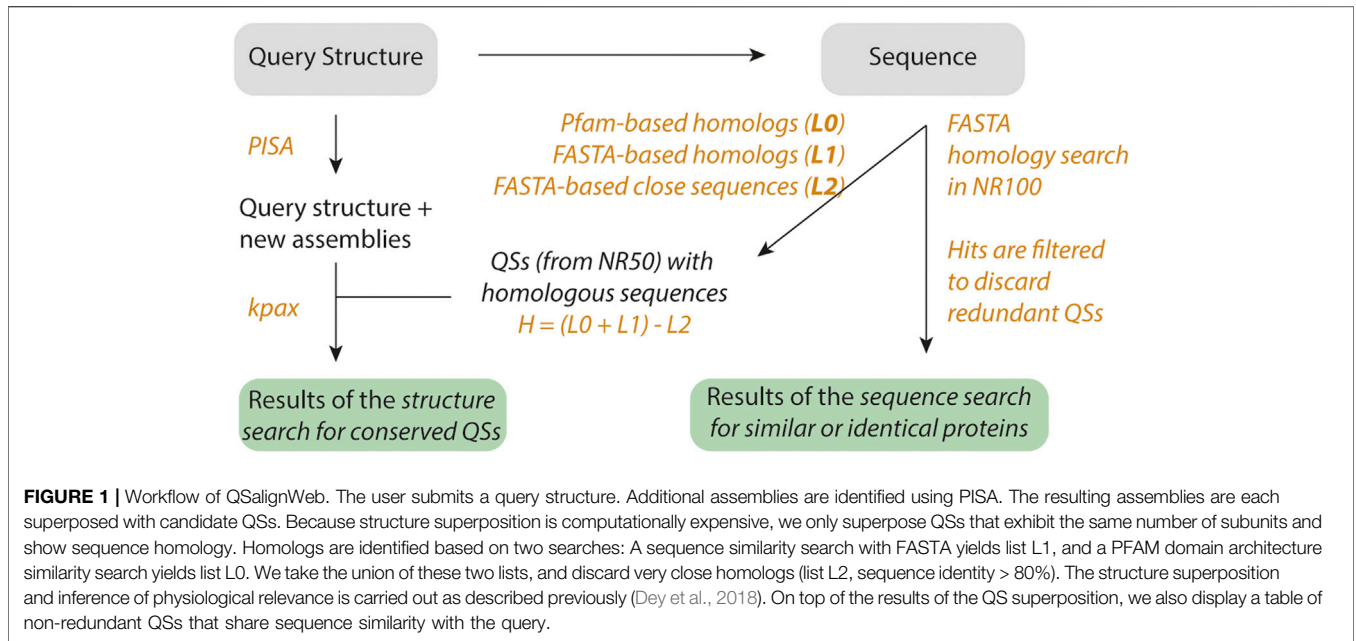


FIGURE 1 | Workflow of QSAIghWeb. The user submits a query structure. Additional assemblies are identified using PISA. The resulting assemblies are each superposed with candidate QSs. Because structure superposition is computationally expensive, we only superpose QSs that exhibit the same number of subunits and show sequence homology. Homologs are identified based on two searches: A sequence similarity search with FASTA yields list L1, and a PFAM domain architecture similarity search yields list L0. We take the union of these two lists, and discard very close homologs (list L2, sequence identity > 80%). The structure superposition and inference of physiological relevance is carried out as described previously (Dey et al., 2018). On top of the results of the QS superposition, we also display a table of non-redundant QSs that share sequence similarity with the query.

QSAIgh [Process](#) [Status](#) [Help](#)

Structure search for conserved QSs

Parameters selected: TM-score > 0.65, Seq. identity with query < 80%, Seq. overlap with query > 60%

Inference: The query structure (4-mer) shows evolutionary conservation with 3ipv_1. Hence, it is likely to be correct.

For query 1fat.pdb (4-mer) we found 2 targets matching the structure.

For PISA-generated 1fat_1 (2-mer) we found 2 targets matching the structure.

