*Bioinformatics Advances*, 2025, **00**, vbaf090
https://doi.org/10.1093/bioadv/vbaf090
Advance Access Publication Date: 24 April 2025
**Application Note**

Structural bioinformatics

# PPIFold: a tool for analysis of protein–protein interaction from AlphaPullDown

**Quentin Rouger[1], Emmanuel Giudice[1], Damien F. Meyer[2,3,*], Kévin Macé[1,*]**

[1]Univ. Rennes, CNRS, Institut de Génétique et Développement de Rennes (IGDR)—UMR6290, Rennes 35000, France
[2]CIRAD, UMR ASTRE, Petit-Bourg, Guadeloupe 97170, France
[3]ASTRE, University Montpellier, CIRAD, INRAE, Montpellier 34398, France

*Corresponding authors. Damien F. Meyer, CIRAD, UMR ASTRE, Petit-Bourg, Guadeloupe 97170, France. E-mail: damien.meyer@cirad.fr; Kévin Macé, Univ. Rennes, CNRS, Institut de Génétique et Développement de Rennes (IGDR)—UMR6290, Rennes 35000, France. E-mail: kevin.mace@univ-rennes.fr.

Associate Editor: Alex Bateman

## Abstract

**Motivation:** Protein structure and protein–protein interaction (PPI) predictions based on coevolution have transformed structural biology, but managing pre-processing and post-processing can be complex and time-consuming, making these tools less accessible.

**Results:** Here, we introduce PPIFold, a pipeline built on the AlphaPulldown Python package, designed to automate file handling and streamline the generation of outputs, facilitating the interpretation of PPI prediction results. The pipeline was validated on the bacterial Type 4 Secretion System nanomachine, demonstrating its effectiveness in simplifying PPI analysis and enhancing accessibility for researchers.

**Availability and implementation:** PPIFold is implemented as a pip package and available at: https://github.com/Qrouger/PPIFold.

## 1 Introduction

Artificial intelligence has revolutionized the field of structural biology with software such as AlphaFold2 (Jumper *et al.* 2021) and RoseTTAFold (Baek *et al.* 2021). In addition to predicting structure, these methods have proven very effective in predicting protein-protein interaction (PPI) (Humphreys *et al.* 2024). In the context of high-throughput PPI prediction, the process often involves a series of redundant and automatable steps, along with essential verification stages to ensure interpretable and reliable results. The AlphaPulldown tool (Yu *et al.* 2023) exemplifies this approach but requires considerable manual intervention to achieve high-confidence predictions. To address these challenges, we present PPIFold, a Python-based tool designed to streamline PPI prediction by minimizing redundant steps and reducing the likelihood of inaccurate results. PPIFold is a more intuitive and specialized tool for PPI analysis, incorporating new interaction scoring metrics and automating the generation of detailed interaction evaluation reports. This enhanced functionality allows for a more comprehensive assessment of predicted PPIs, streamlining the interpretation process and providing deeper insights into interaction dynamics, making it a powerful tool for PPI-focused research.

PPIFold offers an accessible solution for researchers, particularly those without extensive bioinformatics expertise, enabling them to efficiently predict PPIs by eliminating unnecessary steps and rapidly generating the data and figures necessary for publication. The tool is built upon the AlphaPulldown package, which leverages AlphaFold Multimer (Evans *et al.* 2021) for extensive PPI screening across diverse protein combinations. Additionally, this pipeline allows for the separation of feature generation (handled by CPUs) from model generation (handled by GPUs), further optimizing the computational workflow.

## 2 Pipeline description

The PPIFold pipeline is designed for the identification or validation of PPI, including homo-oligomers (Fig. 1). PPIFold is structured into three main modules. First, protein sequences undergo automatic cleaning and verification. Next, the pipeline predicts all potential PPIs using the AlphaPulldown tool. Finally, the predicted interactions are scored, and detailed figures are generated for the PPIs of interest, facilitating a comprehensive analysis and visualization of the results. To assess its performance, the pipeline was applied to the bacterial Type 4 Secretion System nanomachine, showcasing its ability to streamline PPI predictions and provide user-friendly, interpretable results (Supplementary Information).

