



HAL
open science

AlphaFold2 predicts interactions amidst confounding structural compatibility

Juliette Martin

► **To cite this version:**

Juliette Martin. AlphaFold2 predicts interactions amidst confounding structural compatibility. 2023. ⟨hal-04275604⟩

HAL Id: hal-04275604

<https://hal.science/hal-04275604v1>

Preprint submitted on 8 Nov 2023

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.



HAL Authorization

AlphaFold2 predicts interactions amidst confounding structural compatibility

Juliette Martin

Univ Lyon, CNRS, UMR 5086 MMSB, 7 passage du Vercors F-69367, Lyon, France

Abstract

Predicting physical interactions is one of the holy grails of computational biology, galvanized by rapid advancements in deep learning. AlphaFold2, although not developed with this goal, seems promising in this respect. Here, I test the prediction capability of AlphaFold2 on a very challenging data set, where proteins are structurally compatible, even when they do not interact. AlphaFold2 achieves high discrimination between interacting and non-interacting proteins, and the cases of misclassifications can either be rescued by revisiting the input sequences or can suggest false positives and negatives in the data set. AlphaFold2 is thus not impaired by the compatibility between protein structures and has the potential to be applied at large scale.

Introduction

Prediction of protein-protein interactions has profound implications to suggest functions for uncharacterized proteins, understand protein activity and regulation at the molecular level, and more generally, highlight protein functions in the context of global interactomes. Numerous computational methods have been developed to predict whether or not two proteins physically interact, based on their sequences and 3D structures, see the following references for review ¹⁻⁶.

The formidable capability of AlphaFold2 (AF2) to predict protein 3D structures ⁷ has stimulated the creativity of the scientific community to evaluate what are the application range and limits of AF2 predictions ^{8–20} and extend the tool beyond its initial prediction task ^{18, 21–38}. Prediction of protein-protein complex structures, a task traditionally addressed by protein-protein docking, has rapidly been tackled by modification in the input for the AF2 monomer pipeline ²⁴. A specific model for protein-protein complexes is now available, with breakthrough prediction results ³⁹. Note that in this case, predictions are made with the prior knowledge that the proteins physically interact.

Logically then, the capability of AlphaFold2 to predict interaction between proteins has recently been explored, by using the predicted quality of modeled interfaces as prediction criterion, with very encouraging discrimination capability ^{28, 35, 40, 41}.

In this short article, I challenge AF2 on a particular, presumably difficult data set in which non-interacting protein pairs are special cases, in which the two (non-interacting) proteins are structurally similar to available experimental complexes ⁴². This feature should challenge AF2, since the proteins are compatible in terms of structures. Using the iPTM score of AF2, I found that AF2 is very accurate at discriminating interacting from non-interacting pairs, even in this challenging context, attaining an AUC value of 0.93. Interestingly, model recycling did not improve the discriminative power. The analysis of the few misclassified cases provides suggestions to further improve the discrimination and how to use AF2 for pair screening.

Material and Methods

Data set

Protein pairs from *S cerevisiae* are taken from our previous study ⁴², details about these data set can be found in ⁴² and briefly summarized below.

Interacting protein pairs

The initial data set of interacting protein pairs was extracted from three sources: high confidence physical interactions from BioGrid⁴³, direct interactions from the KUPS resource⁴⁴, and high confidence physical interactions detected by yeast-two-hybrid from Ito et al⁴⁵.

Non-interacting protein pairs

The initial negative data set was extracted from three sources : the negative data set used by Yu et al which are simply sampled from pairs without experimental evidence of interaction⁴⁶, pairs of protein from the KUPS resource⁴⁴, which have no evidence of interaction and also distant GO annotations, and the negative data set built by Trabuco et al⁴⁷ from the Ito data set, where the yeast-two-hybrid data set is used to select proteins pairs without interaction but correctly detected in the experiment .

In our previous study, those pairs were compared with known structures. We had screened homology models of *S. cerevisiae* proteins against a non-redundant database of experimentally known dimers and we had selected pairs where the monomers structurally matched with the experimental dimers (TM score >0.8), and, once superimposed on those dimers, could form an interface of reasonable size (> 20 residues) and without extensive clashes (less than 3 between C α s). Using these criteria resulted in a data set of 22 non-interacting and 222 interacting proteins. In this work, I use the data set of 22 non-interacting proteins and a random sample of 22 interacting pairs, see Table S1.

AlphaFold2 Models

AlphaFold2 predictions are computed using LocalColabFold (<https://github.com/YoshitakaMo/localcolabfold>), a local installation of ColabFold^{39,48}. ColabFold replaces the time-consuming step of multiple sequence alignment (MSA) creation by an ultra-fast step with MMseqs2⁴⁹. No templates are used; models are

not minimized; I tested both with and without model recycling, and different modes of sequence pairing for the MSA: unpaired+paired (default), paired only, and unpaired only.

The resulting 5 models are ranked according to the ipTM computed by AF2 and this score is used as a predictor of protein interaction.

Assessing classification performance

The separation between scores of interacting and non-interacting pairs is measured by the AUC value when using the score to predict interaction. This is done for the AF2 ipTM score and the pDockQ, recently introduced by Bryant et al ⁴⁰, which is derived from the pLDDT scores of interface residues. Statistical significance between AUC values is assessed using the non-parametric DeLong's test ⁵⁰ implemented in the pROC package ⁵¹.

The classification performance is measured by the accuracy, i.e., percentage of correctly classified pairs. Statistical difference between accuracies is assessed using the MacNemar test ⁵².

Results

Although AF2 has recently been used to discriminate interacting from non-interacting pairs with promising results ^{28, 35, 40, 41}, it is always worthy of pushing the system to the limits to better know its applicability range. Here, I propose to further test AF2 prediction capability in extreme conditions. I submitted to AF2 prediction a particularly challenging data set from a previous study ⁴². In this data set, all the pairs are supported by structural data: interacting but also non-interacting pairs are compatible in shape, as assessed by their high similarity to experimental dimers, as explained in the Methods section.

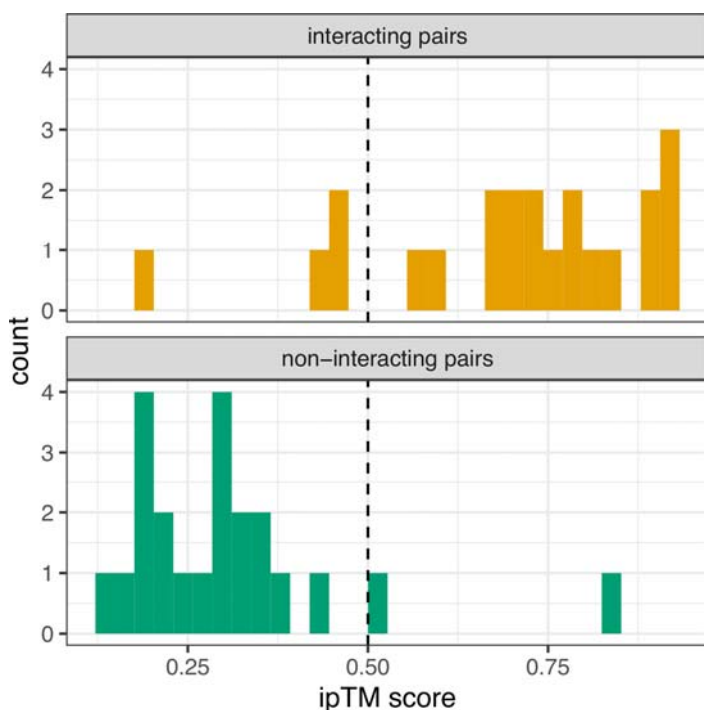


Figure 1. ipTM score distribution for interacting and non-interacting pairs. The vertical dashed line indicates the prediction cutoff (0.5).

The results of AF2 prediction on this challenging data set are impressive, attaining AUC value in the 0.86-0.93 range, see Figure 1, and accuracy around 80-86%. These results are comparable to the recent result of Bryant et al who obtained an AUC equal to 0.87 using the monomer AF2 pipeline on *E. coli* proteins⁴¹. Detailed results are shown in Tables S2 and S3 for each AF2 model, with and without recycling, and different MSA pairing modes. Unfortunately, the limited size of the data set does not allow to statistically differentiate those different settings: the predictions are uniformly good. AF2 ipTM score outperforms the pdockQ score in terms of separation (DeLong's test p-value=0.011, see Table S3). Considering the predictions obtained without recycling, default MSA pairing and the best model out of five according to the ipTM score, 6 pairs are misclassified, corresponding to an accuracy equal to 86%. The misclassified cases are shown in Figures 2 and 3 and discussed below.

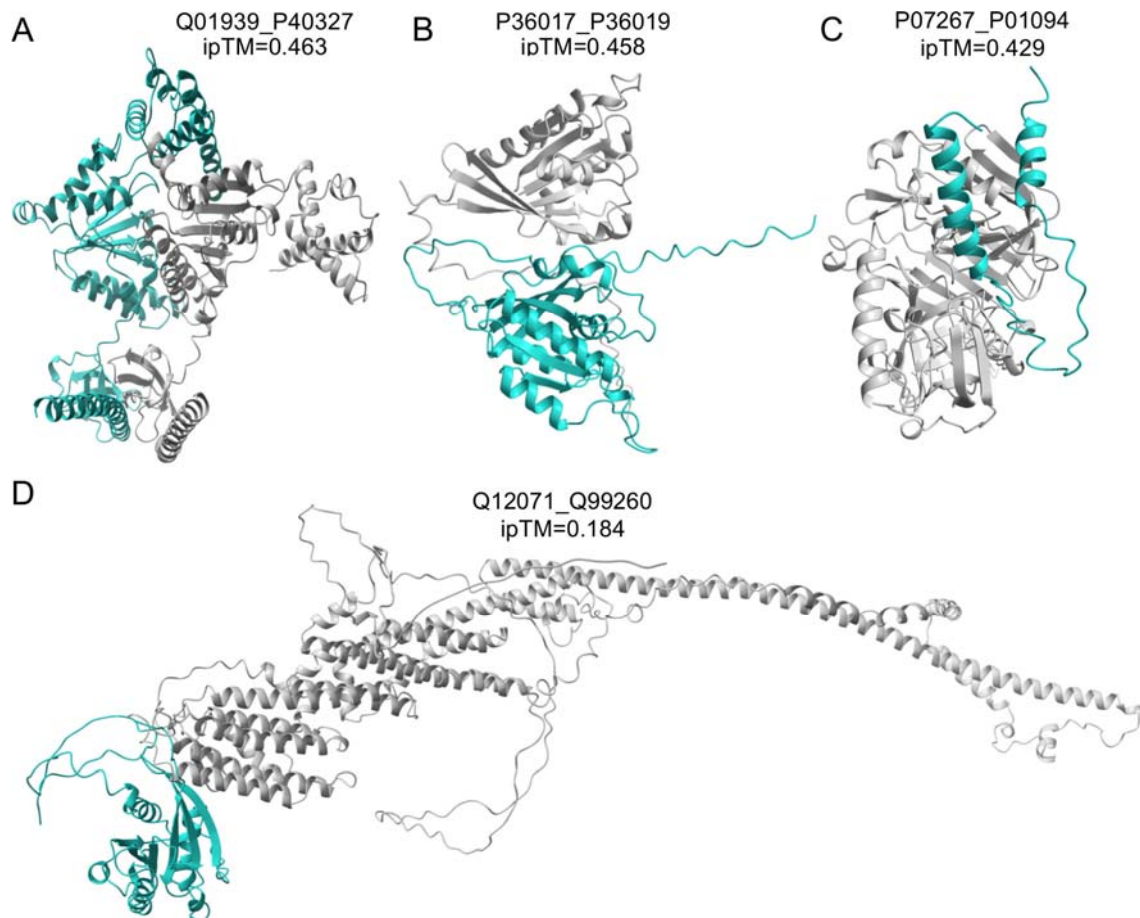


Figure 2. AF2 models of interacting pairs misclassified as non-interacting ($ipTM < 0.5$). A: pair formed by the subunits 8 (Q01939, in blue) and 4 (P40327, in grey) of the 26s proteasome. B: pair formed by the vacuolar protein sorting-associated protein 21 (P36017 in grey) and the GTP-binding protein YPT53 (P36019, in blue). C: pair formed by saccharopepsin (P07267, in grey) and its inhibitor (P01094, in blue). D: pair formed by the vacuolar protein sorting-associated protein 54 (Q12071, in grey) and the GTP-binding protein YPT6 (Q99260, in blue).

Four interacting pairs are incorrectly classified as non-interacting ($ipTM$ score < 0.5). The first case is the dimer formed by the regulatory subunit 8 homolog and subunit 4 homolog (Uniprot ids Q01939 and P40327) of the 26S proteasome, which is an assembly of 47 protein chains. The dimer obtains an $ipTM$ score equal to 0.463, close to the interaction

cutoff, see Figure 2A. The comparison of the AF2 model with the experimental structure of the 26S proteasome (PDB id 3JCO) reveals that the AF2 model is structurally similar to the dimer formed by the subunits 7 and 4, meaning that AF2 wrongly placed the subunit 8 in place of the subunit 7, see Figure S1. This is possible because subunits 8 and 7 are structurally similar (TM score = 0.6 between experimental structures). So, in this case, AF2 prediction was confused by the presence of another interacting chain with similar structure.

The pair formed by the vacuolar protein sorting-associated protein 21 and the GTP-binding protein YPT53 (Uniprot ids P36017 and P36019) obtains an ipTM score equal to 0.458, close to the prediction cutoff. Furthermore, the PAE matrix displays low values between the two chains, suggesting a good confidence in the relative orientation of the two protein chains. The AF2 model, shown in Figure 2B, has an interface involving the disorder C-tail of the GTP protein, which could explain the low score. Interestingly, the model obtained with recycling has a good ipTM score (0.616) with a similar configuration but without the disordered part at the interface (see Figure S2). However, model recycling significantly increases the computation time, which is a limiting factor in the perspective of pair screening. Alternatively to model recycling, the prediction can be run with the disordered part chopped from the sequence, which produces a model similar to the recycled one, with an ipTM score equal to 0.63 (see Figure S2). So, in that case, it is possible to ‘rescue’ the prediction by chopping the sequence.

The pair formed by the saccharopepsin and its inhibitor (Uniprot ids P07267 and P01094) obtains an ipTM score equal to 0.429, see Figure 2C. The full-length sequence of the saccharopepsin contains an N-terminal propeptide of 75 residues that is cleaved upon activation of the enzyme⁵³. The comparison of the AF2 model with the experimental structure (PDB id 3COJ) reveals that the binding cleft where the inhibitor is supposed to bind is occluded by the N-terminal region of the enzyme corresponding to the propeptide. Re-running the prediction after chopping the propeptide sequence results in a model with a good ipTM score equal to 0.63 and in good agreement with the experimental structure (see Figure

S3). So, in this case also, it is possible to rescue the prediction with appropriate sequence chopping.

The pair formed by the vacuolar protein sorting-associated protein 54 and the GTP-binding protein YPT6 (Uniprot ids Q12071 and Q99260) obtains a very low ipTM score equal to 0.184, see Figure 2D. The 5 models are drastically different from each other, with even drastic situations where the proteins are not in contact in two of the models (see Figure S4). There is evidence of physical interaction between these proteins, as detected by affinity purification. However, there is no evidence of direct physical interaction by two-hybrid assay. So there is the possibility that this pair is in fact a false positive case.

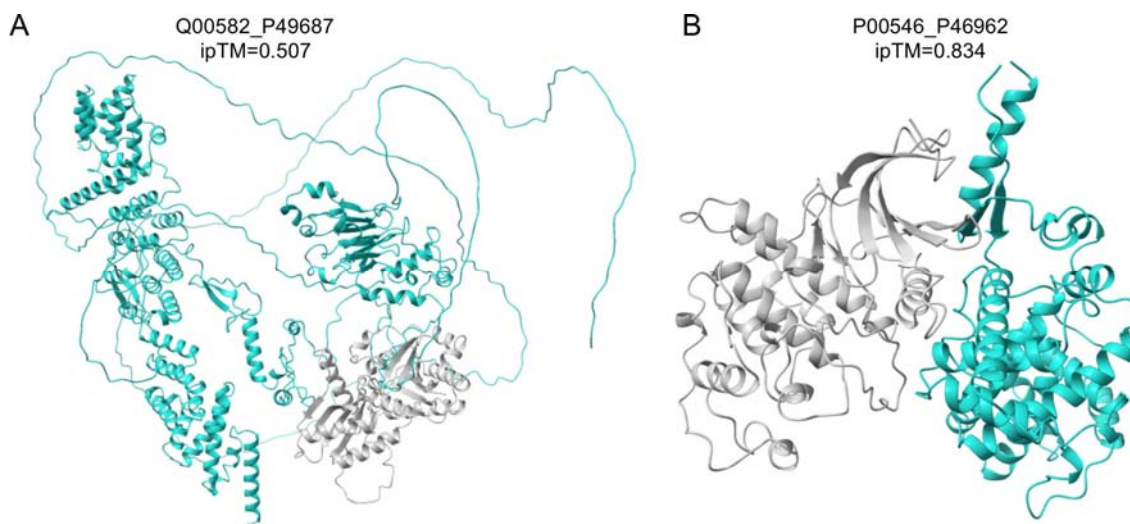


Figure 3: AF2 models of non-interacting pair misclassified as interacting (ipTM>0.5). A: pair formed by the GTP-binding protein GTR1 (Q00582, in grey) and the nucleoporin NUP145 (P49687, in blue). B: pair formed by cyclin-dependent kinase 1 (P00546, in grey) and the CTD kinase subunit beta (P46962, in blue)

Two pairs of non-interacting proteins are incorrectly classified as interacting. The pair between GTP-binding protein GTR1 and nucleoporin NUP145, a component of the nuclear pore complex (Uniprot id Q00582 and P49687) obtains a border line ipTM score equal to

0.507, see Figure 1F. The examination of the AF2 model reveals that the nucleoporin has a disordered N-terminal region that mediates the interaction with GTR1. After chopping the disordered parts of the nucleoporin, the best model obtains an ipTM score equal to 0.23. So in this case, the initial ipTM score was meaningless because of the disordered parts, and the interaction can be excluded by adequate sequence chopping.

The pair formed by the cyclin-dependent kinase 1 and the CTD kinase subunit beta (Uniprot id P00546 and P46962) achieves a very high ipTM score equal to 0.834, see Figure 1G. In this case, all models have high ipTM scores (>0.8) and are structurally similar (data not shown). Although there is no evidence of direct interaction between these two proteins, they are reported as interacting in the STRING resource⁵⁴, having, among other things, a direct interaction between homologs in *Drosophila* measured by yeast-two-hybrid assay⁵⁵. This suggests that this case could be a false negative pair.

In summary, out of 6 misclassified cases, three could be corrected by sequence chopping, two can be questioned as false positive/negative, and one highlights a phenomenon of confusion between chains in interfaces in a macromolecular assembly.

Conclusion

A reduced but challenging data set was submitted to AF2 in order to discriminate interacting from non-interacting pairs, resulting in very high prediction accuracy. Several misclassified cases could be rescued by appropriate sequence chopping, and some others are suggestive of incorrect annotations (false positive or false negative). A potential limitation of AF2 was observed in a case where several protein chains with structural similarity form a supra-molecular assembly. The fact that no recycling is required opens the possibility to apply this procedure at large scale. To conclude, AF2 seems a promising technology for predicting protein-protein interactions, even capable of discriminating interacting from non-interacting pairs in presence of confounding structural compatibility.

References

1. Hu X, Feng C, Ling T, Chen M. Deep learning frameworks for protein–protein interaction prediction. *Comput Struct Biotechnol J* 2022;20:3223–3233.
2. Dunham B, Ganapathiraju MK. Benchmark Evaluation of Protein-Protein Interaction Prediction Algorithms. *Mol Basel Switz* 2021;27(1):41.
3. Casadio R, Martelli PL, Savojardo C. Machine learning solutions for predicting protein–protein interactions. *WIREs Comput Mol Sci*;n/a(n/a):e1618.
4. Zahiri J, Bozorgmehr JH, Masoudi-Nejad A. Computational Prediction of Protein–Protein Interaction Networks: Algo-rithms and Resources. *Curr Genomics* 2013;14(6):397–414.
5. Shoemaker BA, Panchenko AR. Deciphering Protein–Protein Interactions. Part II. Computational Methods to Predict Protein and Domain Interaction Partners. *PLoS Comput Biol* 2007;3(4):e43.
6. Wass MN, David A, Sternberg MJ. Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol* 2011.
7. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–589.
8. Johansson-Åkhe I, Wallner B. Benchmarking Peptide-Protein Docking and Interaction Prediction with AlphaFold-Multimer. 2021 November 18. 2021.11.16.468810 p.
9. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A, Zhu W, Kundrotas P, Serra VR, Rodrigues CHM, Dunham AS, Burke D, Borkakoti N, Velankar S, Frost A, Basquin J, Lindorff-Larsen K, Bateman A, Kajava AV, Valencia A, Ovchinnikov S, Durairaj J, Ascher DB, Thornton JM, Davey NE, Stein A, Elofsson A, Croll TI, Beltrao P. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* 2022;29(11):1056–1067.
10. Pozzati G, Zhu W, Bassot C, Lamb J, Kundrotas P, Elofsson A. Limits and potential of combined folding and docking. *Bioinformatics* 2021:btab760.
11. Saldaño T, Escobedo N, Marchetti J, Zea DJ, Donagh JM, Rueda AJV, Gonik E, Melani AG, Nechcoff JN, Salas MN, Peters T, Demitroff N, Alberti SF, Palopoli N, Fornasari MS, Parisi G. Impact of protein conformational diversity on AlphaFold predictions. 2021 October 28. 2021.10.27.466189 p.
12. Fowler NJ, Williamson MP. The accuracy of protein structures in solution determined by AlphaFold and NMR. *Structure* 2022;30(7):925-933.e2.
13. Chakravarty D, Porter LL. AlphaFold2 fails to predict protein fold switching. *bioRxiv*; 2022. p 2022.03.08.483439.
14. Sawicki LR, Benitez G, Carletti M, Palopoli N, Fornasari MS, Parisi G. Conformational epistasis impairs AlphaFold structural predictions. *bioRxiv*; 2022. p 2022.11.15.516638.
15. Wong F, Krishnan A, Zheng EJ, Stärk H, Manson AL, Earl AM, Jaakkola T, Collins JJ. Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery. *Mol Syst Biol* 2022;18(9):e11081.
16. Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci* 2022;31(8):e4379.
17. Holcomb M, Chang Y-T, Goodsell DS, Forli S. Evaluation of AlphaFold2 structures as docking targets. *Protein Sci Publ Protein Soc* 2023;32(1):e4530.
18. Bryant P, Pozzati G, Zhu W, Shenoy A, Kundrotas P, Elofsson A. Predicting the structure of large protein complexes using AlphaFold and sequential assembly. *bioRxiv*; 2022. p 2022.03.12.484089.
19. McCafferty CL, Pennington EL, Papoulas O, Taylor DW, Marcotte EM. Does AlphaFold2 model proteins' intracellular conformations? An experimental test using cross-linking

- mass spectrometry of endogenous ciliary proteins. *bioRxiv*; 2022. p 2022.08.25.505345.
20. Martin J. When AlphaFold2 predictions go wrong for protein–protein complexes, is there something to be learnt? *Q Rev Biophys* 2022;55:e6.
 21. Bruley A, Bitard-Feildel T, Callebaut I, Duprat E. A sequence-based foldability score combined with AlphaFold2 predictions to disentangle the protein order/disorder continuum. *Proteins Struct Funct Bioinforma*;n/a(n/a).
 22. Jussupow A, Kaila VRI. Effective Molecular Dynamics from Neural- Network Based Structure Prediction Models. :37.
 23. Alamo D del, Sala D, Mchaourab HS, Meiler J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* 2022;11:e75751.
 24. Ghani U, Desta I, Jindal A, Khan O, Jones G, Kotelnikov S, Padhorny D, Vajda S, Kozakov D. Improved Docking of Protein Models by a Combination of AlphaFold2 and ClusPro. *bioRxiv* 2021:2021.09.07.459290.
 25. Roney JP, Ovchinnikov S. State-of-the-art estimation of protein model accuracy using AlphaFold. *bioRxiv*; 2022. p 2022.03.11.484043.
 26. Sala D, Hildebrand PW, Meiler J. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Front Mol Biosci* 2023;10.
 27. Stein RA, Mchaourab H. Modeling Alternate Conformations with AlphaFold2 via Modification of the Multiple Sequence Alignment. 2021 November 29. 2021.11.29.470469 p.
 28. Yu D, Chojnowski G, Rosenthal M, Kosinski J. AlphaPulldown – a Python package for protein-protein interaction screens using AlphaFold-Multimer. *bioRxiv*; 2022. p 2022.08.05.502961.
 29. Andorf CM, Sen S, Hayford RK, Portwood JL, Cannon EK, Harper LC, Gardiner JM, Sen TZ, Woodhouse MR. FASSO: An AlphaFold based method to assign functional annotations by combining sequence and structure orthology. *bioRxiv*; 2022. p 2022.11.10.516002.
 30. Monzon V, Paysan-Lafosse T, Wood V, Bateman A. Reciprocal Best Structure Hits: Using AlphaFold models to discover distant homologues. *bioRxiv*; 2022. p 2022.07.04.498216.
 31. Baltzis A, Mansouri L, Jin S, Langer BE, Erb I, Notredame C. Improving sequence alignments with AlphaFold2 regardless of structural modeling accuracy. *bioRxiv*; 2022. p 2022.05.24.492699.
 32. Bordin N, Sillitoe I, Nallapareddy V, Rauer C, Lam SD, Waman VP, Sen N, Heinzinger M, Littmann M, Kim S, Velankar S, Steinegger M, Rost B, Orengo C. AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms. *bioRxiv*; 2022. p 2022.06.02.494367.
 33. Brems MA, Runkel R, Yeates TO, Virnau P. AlphaFold predicts the most complex protein knot and composite protein knots. *Protein Sci* 2022;31(8):e4380.
 34. Wayment-Steele HK, Ovchinnikov S, Colwell L, Kern D. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv*; 2022. p 2022.10.17.512570.
 35. Gao M, Nakajima An D, Parks JM, Skolnick J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun* 2022;13(1):1744.
 36. Chang L, Perez A. Ranking Peptide Binders by Affinity with AlphaFold**. *Angew Chem Int Ed* 2023;62(7):e202213362.
 37. Tsaban T, Varga JK, Avraham O, Ben-Aharon Z, Khramushin A, Schueler-Furman O. Harnessing protein folding neural networks for peptide-protein docking. *Nat Commun* 2022;13(1):176.
 38. Wallner B. AFsample: Improving Multimer Prediction with AlphaFold using Aggressive Sampling. *bioRxiv*; 2022. p 2022.12.20.521205.
 39. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstern S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021:2021.10.04.463034.

40. Bryant P, Noe F. Rapid protein-protein interaction network creation from multiple sequence alignments with Deep Learning. *bioRxiv*; 2023. p 2023.04.15.536993.
41. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 2022;13(1):1265.
42. Launay G, Ceres N, Martin J. Non-interacting proteins may resemble interacting proteins: prevalence and implications. *Sci Rep* 2017;7:40419.
43. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci Publ Protein Soc* 2021;30(1):187–200.
44. Chen X -w., Jeong JC, Dermeyer P. KUPS: constructing datasets of interacting and non-interacting protein pairs with associated attributions. *Nucleic Acids Res* 2010(Database Issue):D750-754.
45. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci* 2001;98(8):4569–4574.
46. Yu J, Guo M, Needham CJ, Huang Y, Cai L, Westhead DR. Simple sequence-based kernels do not predict protein-protein interactions. *Bioinformatics* 2010;26(20):2610–2614.
47. Trabuco LG, Betts MJ, Russell RB. Negative protein-protein interaction datasets derived from large-scale two-hybrid experiments. *Methods San Diego Calif* 2012;58(4):343–348.
48. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;19(6):679–682.
49. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35(11):1026–1028.
50. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
51. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
52. Patrick Walters W. Comparing classification models—a practical tutorial. *J Comput Aided Mol Des* 2022;36(5):381–389.
53. Dreyer T, Halkier B, Svendsen I, Ottesen M. Primary structure of the aspartic proteinase A from *Saccharomyces cerevisiae*. *Carlsberg Res Commun* 1986;51(1):27.
54. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021;49(D1):D605–D612.
55. Stanyon CA, Liu G, Mangiola BA, Patel N, Giot L, Kuang B, Zhang H, Zhong J, Finley RL. A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol* 2004;5(12):R96.