Download supporting files

Sequence search for similar or identical proteins

Query 1fat.pdb (4-mer) - Table for chain A

Target	% seq. ident.	#subunits	Symmetry	QS confidence	Prediction
1g7y_3	64.19	2	C2	low confidence	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
3ipv_1	63.52	4	D2	high confidence	Your query has a similar sequence to this target. It has a 80% probability to have the same QS
5kad_1	56.96	4	D2	high confidence	Your query has a similar sequence to this target. It has a 80% probability to have the same QS
1avb_1	55.74	2	C2	high confidence	Your query has a similar sequence to this target. It has a 80% probability to have the same QS
1dbn_1	54.17	2	C2	low confidence	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
6byn_1	50.44	2	C2	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
6ahg_1	50.44	3	C3	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
1mxd_7	50.44	2	C2	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
2ow4_1	50.44	1	NPS	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
1qgl_1	50.44	2	C3	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
555p_1	50.44	2	C2	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
2d3r_1	49.56	4	D2	high confidence	Your query's sequence is fairly distant to this target. It has a 40-60% probability to have the same QS
1ioa_2	49.14	1	NPS	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
3n3h_1	48.94	1	NPS	unknown	We are uncertain about the physiological relevance of this target's QS. We cannot make a prediction based on it.
1sfy_2	48.94	2	C2	high confidence	Your query's sequence is fairly distant to this target. It has a 40-60% probability to have the same QS
2bqp_1	47.83	2	C2	high confidence	Your query's sequence is fairly distant to this target. It has a 40-60% probability to have the same QS
1hqj_1	45.68	4	D2	high confidence	Your query's sequence is fairly distant to this target. It has a 40-60% probability to have the same QS

FIGURE 2 | Results of QSAIghWeb. The result page of QSAIghWeb describes the prediction made by QSAIgh based on the superposition with homologous QSs. The superposition of the two QSs based on which the prediction is made is shown on the right-hand side. A table displays the result of the search of non-redundant QSs with a sequence similar to the query. In this list, the closest homolog with a high-confidence QS is highlighted in green and represents the structure we judge best for homology modeling.

2008), packing (Bahadur et al., 2004; Zhu et al., 2006; Tsuchiya et al., 2008), sequence conservation (Elcock and McCammon, 2001; Guharoy and Chakrabarti, 2005; Baskaran et al., 2014) or structure conservation (Xu and Dunbrack, 2011; Xu and Dunbrack, 2020) can discriminate physiological interfaces from crystal contacts. Several methods also integrated multiple features to train a classifier (Zhu et al., 2006; Bernauer et al., 2008;

Mitra and Pal, 2011; Silva et al., 2015; Hu et al., 2018; Fukasawa and Tomii, 2019; Jiménez-García et al., 2019). Features range from types of atomic contacts at the interface to amino acid interface propensity scores, contact preferences, packing, co-evolution, and more.

While these works focused on individual interfaces, PQS (Henrick and Thornton, 1998) enabled predicting the full QS,

which could involve more than two chains and multiple distinct interfaces (e.g., in the case of a tetramer with dihedral symmetry). Today PQS has been succeeded by PISA (Krissinel and Henrick, 2007), which predicts the stability of QSs compatible with the crystal lattice of a protein. Like PISA, EPPIC (Baskaran et al., 2014) predicts full QSs while placing emphasis on the conservation of amino acids at interfaces to infer their physiological relevance. QSalgn (Dey et al., 2018; Dey and Levy, 2021) also employs evolutionary conservation, but it does not rely on sequence conservation. Instead, it searches for structural conservation of the structure of a QS across homologs. For example, if the QSs of two homologous homotetramers match, QSalgn annotates both as physiologically relevant. An advantage of this approach is its accuracy: While EPPIC and PISA reach accuracies of ~85% for homo-oligomers, the accuracy of QSalgn was 96%. At the same time, a drawback of the methodology is its coverage. Homologous QSs are necessary for the annotation and as a result, only about half of the QSs in the PDB can be annotated with this strategy. Nevertheless, as more structures are solved, coverage will increase. We previously used this strategy to annotate structures from the PDB. The implementation of QSalgn relied on the 3DComplex database (Levy et al., 2005), which is a classification of protein complexes of known structure. This dependency placed a barrier to generalizing this approach to any structure of choice. Here, we report the web server version of QSalgn where users can upload a query structure of their choice.

DESIGN, USE AND PERFORMANCE OF THE QSALIGNWEB SERVER

Server Input

The user provides two pieces of information to the server. First, an email address is required to send the results to the user. Second, a structure in PDB format is uploaded. The structure may contain a particular QS with multiple chains or may also include a single chain. The way in which the QSalgn strategy works has been described and benchmarked before and we refer the reader to the original paper (Dey et al., 2018) for details.

Server Processing

Upon submission of the request by the user, three main actions are performed, as described in **Figure 1**. First, the server analyzes the query structure to identify the number of subunits. If two or more subunits are found, the query structure is considered an assembly, referred to as A0. Then, we use PISA (Krissinel and Henrick, 2007) to generate additional possible assemblies, referred to as A1, A2, ... etc. We also record the number of subunits for each assembly.

Second, the sequence of the structure is extracted, and PFAM domains are predicted based on version 33 (El-Gebali et al., 2019). The PFAM domain architecture of the query sequence is subsequently used to search for homologs with the same domain architecture. This search is executed on a non-redundant set of QSs that we call NR50 (described below). The search by domain architecture yields a list of candidate proteins called L0.

Additionally, we execute a FASTA (Pearson, 1990) search on sequences from the NR50 set and retrieve two lists: L1, the list of homologs showing less than 80% sequence identity (this cut-off can be adjusted in the submission form) together with a sequence-coverage >70% (this cut-off can be adjusted in the submission form). We also retrieve L2, a list of structures with a sequence similar to that of the query (>80% identity). The final list of candidate QSs used in the next step results from the union of L0 and L1 after removing structures from L2. This list contains valid homologs, and we call it H. Each query assembly is then superposed using Kpax (Ritchie, 2016) onto all the QSs from list H that share the same number of subunits. If a superposition yields a TM score above 0.65 (this cut-off can be adjusted in the submission form), we infer that the query QS is conserved and likely physiological.

Third, we search for non-redundant QSs whose sequence is similar to the query sequence. This search is executed on a dataset we call NR100 (described below). The hits are then ordered by decreasing sequence identity, and only the closest target is kept per NR50 group, yielding a list of N closest distinct QS.

The NR50 and NR100 sets consist of non-redundant sets of QSs. Classically, redundancy is removed at the sequence level only. In this case, a homodimer and a monomer sharing 60% sequence identity would be grouped in the same NR50 cluster. In contrast, the non-redundant sets available in 3DComplex compare both graph topology created by connected chains along with sequence identity (Levy et al., 2005). Since a monomer and a dimer yield different graphs, they would end up in two classes. However, two homodimers can be similar at the sequence level while interacting with a different interface. In this case, they would be wrongly grouped based on their graph topology. To avoid this, we created new non-redundant sets following the procedure from 3DComplex, while also imposing that QSs in the same group show a TM-score above 0.65.

Server Output

The server consists of both a visual output, and a downloadable archive containing the information computed by the server. The visual output is illustrated in **Figure 2**. It includes an inference on the validity of the query based on the results of the QS superposition. Additionally, superposed structures are displayed using Mol* (Sehnal et al., 2021). Each QS of the superposed pair can be interactively toggled on or off to facilitate comparing their structure.

The QSs resulting from the search of the NR100 dataset are shown as a table and provide an overview of all QS types that exist among homologs, while prioritizing closely related sequences (even identical sequences). Those QSs are often annotated by QSalgn, and we provide a confidence estimate for these QSs based on their QSalgn annotation. The “high confidence” QS with the closest sequence to the query is highlighted and represents the best candidate for homology modeling. For example, if a user submits a single-chain structure for which no stable assembly is found with PISA, a closely related dimer annotated with high-confidence could serve as a template for modeling the query structure as a dimer.

The downloadable archive contains a list of homologs for each assembly as well as their structure in PDB format files. It also provides aligned coordinates for the query and target to enable comparing their structure using a local visualization software.

Server Implementation

The server is hosted by the Weizmann Institute on a 64-bit machine running Linux CentOS. The backend runs on Perl and MySQL, the frontend uses Perl, JavaScript and the PDB implementation of Mol* (<https://github.com/PDBEurope/pdbe-molstar>) to render the 3D representation of the molecule (Sehnal et al., 2021). QSalgnWeb is freely available and requires an email address for results to be sent when they become available. For details on the QSalgn methodology we refer the reader to the original paper (Dey et al., 2018).

Server Performance

The performance of the prediction pipeline of QSalgnWeb was benchmarked in previous work (Dey et al., 2018). According to this benchmark, the validation of a QS based on the conservation of its geometry yields predictions with an accuracy of 97%. Thus, the user can be confident about an inference of the server when homologs are found with a conserved QS. Also, according to the same benchmark, the correction of a QS by transitivity (i.e., query QS shows the same sequence but a different geometry to a valid QS) is more error-prone, with an accuracy of 89%.

CONCLUSION

Sequence-based tertiary structure predictions have seen a recent breakthrough, notably with Alphafold2 (Cramer, 2021; Jumper et al., 2021) and Rosettafold (Baek et al., 2021). At the same time, predicting the QS of a protein remains challenging even when the tertiary structure is known. The use of evolution and coevolution information together with deep learning has also seen recent developments for scoring and predicting protein-protein interactions (Andreani et al., 2020; Quadir et al., 2021; Quignot et al., 2021; Yan and Huang, 2021). Complementary to such residue-level information, the use of subunit interaction geometry conservation as evidence of a QS being physiological is a powerful approach, which yields accurate predictions (Dey et al., 2018; Dey and Levy, 2021). In that respect, we hope that

making this prediction pipeline available as a web server will help biologists identify relevant QS of proteins and will help them investigate the QS of specific proteins. Together, these approaches will help bridge the gap between the sequence and structure space by adding a third dimension to proteomes and interactomes (Aloy and Russell, 2006; Rolland et al., 2014; Elofsson, 2021; Postic et al., 2021; Sali, 2021), thus making 3D proteomics or “structuromics” (Levy and Vogel, 2021) accessible.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SD and JP developed the server with input from EL. SD and EL wrote the manuscript with help from JP.

FUNDING

This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 819318), by the Israel Science Foundation (Grant No. 1452/18), by a research grant from A.-M. Boucher, by research grants from the Estelle Funk Foundation, the Estate of Fannie Sherr, the Estate of Albert Deligher, the Merle S. Cahn Foundation, Mildred S. Gosden, the Estate of Elizabeth Wachsman, the Arnold Bortman Family Foundation.

ACKNOWLEDGMENTS

We thank Eugene Krissinel for making PISA freely available, and Harry Greenblatt for help with the computer infrastructure. SD is acknowledging support from the Ramalingaswami re-entry fellowship (NO. BT/HRD/35/02/2006 dated March 30th 2021) by DBT, GOVT. of India.

REFERENCES

- Aloy, P., and Russell, R. B. (2006). Structural Systems Biology: Modelling Protein Interactions. *Nat. Rev. Mol. Cell Biol.* 7, 188–197. doi:10.1038/nrm1859
- Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008). Choose Your Partners: Dimerization in Eukaryotic Transcription Factors. *Trends Biochem. Sci.* 33, 220–229. doi:10.1016/j.tibs.2008.02.002
- Andreani, J., Quignot, C., and Guerois, R. (2020). Structural Prediction of Protein Interactions and Docking Using Conservation and Coevolution. *Wires Comput. Mol. Sci.* 10, e1470. doi:10.1002/wcms.1470
- Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Gutmanas, A., Anyango, S., Choudhary, P., et al. (2019). PDB: Improved Findability of Macromolecular

Structure Data in the PDB. *Nucleic Acids Res.* 48, D335–D343. doi:10.1093/nar/gkz990

- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Baskaran, K., Duarte, J. M., Biyani, N., Bliven, S., and Capitani, G. (2014). A PDB-wide, Evolution-Based Assessment of Protein-Protein Interfaces. *BMC Struct. Biol.* 14, 22. doi:10.1186/s12900-014-0022-0
- Bernauer, J., Bahadur, R. P., Rodier, F., Janin, J., and Poupon, A. (2008). DiMoVo: a Voronoi Tessellation-Based Method for Discriminating Crystallographic and Biological Protein-Protein Interactions. *Bioinformatics* 24, 652–658. doi:10.1093/bioinformatics/btn022

- Capitani, G., Duarte, J. M., Baskaran, K., Bliven, S., and Somody, J. C. (2016). Understanding the Fabric of Protein Crystals: Computational Classification of Biological Interfaces and Crystal Contacts. *Bioinformatics* 32, 481–489. doi:10.1093/bioinformatics/btv622
- Cramer, P. (2021). AlphaFold2 and the Future of Structural Biology. *Nat. Struct. Mol. Biol.* 28, 704–705. doi:10.1038/s41594-021-00650-1
- Da Silva, F., Desaphy, J., Bret, G., Rognan, D., and Rognan, D. (2015). IChemPIC: A Random Forest Classifier of Biological and Crystallographic Protein-Protein Interfaces. *J. Chem. Inf. Model.* 55, 2005–2014. doi:10.1021/acs.jcim.5b00190
- Dey, S., and Levy, E. D. (2018). “Inferring and Using Protein Quaternary Structure Information from Crystallographic Data,” in *Protein Complex Assembly: Methods and Protocols*. Editor J. A. Marsh (New York, NY: Springer New York), 357–375. doi:10.1007/978-1-4939-7759-8_23
- Dey, S., and Levy, E. D. (2021). PDB-wide Identification of Physiological Hetero-Oligomeric Assemblies Based on Conserved Quaternary Structure Geometry. *Structure* 29, 1303–1311. doi:10.1016/j.str.2021.07.012
- Dey, S., Ritchie, D. W., and Levy, E. D. (2018). PDB-wide Identification of Biological Assemblies from Conserved Quaternary Structure Geometry. *Nat. Methods* 15, 67–72. doi:10.1038/nmeth.4510
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam Protein Families Database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995
- Elcock, A. H., and McCammon, J. A. (2001). Identification of Protein Oligomerization States by Analysis of Interface Conservation. *Proc. Natl. Acad. Sci.* 98, 2990–2994. doi:10.1073/pnas.061411798
- Elez, K., Bonvin, A. M. J. J., and Vangone, A. (2020). Biological vs. Crystallographic Protein Interfaces: An Overview of Computational Approaches for Their Classification. *Crystals* 10, 114. doi:10.3390/cryst10020114
- Elofsson, A. (2021). Toward Characterising the Cellular 3D-Proteome. *Front. Bioinform.* 1, 2. doi:10.3389/fbinf.2021.598878
- Franzosa, E. A., and Xia, Y. (2009). Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. *Mol. Biol. Evol.* 26, 2387–2395. doi:10.1093/molbev/msp146
- Fukasawa, Y., and Tomii, K. (2019). Accurate Classification of Biological and Non-biological Interfaces in Protein Crystal Structures Using Subtle Covariation Signals. *Sci. Rep.* 9, 12603. doi:10.1038/s41598-019-48913-8
- García-Seisdedos, H., Empereur-Mot, C., Elad, N., and Levy, E. D. (2017). Proteins Evolve on the Edge of Supramolecular Self-Assembly. *Nature* 548, 244–247. doi:10.1038/nature23320
- Goodsell, D. S., and Olson, A. J. (2000). Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* 29, 105–153. doi:10.1146/annurev.biophys.29.1.105
- Guharoy, M., and Chakrabarti, P. (2005). Conservation and Relative Importance of Residues across Protein-Protein Interfaces. *Proc. Natl. Acad. Sci.* 102, 15447–15452. doi:10.1073/pnas.0505425102
- Henrick, K., and Thornton, J. M. (1998). PQS: a Protein Quaternary Structure File Server. *Trends Biochem. Sci.* 23, 358–361. doi:10.1016/s0968-0004(98)01253-5
- Hu, J., Liu, H.-F., Sun, J., Wang, J., and Liu, R. (2018). Integrating Co-evolutionary Signals and Other Properties of Residue Pairs to Distinguish Biological Interfaces from crystal Contacts. *Protein Sci.* 27, 1723–1735. doi:10.1002/pro.3448
- Janin, J. (1997). Specific versus Non-specific Contacts in Protein Crystals. *Nat. Struct. Mol. Biol.* 4, 973–974. doi:10.1038/nsb1297-973
- Jiménez-García, B., Elez, K., Koukos, P. I., Bonvin, A. M., and Vangone, A. (2019). PRODIGY-crystal: a Web-Tool for Classification of Biological Interfaces in Protein Complexes. *Bioinformatics* 35, 4821–4823. doi:10.1093/bioinformatics/btz437
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 372, 774–797. doi:10.1016/j.jmb.2007.05.022
- Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2005). 3D Complex: a Structural Classification of Protein Complexes. *Plos Comp. Biol.* 2, e155. doi:10.1371/journal.pcbi.0020155
- Levy, E. D., and Teichmann, S. A. (2013). Structural, Evolutionary, and Assembly Principles of Protein Oligomerization. *Prog. Mol. Biol. Transl. Sci.* 117, 25–51. doi:10.1016/B978-0-12-386931-9.00002-7
- Levy, E. D., and Vogel, C. (2021). “Structuromics”: Another Step toward a Holistic View of the Cell. *Cell* 184, 301–303. doi:10.1016/j.cell.2020.12.030
- Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004). The Power of Two: Protein Dimerization in Biology. *Trends Biochem. Sci.* 29, 618–625. doi:10.1016/j.tibs.2004.09.006
- Marsh, J. A., and Teichmann, S. A. (2015). Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* 84, 551–575. doi:10.1146/annurev-biochem-060614-034142
- Mitra, P., and Pal, D. (2011). Combining Bayes Classification and Point Group Symmetry under Boolean Framework for Enhanced Protein Quaternary Structure Inference. *Structure* 19, 304–312. doi:10.1016/j.str.2011.01.009
- Pearson, W. R. (1990). “[5] Rapid and Sensitive Sequence Comparison with FASTP and FASTA,” in *Methods in Enzymology* (Cambridge, UK: Academic Press), 63–98. doi:10.1016/0076-6879(90)83007-V
- Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960). Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* 185, 416–422. doi:10.1038/185416a0
- Ponstingl, H., Henrick, K., and Thornton, J. M. (2000). Discriminating between Homodimeric and Monomeric Proteins in the Crystalline State. *Proteins* 41, 47–57. doi:10.1002/1097-0134(20001001)41:1<47:aid-prot80>3.0.co;2-8
- Postic, G., Andreani, J., Marcoux, J., Reys, V., Guerois, R., Rey, J., et al. (2021). Proteo3Dnet: a Web Server for the Integration of Structural Information with Interactomics Data. *Nucleic Acids Res.* 49, W567–W572. doi:10.1093/nar/gkab332
- Prasad Bahadur, R., Chakrabarti, P., Rodier, F., and Janin, J. (2004). A Dissection of Specific and Non-specific Protein-Protein Interfaces. *J. Mol. Biol.* 336, 943–955. doi:10.1016/j.jmb.2003.12.073
- Quadri, F., Roy, R. S., Halfmann, R., and Cheng, J. (2021). DNCON2_Inter: Predicting Interchain Contacts for Homodimeric and Homomultimeric Protein Complexes Using Multiple Sequence Alignments of Monomers and Deep Learning. *Sci. Rep.* 11, 12295. doi:10.1038/s41598-021-91827-7
- Quignot, C., Granger, P., Chacón, P., Guerois, R., and Andreani, J. (2021). Atomic-level Evolutionary Information Improves Protein-Protein Interface Scoring. *Bioinformatics* 37, 3175–3181. doi:10.1093/bioinformatics/btab254
- Ritchie, D. W. (2016). Calculating and Scoring High Quality Multiple Flexible Protein Structure Alignments. *Bioinformatics* 32, 2650–2658. doi:10.1093/bioinformatics/btw300
- Rolland, T., Taşan, M., Charlotheaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., et al. (2014). A Proteome-Scale Map of the Human Interactome Network. *Cell* 159, 1212–1226. doi:10.1016/j.cell.2014.10.050
- Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2017). The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nucleic Acids Res.* 45, D271–D281. doi:10.1093/nar/gkw1000
- Sali, A. (2021). From Integrative Structural Biology to Cell Biology. *J. Biol. Chem.* 296, 100743. doi:10.1016/j.jbc.2021.100743
- Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., et al. (2021). Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures. *Nucleic Acids Res.* 49, W431–W437. doi:10.1093/nar/gkab314
- Tsuchiya, Y., Nakamura, H., and Kinoshita, K. (2008). Discrimination between Biological Interfaces and crystal-packing Contacts. *Aabc* 1, 99–113. doi:10.2147/aabc.s4255
- 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. doi:10.1016/j.sbi.2019.03.006
- Xu, Q., and Dunbrack, R. L., Jr (2020). ProtCID: a Data Resource for Structural Information on Protein Interactions. *Nat. Commun.* 11, 711. doi:10.1038/s41467-020-14301-4
- Xu, Q., and Dunbrack, R. L., Jr (2011). The Protein Common Interface Database (ProtCID)-A Comprehensive Database of Interactions of Homologous Proteins in Multiple crystal Forms. *Nucleic Acids Res.* 39, D761–D770. doi:10.1093/nar/gkq1059

- Yan, Y., and Huang, S.-Y. (2021). Accurate Prediction of Inter-protein Residue-Residue Contacts for Homo-Oligomeric Protein Complexes. *Brief. Bioinform.* 22, bbab038. doi:10.1093/bib/bbab038
- Yates, C. M., and Sternberg, M. J. E. (2013). The Effects of Non-synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein-Protein Interactions. *J. Mol. Biol.* 425, 3949–3963. doi:10.1016/j.jmb.2013.07.012
- Zhu, H., Domingues, F. S., Sommer, I., and Lengauer, T. (2006). NOXclass: Prediction of Protein-Protein Interaction Types. *BMC Bioinformatics* 7, 27. doi:10.1186/1471-2105-7-27

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Dey, Prilusky and Levy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.