### 2.1 Sequence and verification

Inputs sequences are analyzed using SignalP5 (Almagro Armenteros *et al.* 2019) to identify the presence of signal peptides characteristic of the organism, and if yes, the sequences are removed from the original protein sequences to obtain the mature protein structure post-translocation into the periplasm. Next, the feature generation and multiple sequence alignment (MSA) for all proteins are performed using MMseq2 or HMMER *via* AlphaPulldown package. Then PPIFold generates an MSA depth figure to assess the quality

**Received:** January 10, 2025; **Revised:** March 28, 2025; **Editorial Decision:** April 2, 2025; **Accepted:** April 22, 2025

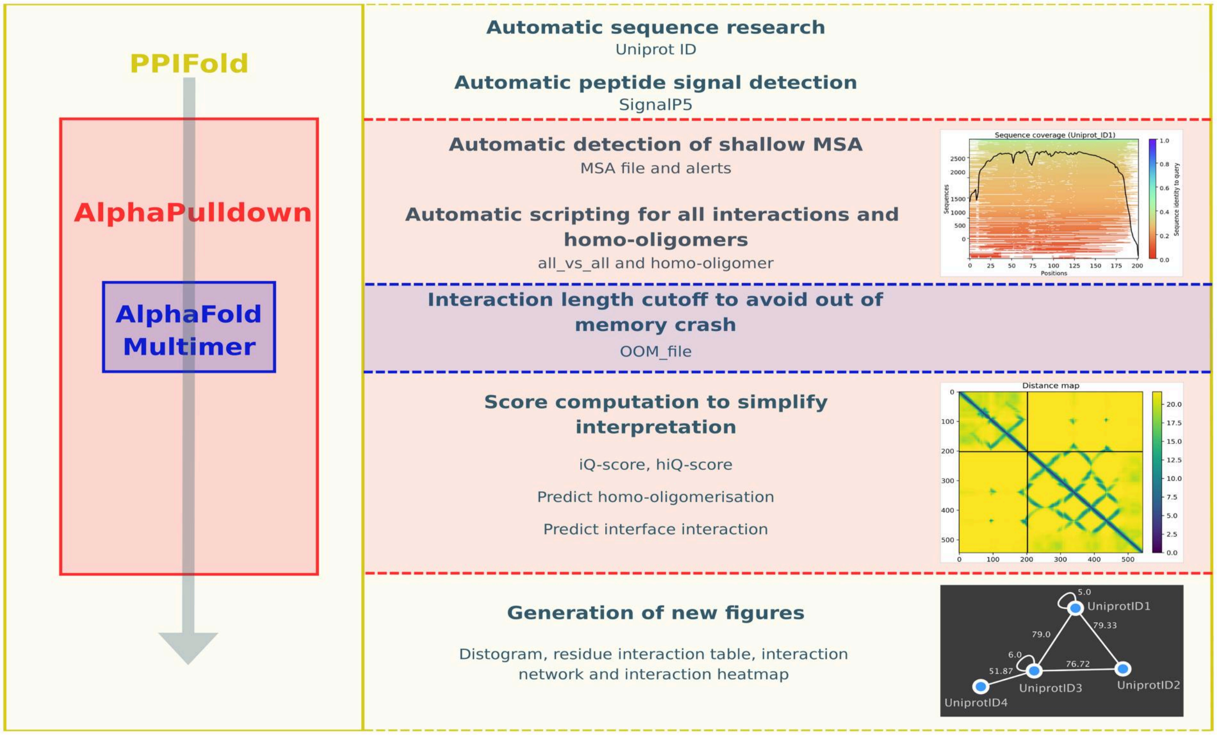
PPIFold
AlphaPulldown
AlphaFold Multimer
Automatic sequence research
Uniprot ID
Automatic peptide signal detection
SignalP5
Automatic detection of shallow MSA
MSA file and alerts
Automatic scripting for all interactions and homo-oligomers
all_vs_all and homo-oligomer
Sequence coverage (Uniprot_ID1)
Sequences
Positions
Sequence identity to query
Interaction length cutoff to avoid out of memory crash
OOM_file
Score computation to simplify interpretation
iQ-score, hiQ-score
Predict homo-oligomerisation
Predict interface interaction
Distance map
Generation of new figures
Distogram, residue interaction table, interaction network and interaction heatmap
UniprotID1
UniprotID2
UniprotID3
UniprotID4
5.0
79.33
79.0
6.0
76.72
51.87

**Figure 1.** Depiction of the PPIFold pipeline. The PPIFold pipeline consists of three major modules. The first component involves sequence generation and verification, including the creation of scripts for AlphaPulldown. The second component focuses on model generation using AlphaPulldown, which provides initial scores for predicted PPIs. The final component involves calculating refined scores and generating various figures and tables for data visualization and interpretation.

of the alignment, as a shallow MSA can negatively impact prediction accuracy, the alignment depth is adequate to capture co-evolutionary signals, which are crucial for predicting both intra- and inter-protein interactions. Proteins with poor MSA quality are recorded in a file named shallow_MSA.txt, and predictions for these proteins are flagged as unreliable for validating or invalidating PPIs.

Additionally, PPIFold provides two .txt files of interaction prediction to do, all-against-all and homo-oligomers, tailored according to the available GPU memory (https://www.rbvi.ucsf.edu/chimerax/data/alphafold-jan2022/afspeed.html). Indeed, to prevent out-of-memory (OOM) errors, interactions too large to model are excluded, and these cases are listed in the OOM_int.txt file, allowing users to track interactions that cannot be processed due to memory constraints.

## 2.2 Structure prediction by AlphaPulldown

The custom mode and homo-oligomer functions from AlphaPulldown tool *via* AlphaFold Multimer are utilized to generate structural models and assign various scores to rank predicted PPIs. For each interaction, five different models are generated and ranked based on their ipTM scores (Gao *et al.* 2022), with only the highest-ranked model undergoing further scoring. This step, which is the most time-intensive, is GPU-dependent. quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

## 2.3 Scores

Interaction within hetero-oligomeric models are evaluated using three distinct scores generated by AlphaPulldown tool: pi-score (Malhotra *et al.* 2021), ipTM_pTM scores, and pDockQ (Bryant *et al.* 2022). On the other hand, interaction within homo-oligomeric models are assessed using two scores: pi-score and ipTM_pTM. While each score provides valuable insights into a PPI, relying on a single score is insufficient, moreover evaluating multiple scores can be difficult and time-consuming. To address this, we introduced the iQ-score, a weighted linear combination of these individual scores, with an emphasis on the pi-score. The iQ-score was designed to improve the selection of relevant interactions by integrating three existing metrics from AlphaPulldown, with two main objectives: (i) Facilitating result interpretation by providing a single score metric, the iQ-score simplifies analysis for non-expert users while retaining individual scores in the output files for transparency; and (ii) Enhancing interaction assessment, as relying on a single metric may sometimes lead to misclassification of interactions if the score falls below or above a defined threshold. By combining three scores, the iQ-score mitigates this risk and provides a more balanced assessment. Similarly, for homo-oligomerization predictions, we introduce the hiQ-score.

$$\text{iQ-score} = \Big((\text{pi-score}+2.63)/5.26\Big)*40+\text{iptm\_ptm}*30+\text{pDockQ}*30$$

$$\text{iQ-score}_{\text{cutoff}} >= \Big((0.05+2.63)/5.26\Big)*40+0.5*30+0.5*30$$

$$\text{iQ-score}_{\text{cutoff}} >= 50.38$$

$$\text{hiQ-score} = \Big(\Big(\big(\textstyle\sum \text{pi-score}\big)/\text{n}+2.63\Big)/5.26\Big)*60+\text{iptm\_ptm}*40$$

$$\text{hiQ-score}_{\text{cutoff}} >= \Big((0.05+2.63)/5.26\Big)*60+0.5*40$$

$$\text{hiQ-score}_{\text{cutoff}} >= 50.57$$

These scores are calculated exclusively for the best model (highest ipTM) and for interactions with a predicted alignment error (PAE) above a threshold of 10, which is directly influenced by the depth of the multiple sequence alignment (MSA). All results are sorted based on specific cutoff values (iQ-score and hiQ-score >50), ensuring that only the most probable interactions are retained. Since the iQ-score is based on three pre-existing metrics, its cutoff value was determined based on previously published studies (Malhotra *et al.* 2021, Bryant *et al.* 2022, Gao *et al.* 2022). Unlike other metrics such as ipTM_pTM or pi-score, which can have varying scales, including negative values, the iQ and hiQ-scores provide a more intuitive interpretation, particularly for non-specialists. Their consistent scale from 0 (lowest) to 100 (highest), with a threshold set at 50 to identify the most probable interactions, makes them easier to understand and use.

### 2.3.1 Benchmarking of the iQ-score

In addition, we carried out a benchmarking analysis using publicly available datasets (Supplementary Information). This benchmark indicates that the iQ-score is more sensitive than pDockQ but less specific, and it is also more sensitive than pi-score and ipTM_pTM, though still less specific. The iQ-score shows the highest percentage of true positives among the predicted positives but has a lower percentage of true negatives than pi-score and ipTM_pTM. Minimizing false positive and maximizing true negative predictions remains a crucial challenge in PPI prediction to ensure accurate results. Beyond facilitating the interpretation and use of interaction scores, the iQ-score proves to be the most effective in limiting false positives in interaction predictions. It is worth noting that the score is entirely independent of any optimization.

## 2.4 Figures generation

Once the filtering is complete, PPIFold performs further analysis and generates key visual outputs, including a distogram figure, a residue interaction table, a heatmap, and a protein interaction network figure (Supplementary Information).

### 2.4.1 Distogram

The distogram figure provides a rapid and schematic visualization of the interface involved in the PPI. It represents interactions into two proteins where the value on the left indicates the protein sequence length in amino acids. Points near the diagonal symmetry line, along with pixels in black squares, represent residues in contact within the same protein, while points outside this region denote residues in contact between different proteins. Colors correspond to the distance between residue pairs, measured in Ångströms, with dark blue points indicating shorter distances. This distogram is generated directly from the PDB file and is displayed only for PPIs that have passed the cutoff scores.

### 2.4.2 Residue interaction table

For each interaction, all residues in contact at the interface are listed in the table, providing a more detailed representation of the distogram. This additional information is particularly useful for designing mutations for wet lab validation. The table is generated exclusively for the top-ranked PPIs that meet the cutoff criteria.

### 2.4.3 Heatmap and interaction network

The iQ-score heatmap allows a comprehensive visualization of interaction scores for all proteins, highlighting those with either low or high average scores. The interaction network figure represents all proteins that have at least one interaction with other proteins within the system indicated by iQ-score. It also illustrates the homo-oligomerization of each protein indicated by hiQ-score. Thus, the interaction network provides a comprehensive overview of the PPIs, facilitating the generation of hypotheses regarding the system's structure and function. By examining the network, researchers can explore functional relationships and quickly identify potential incompatibilities between interactions. These observations can be further supported by detailed analyses of interaction interfaces on the PDB models, such as when two proteins interact at the same interface area of a third partner. The scoring system serves as an indicator of interaction strength, offering a valuable tool for predicting the stability and likelihood of protein interactions within the system.

## 3 Conclusion

PPIFold builds on the strengths of AlphaPulldown and AlphaFold Multimer, automating and simplifying both pre- and post-processing steps for large-scale PPI predictions. By streamlining the workflow, PPIFold reduces manual intervention, making PPI analysis more efficient and accessible to a broader community of researchers. This pipeline accelerates the *in silico* exploration of PPIs, offering a fast and efficient tool for generating biological hypotheses, improving experimental design, and contributing to a deeper understanding of molecular interactions. Moreover, PPIFold stands out as an essential resource for researchers seeking to harness cutting-edge computational power, ultimately enabling groundbreaking discoveries in structural and systems biology. Looking ahead, PPIFold is designed with the flexibility to be adapted to AlphaFold3 or any future prediction software advanced.

## