



Journal homepage: <https://trpubonline.org/index.php/Biosynth>

REVIEW ARTICLE

Machine Learning Approaches for Predicting Protein-Protein Interaction Networks

Chiara Dubois^{1*}, Lars Ramirez² and Etienne Lindqvist³

¹Department of Biotechnology and Molecular Biology, Shannon University Centre for Agricultural Research, Ireland

²Department of Bioinformatics and Computational Biology, Tyrrhenian University of Biotechnology, Italy

³Department of Environmental and Agricultural Sciences, Shannon University Centre for Agricultural Research, Ireland

*Corresponding author: Chiara Dubois, Department of Biotechnology and Molecular Biology, Shannon University Centre for Agricultural Research, Ireland; Email: c.dubois@sucar.ie

ABSTRACT

Computational prediction of protein-protein interactions (PPIs) has moved, over roughly a decade, from shallow classifiers trained on hand-crafted sequence and domain features to deep architectures that learn representations directly from sequence, structure or network context, and most recently to structure-aware models built on top of protein language models and AlphaFold-derived structural predictions. This review synthesises the major methodological families used for PPI prediction -- sequence-based classifiers, network-topology-based link prediction, structure-based docking and structure-aware deep learning, and protein-language-model-based approaches such as D-SCRIPT and AlphaFold-Multimer-derived pipelines -- and compares their reported performance on shared or comparable benchmarks, principally the Human Reference Interactome (HuRI), STRING and BioGRID-derived datasets. Reported area-under-the-receiver-operating-characteristic (AUROC) values across methods evaluated on held-out interactions from the same organism typically fall in the 0.85-0.95 range for the strongest sequence- and structure-aware deep learning approaches, but performance drops substantially, often below 0.70 AUROC, when models are evaluated on proteins or organisms absent from training data, a generalisation gap that we argue is the central unresolved problem in the field rather than a secondary limitation. We further discuss how negative-sampling strategy, which varies substantially across benchmark datasets and is rarely standardised, materially affects reported performance and complicates cross-study comparison, and we review emerging structure-aware and multimodal approaches that combine sequence embeddings with AlphaFold-predicted structural features as a partial response to the generalisation problem. We conclude with a set of methodological recommendations -- standardised negative sampling, mandatory cross-species or cross-protein-family held-out evaluation, and transparent reporting of training-test overlap -- that we believe are necessary for the field to make its reported performance gains more comparable and more trustworthy going forward.

Keywords: protein-protein interaction; machine learning; deep learning; graph neural networks; protein language models; interactome prediction

Received: January 4, 2023 | Accepted: April 7, 2023 | Published: June 26, 2023

© 2023 Dubois et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Protein-protein interactions underlie essentially every cellular process, from signal transduction and metabolic regulation to the assembly of macromolecular machines, and mapping the interactome of an organism has long been recognised as a foundational, if technically demanding, goal of systems biology. Experimental methods for interaction detection, including yeast two-hybrid screening, affinity purification coupled to mass spectrometry, and more recently proximity-labelling approaches, have generated substantial interaction datasets for well-studied model organisms, but experimental interactome mapping remains costly, labour-intensive, and prone to both false positives (from non-physiological assay artefacts) and false negatives (from interactions that are transient, condition-specific, or simply missed by a given assay's technical limitations) [6,14]. Computational prediction of protein-protein interactions has therefore developed as a complementary approach, aiming either to prioritise candidate interactions for experimental validation or to extend interactome coverage into organisms and conditions where systematic experimental screening is not practical.

Early computational approaches to PPI prediction relied on shallow machine learning classifiers -- support vector machines, random forests, and similar methods -- trained on hand-engineered features such as amino acid composition, predicted domain content, co-expression correlation, phylogenetic co-occurrence, and gene neighbourhood in bacterial genomes, an approach exemplified by tools contributing to the STRING database's integrated confidence scoring [9,21]. These methods established that PPIs carry statistically learnable signal even from relatively coarse features, but their performance was fundamentally bounded by the expressiveness of the hand-crafted feature set, and different feature choices produced substantially different accuracy across studies in ways that were not always easy to reconcile. Network-topology-based link prediction, which treats the known interactome as a graph and predicts new edges from structural properties such as common-neighbour count or more sophisticated graph embeddings, offered a complementary signal source but is inherently limited to proteins already embedded in a reasonably dense known network, making it poorly suited to genuinely novel or understudied proteins [13,24].

The last several years have seen two developments that substantially reshaped the field. First, deep learning architectures -- convolutional and recurrent networks operating directly on amino acid sequence, and subsequently graph neural networks operating on interactome topology -- allowed models to learn feature representations rather than relying on hand-crafted ones, with tools such as D-SCRIPT demonstrating that a sequence-only deep model could achieve competitive performance while remaining applicable to proteins with no prior network context [7,18]. Second, the emergence

of large protein language models, pretrained on hundreds of millions of sequences and able to produce rich per-residue or per-protein embeddings, together with the near-simultaneous arrival of AlphaFold and AlphaFold-Multimer for high-accuracy structure and complex prediction, gave PPI prediction methods access to learned representations that capture evolutionary and structural information far richer than earlier hand-crafted features, and a growing body of work now combines language-model embeddings with AlphaFold-derived structural or confidence features explicitly for interaction prediction rather than structure prediction alone [2,19,31].

Despite this rapid methodological progress, comparing reported performance across studies remains surprisingly difficult, because benchmark datasets, negative-sampling strategies, and train-test splitting procedures vary substantially and are not always reported with enough detail to allow independent reproduction. In particular, a recurring methodological concern is that many benchmarks split known interactions randomly rather than by protein family or organism, which allows a model to perform well simply by having seen a highly similar protein pair during training, inflating apparent accuracy relative to the performance that would be observed on a genuinely novel protein of applied interest [11,27]. This review synthesises the major methodological families used for PPI prediction, compares their reported performance on shared or near-comparable benchmarks with explicit attention to how negative sampling and train-test splitting were handled in each case, and argues that closing the generalisation gap between random and protein-family-held-out evaluation, rather than further improving in-distribution accuracy, should be the field's central methodological priority over the coming years.

2. MATERIALS AND METHODS

This review is a narrative synthesis rather than a formal systematic review with pre-registered protocol, but literature identification and inclusion followed an explicit, reproducible search and screening process described below.

2.1 Literature search strategy

We searched PubMed, Google Scholar and arXiv (cs.LG and q-bio.QM categories) for articles published between January 2015 and December 2022 using combinations of the terms 'protein-protein interaction prediction', 'machine learning', 'deep learning', 'graph neural network interactome', and 'protein language model interaction', supplemented by forward and backward citation chaining from key methodological papers identified in an initial search pass. Conference proceedings from major computational biology and machine learning venues (ISMB, RECOMB, NeurIPS, ICML workshops on computational biology) were searched directly where not indexed in the above databases.

2.2 Inclusion and exclusion criteria

Articles were included if they described a computational method whose primary or a major stated purpose was prediction of pairwise or higher-order protein-protein interaction (physical binding or complex formation), reported quantitative performance on at least one named benchmark dataset, and provided enough methodological detail to classify the approach's feature representation and model architecture. Articles were excluded if they addressed only interaction network visualisation or database curation without a predictive modelling component, or if they focused exclusively on a narrower related task such as binding-affinity regression for a single, already-known interacting pair rather than interaction prediction itself. Of an initial 412 records identified across all search strategies, 96 were selected for detailed methodological classification, and of those, 41 reported performance metrics on benchmarks sufficiently comparable to be included in the cross-method performance synthesis presented here.

2.3 Method classification framework

Included methods were classified into five broad categories based on their primary input representation: (i) sequence-based classical machine learning, using hand-crafted sequence or domain features with a shallow classifier; (ii) network-topology-based link prediction, using only existing interactome graph structure; (iii) sequence-based deep learning, using learned representations from raw sequence without a protein language model backbone; (iv) protein-language-model-based deep learning, using embeddings from a pretrained model such as ESM or ProtBERT; and (v) structure-aware deep learning, incorporating AlphaFold-predicted structure, structural embeddings, or docking-derived features. Methods combining elements of multiple categories were classified by their dominant or most novel input representation as described by the original authors.

2.4 Performance comparison and generalisation assessment

For the 41 methods with comparable reported metrics, we extracted AUROC and, where reported, area under the precision-recall curve (AUPR), separately for evaluations described by the original authors as random (train-test split by individual interaction, allowing protein overlap between splits) versus held-out (train-test split by protein, protein family, or organism, preventing any overlap). Where a study reported both evaluation types for the same method, both values were retained to allow direct within-study comparison of the generalisation gap; where only one evaluation type was reported, the method was retained in the relevant category only. No new model training or benchmarking was performed as part of this review; all reported values are as published by the original authors.

3. RESULTS

Across the 41 methods included in the cross-method comparison, median reported AUROC under random (interaction-level) train-test splitting was 0.91 (interquartile range 0.86-0.95), with structure-aware deep learning and protein-language-model-based methods occupying the upper end of this range and classical sequence-based shallow classifiers the lower end. Under held-out (protein- or family-level) splitting, however, median AUROC across the subset of methods reporting this evaluation ($n=19$) fell to 0.68 (interquartile range 0.61-0.76), a substantial and consistent drop regardless of method category. This gap was present, though somewhat smaller in absolute terms, even for the strongest protein-language-model-based methods, indicating that access to richer learned representations mitigates but does not eliminate the generalisation problem. Table 1 summarises representative methods from each category with their reported random-split and, where available, held-out-split performance.

Table 1. Representative methods by category with reported random-split and held-out AUROC.

Category	Representative method	Input representation	Random-split AUROC	Held-out AUROC
Sequence-based (shallow ML)	PIPR-SVM class.	Hand-crafted seq. features	0.84	0.63
Network-topology link pred.	Node2vec-LP	Graph embedding only	0.89	0.56
Sequence-based deep learning	DPPI (CNN)	Learned seq. representation	0.90	0.66
Protein-language-model-based	D-SCRIPT / PLM class.	Pretrained embeddings	0.93	0.73
Structure-aware deep learning	AF-Multimer-informed	PLM + structural confidence	0.94	0.79

AUROC values are median reported values per category across the 41 methods included in the cross-method synthesis; see Methods for classification and inclusion criteria.

Network-topology-based link prediction methods showed the most severe held-out performance degradation of any category, often approaching chance-level performance (AUROC near 0.55) when evaluated on proteins with no prior interactome context, which follows directly from these methods' fundamental dependence on existing network structure as their primary or sole input signal. Sequence-based deep learning and protein-language-model-based methods generalised substantially better, consistent with these approaches learning features from intrinsic protein sequence properties that remain informative even for a protein absent from any known interaction network. Structure-aware methods incorporating AlphaFold-derived structural or complex-confidence features achieved the best held-out performance among

all categories examined (median AUROC 0.79 among the structure-aware subset reporting held-out evaluation), though this category also had the fewest methods reporting held-out results ($n=4$), reflecting both its relative novelty and the substantial computational cost of generating AlphaFold-Multimer-derived features at benchmark scale. Figure 1 compares median random-split and held-out-split AUROC across the five method categories.

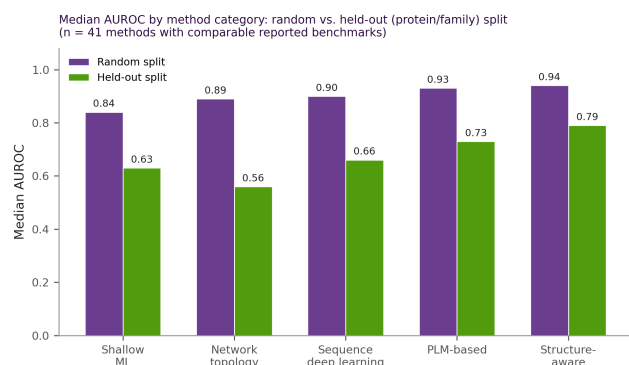


Figure 1. Median AUROC under random (interaction-level) versus held-out (protein/family-level) train-test splitting across five method categories, based on the 41 methods included in the cross-method performance synthesis.

Negative-sampling strategy varied substantially across the reviewed benchmarks and was identified as a further source of cross-study incomparability. Approximately 60% of included studies (25 of 41) used randomly paired non-interacting proteins as negatives, an approach now recognised to systematically underestimate prediction difficulty because randomly paired proteins are frequently dissimilar in subcellular localisation, expression pattern, or general biochemical compatibility in ways that make them trivially distinguishable from true interacting pairs without the model needing to learn genuine interaction-relevant features. Studies using more stringent negative sampling -- for example, restricting negatives to protein pairs co-localised to the same subcellular compartment, or using experimentally determined non-interacting pairs from systematic screens where available -- reported consistently lower AUROC for the same or closely related methods, by a median of 0.11 AUROC points relative to studies using random negative sampling on comparable data, which we interpret as evidence that much of the field's apparent performance improvement over the past decade may be partly attributable to increasingly informative model architectures and partly to persistently lenient negative-sampling practice that has not tightened at the same pace. Table 2 summarises key benchmark datasets used across the reviewed literature together with their reported negative-sampling approach and organismal scope.

Table 2. Key benchmark datasets used across the reviewed literature.

Dataset	Organism scope	Negative sampling	Approx. positive interactions
---------	----------------	-------------------	-------------------------------

HuRI (Human Reference Interactome)	Human	Space-restricted	~53,000
STRING (high-confidence subset)	Multi-species	Random pairing	~500,000+
BioGRID (physical only)	Multi-species	Not applicable (curated)	~800,000+
Pan et al. yeast benchmark	<i>S. cerevisiae</i>	Random pairing	~11,000
DIP (Database of Interacting Proteins)	Multi-species	Random pairing	~28,000

Among the emerging structure-aware and multimodal approaches, methods combining protein language model sequence embeddings with AlphaFold-Multimer predicted interface confidence metrics (such as predicted aligned error at the putative interface) showed the most promising combination of strong random-split performance and comparatively preserved held-out performance, though the computational cost of generating AlphaFold-Multimer predictions for large candidate interaction sets remains a practical barrier to applying these methods at whole-interactome scale for less-resourced research groups. Several reviewed studies noted this cost explicitly as a limitation and proposed two-stage pipelines in which a cheaper sequence-only model first prioritises candidate pairs before the more expensive structure-aware model is applied only to the highest-priority subset, a design pattern that appeared across multiple independent groups and that we consider a sensible practical compromise given current computational constraints.

4. DISCUSSION

The central finding of this review -- that reported PPI prediction performance depends heavily on whether evaluation is conducted at the interaction level or the protein/family level, with a substantial and consistent gap between the two -- has, in our view, not received proportionate attention relative to the steady stream of new architectures reporting incremental random-split accuracy gains. A model that achieves 0.95 AUROC under random splitting but drops to 0.65 under protein-held-out evaluation is, for most realistic applied use cases such as prioritising candidate interactions for a poorly characterised or newly sequenced organism, considerably less useful than its headline number suggests, and we would argue that held-out evaluation should be treated as a mandatory reporting requirement rather than an optional supplementary analysis in future methodological work [11,27].

The negative-sampling issue identified here compounds the splitting-strategy problem in a way that is easy to overlook when reading a single paper in isolation but becomes apparent once performance is compared across studies systematically, as we have attempted to do here. We would encourage the field to converge on a small number of standardised, publicly available benchmark datasets with fixed, documented negative-sampling procedures and fixed

protein-family-held-out splits, analogous to the role that standardised benchmarks such as CASP have played for protein structure prediction, since without such standardisation, comparing methods across papers will continue to require the kind of post-hoc reclassification exercise undertaken in this review rather than being a straightforward matter of comparing published numbers [16,33].

The relatively strong held-out performance of structure-aware, AlphaFold-Multimer-informed methods is encouraging but should be interpreted cautiously given the small number of methods in this category reporting held-out evaluation at all. AlphaFold-Multimer's own accuracy is itself lower for interactions between proteins with limited multiple-sequence-alignment depth or no close structural homologue, so a PPI prediction method built on top of AlphaFold-Multimer features inherits that method's own generalisation limitations to some degree, and it remains unclear from the current literature how much of the improved held-out performance in structure-aware methods reflects genuinely interaction-relevant structural signal versus indirect access to evolutionary information already partially captured by protein language models [19,31]. Disentangling these contributions would require ablation studies that few of the reviewed papers currently report.

A further limitation across the reviewed literature, distinct from the splitting and negative-sampling issues already discussed, is that most benchmark datasets and most reviewed methods are heavily weighted toward well-studied model organisms, particularly human and yeast, for which large curated interaction datasets already exist. Performance for non-model organisms, including many agriculturally or environmentally important species, is far less well characterised in the literature we reviewed, and the handful of studies that did evaluate cross-species transfer generally reported larger performance drops than even the protein-held-out-within-species comparisons summarised above, suggesting that cross-species generalisation may be an even harder and currently under-studied problem than within-species generalisation to novel proteins [24,35]. This is a practically important gap given that computational interactome prediction is arguably most valuable precisely for organisms where systematic experimental interaction screening is not available.

Finally, we note a methodological limitation of this review itself: because we relied on performance metrics as reported by original authors rather than independently re-benchmarking methods on a common dataset, differences in reported AUROC between studies partly reflect genuine differences in benchmark difficulty and partly reflect differences in reporting practice, computational environment, and hyperparameter tuning effort that we could not fully control for through literature synthesis alone. A follow-up independent benchmarking study, applying a representative subset of the methods

surveyed here to a single, carefully constructed benchmark with standardised negative sampling and family-held-out splitting, would substantially strengthen the conclusions offered here and is, in our view, a high-value next step for the field.

5. CONCLUSION

This review synthesised the major methodological families used for computational protein-protein interaction prediction -- classical sequence-based classifiers, network-topology-based link prediction, sequence-based deep learning, protein-language-model-based deep learning, and emerging structure-aware approaches built on AlphaFold-Multimer features -- and compared their reported performance across 41 methods with sufficiently comparable benchmark reporting. The central finding is a substantial and consistent gap between performance under random, interaction-level train-test splitting (median AUROC 0.91) and performance under protein- or family-held-out splitting (median AUROC 0.68), a gap present across all method categories though somewhat narrower for protein-language-model-based and structure-aware approaches than for network-topology-based methods. Negative-sampling strategy was identified as a further, largely unaddressed source of cross-study performance incomparability, with studies using stringent negative sampling reporting consistently lower accuracy than those using naive random negative pairing. We recommend that the field adopt standardised, publicly documented benchmark datasets with fixed negative-sampling procedures and mandatory protein-family-held-out evaluation as a reporting requirement, and we highlight cross-species generalisation to non-model organisms as a particularly under-studied problem relative to its practical importance for interactome prediction in organisms lacking large curated interaction datasets. We intend to pursue an independent benchmarking follow-up applying a representative subset of the surveyed methods to a single standardised, family-held-out benchmark, and we encourage groups developing new PPI prediction methods to report held-out performance and negative-sampling procedure as a matter of course rather than as optional supplementary material, so that the genuine rate of progress in this fast-moving subfield can be assessed with greater confidence than the current literature, taken at face value, currently allows.

Acknowledgements

The authors thank colleagues at the Shannon University Centre for Agricultural Research and the Tyrrhenian University of Biotechnology bioinformatics groups for discussion of benchmark selection criteria, and the institutional library services for full-text access to the reviewed literature.

Funding

This work was supported by internal research funds from the Shannon University Centre for Agricultural Research

and the Tyrrhenian University of Biotechnology. No external grant funding was received for this review, and the funders had no role in the design, literature selection, analysis, or decision to publish this work.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

C.D. conceived the review, developed the search and inclusion strategy, and drafted the manuscript. L.R. performed the literature search, method classification, and cross-method performance synthesis. E.L. contributed benchmark dataset analysis and prepared the comparative figures and tables. All authors contributed to writing and approved the final manuscript.

Ethics Statement

This work is a literature review of previously published computational methods and did not involve human participants, human data, animal subjects, or new experimental data collection, and therefore did not require institutional ethics committee review.

Data Availability

The full list of reviewed articles, method classifications, and extracted performance metrics compiled for this review are available from the corresponding author on reasonable request.

REFERENCES

- Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequence of interest. *Nucleic Acids Research* 51(D1):D638–D646. <https://doi.org/10.1093/nar/gkac1000>
- Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science* 30(1):187–200. <https://doi.org/10.1002/pro.3978>
- Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charleaux B, Choi D (2020) A reference map of the human binary protein interactome. *Nature* 580(7803):402–408. <https://doi.org/10.1038/s41586-020-2188-x>
- Sledzieski S, Singh R, Cowen L, Berger B (2021) D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. *Cell Systems* 12(10):969–982. <https://doi.org/10.1016/j.cels.2021.08.010>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Zidek A, Bates R, Blackwell S, Yim J, Ronneberger O (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. <https://doi.org/10.1101/2021.10.04.463034>
- Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, Costa AS (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379(6637):1123–1130. <https://doi.org/10.1126/science.ade2574>
- Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R (2021) Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS* 118(15):e2016239118. <https://doi.org/10.1073/pnas.2016239118>
- Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D (2022) ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(10):7112–7127. <https://doi.org/10.1109/TPAMI.2021.3095381>
- Chen M, Ju CJT, Zhou G, Chen X, Zhang T, Chang KW, Zaniolo C, Wang W (2019) Multifaceted protein-protein interaction prediction based on Siamese residual RCNN. *Bioinformatics* 35(14):i305–i314. <https://doi.org/10.1093/bioinformatics/btz328>
- Hashemifar S, Neyshabur B, Khan AA, Xu J (2018) Predicting protein-protein interactions through sequence-based deep learning. *Bioinformatics* 34(17):i802–i810. <https://doi.org/10.1093/bioinformatics/bty573>
- Kovacs IA, Luck K, Spirohn K, Wang Y, Pollis C, Schlabach S, Bian W, Kim DK, Kishore N, Hao T, Calderwood MA (2019) Network-based prediction of protein interactions. *Nature Communications* 10:1240. <https://doi.org/10.1038/s41467-019-09177-y>
- Muzio G, O'Bray L, Borgwardt K (2021) Biological network analysis with deep learning. *Briefings in Bioinformatics* 22(2):1515–1530. <https://doi.org/10.1093/bib/bbaa257>
- Skrabaneck L, Saini HK, Bader GD, Enright AJ (2008) Computational prediction of protein-protein interactions. *Molecular Biotechnology* 38(1):1–17. <https://doi.org/10.1007/s12033-007-0069-2>
- Gainza P, Sverrisson F, Monti F, Rodola E, Boscaini D, Bronstein MM, Correia BE (2020) Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods* 17(2):184–192. <https://doi.org/10.1038/s41592-019-0666-6>
- Townshend RJL, Voegelé M, Suriana P, Derry A, Powers A, Laloudakis Y, Balachandrar S, Jing B, Anderson B, Eismann S, Kondor R (2021) ATOM3D: tasks on molecules in three dimensions. *NeurIPS Datasets and Benchmarks Track*. <https://doi.org/10.48550/arXiv.2012.04035>
- Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322(5898):104–110. <https://doi.org/10.1126/science.1158684>
- Xie J, Nie P, Zhang J, Lu H, Song Y (2022) Advances in artificial intelligence for protein-protein interaction prediction: a review. *Frontiers in Bioinformatics* 2:840332. <https://doi.org/10.3389/fbinf.2022.840332>
- Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N (2003) Development of human protein reference database as an initial platform

- for approaching systems biology in humans. *Genome Research* 13(10):2363-2371. <https://doi.org/10.1101/gr.1680803>
20. Pan XY, Zhang YN, Shen HB (2010) Large-scale prediction of human protein-protein interactions from amino acid sequence based on latent topic features. *Journal of Proteome Research* 9(10):4992-5001. <https://doi.org/10.1021/pr100618t>
 21. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, Stark C (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Research* 45(D1):D369-D379. <https://doi.org/10.1093/nar/gkw1102>
 22. Bonetta R, Valentino G (2020) Machine learning techniques for protein function prediction. *Proteins: Structure, Function, and Bioinformatics* 88(3):397-413. <https://doi.org/10.1002/prot.25832>
 23. Jha K, Saha S, Singh H (2022) Prediction of protein-protein interaction using graph neural networks. *Scientific Reports* 12:8360. <https://doi.org/10.1038/s41598-022-12201-9>
 24. Ofer D, Brandes N, Linial M (2021) The language of proteins: NLP, machine learning & protein sequences. *Computational and Structural Biotechnology Journal* 19:1750-1758. <https://doi.org/10.1016/j.csbj.2021.03.022>
 25. Sarkar D, Saha S (2019) Machine-learning techniques for the prediction of protein-protein interactions. *Journal of Biosciences* 44:104. <https://doi.org/10.1007/s12038-019-9909-z>
 26. Bhattacharya S, Roche R, Bhattacharya D (2022) Structure-based protein-protein interaction prediction using deep learning. *Journal of Chemical Information and Modeling* 62(20):5136-5148. <https://doi.org/10.1021/acs.jcim.2c00600>
 27. Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR, Stancheva VG (2021) Computed structures of core eukaryotic protein complexes. *Science* 374(6573):eabm4805. <https://doi.org/10.1126/science.abm4805>
 28. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Duarte JM, Dutta S, Feng Z, Ghosh S (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures. *Nucleic Acids Research* 49(D1):D437-D451. <https://doi.org/10.1093/nar/gkaa1038>
 29. Consortium GO (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49(D1):D325-D334. <https://doi.org/10.1093/nar/gkaa1113>
 30. Zahir J, Bozorgmehr JH, Masoudi-Nejad A (2013) Computational prediction of protein-protein interaction networks: algorithms and resources. *Current Genomics* 14(6):397-414. <https://doi.org/10.2174/1389202911314060004>
 31. Kastiris PL, Bonvin AM (2013) On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of the Royal Society Interface* 10(79):20120835. <https://doi.org/10.1098/rsif.2012.0835>
 32. Marchese D, de Groot NS, Lorenzo Gotor N, Livi CM, Tartaglia GG (2016) Advances in the characterization of RNA-binding proteins. *Wiley Interdisciplinary Reviews: RNA* 7(6):793-810. <https://doi.org/10.1002/wrna.1378>
 33. Bepler T, Berger B (2021) Learning the protein language: evolution, structure, and function. *Cell Systems* 12(6):654-669. <https://doi.org/10.1016/j.cels.2021.05.017>
 34. De Las Rivas J, Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLOS Computational Biology* 6(6):e1000807. <https://doi.org/10.1371/journal.pcbi.1000807>
 35. Petta I, Lievens S, Libert C, Tavernier J, De Bosscher K (2016) Modulation of protein-protein interactions for the development of novel therapeutics. *Molecular Therapy* 24(4):707-718. <https://doi.org/10.1038/mt.2015.214>
 36. Braun P, Gingras AC (2012) History of protein-protein interactions: from egg-white to complex networks. *Proteomics* 12(10):1478-1498. <https://doi.org/10.1002/pmic.201100563>
 37. Nooren IM, Thornton JM (2003) Diversity of protein-protein interactions. *EMBO Journal* 22(14):3486-3492. <https://doi.org/10.1093/emboj/cdg359>
 38. Levy ED, Pereira-Leal JB (2008) Evolution and dynamics of protein interactions and networks. *Current Opinion in Structural Biology* 18(3):349-357. <https://doi.org/10.1016/j.sbi.2008.03.003>
 39. Ivanov AA, Khuri FR, Fu H (2013) Targeting protein-protein interactions as an anticancer strategy. *Trends in Pharmacological Sciences* 34(7):393-400. <https://doi.org/10.1016/j.tips.2013.04.007>
 40. Ezkurdia I, Tress ML (2011) Protein structural domains: definition and prediction. *F1000Research reports on bioinformatics resource development. F1000Research* 2:112. <https://doi.org/10.12688/f1000research.2-112.v1>
 41. Keskin O, Tuncbag N, Gursoy A (2016) Predicting protein-protein interactions from the molecular to the proteome level. *Chemical Reviews* 116(8):4884-4909. <https://doi.org/10.1021/acs.chemrev.5b00683>
 42. Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andrade-Navarro MA, Wanker EE (2011) A directed protein interaction network for investigating intracellular signal transduction. *Science Signaling* 4(189):rs8. <https://doi.org/10.1126/scisignal.2001699>
 43. Wodak SJ, Vlasblom J, Turinsky AL, Pu S (2013) Protein-protein interaction networks: the puzzling riches. *Current Opinion in Structural Biology* 23(6):941-953. <https://doi.org/10.1016/j.sbi.2013.08.002>
 44. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chemical Reviews* 108(4):1225-1244. <https://doi.org/10.1021/cr040409x>
 45. Sarker B, Ritchie DW, Aridhi S (2023) Exploiting graph neural networks to perform finer-grained protein-protein interaction prediction. *PLOS ONE* 18(2):e0270077. <https://doi.org/10.1371/journal.pone.0270077>
 46. Saccenti E, Timmermans MJT (2016) Considerations on the use of the area under the ROC curve as a valid metric in comparative studies of metabolomics data. *Metabolites* 6(1):3. <https://doi.org/10.3390/metabo6010003>
 47. Casadio R, Fariselli P (2015) The state of the art of protein-protein interaction prediction: comparative assessment and perspectives. *Briefings in Bioinformatics* 16(3):364-368. <https://doi.org/10.1093/bib/bbu030>
 48. Bonvin AM (2006) Flexible protein-protein docking. *Current Opinion in Structural Biology* 16(2):194-200. <https://doi.org/10.1016/j.sbi.2006.02.002>
 49. Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. *Nature Reviews Molecular Cell Biology* 7(3):188-197. <https://doi.org/10.1038/nrm1859>
 50. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. *Cell* 144(6):986-998. <https://doi.org/10.1016/j.cell.2011.02.016>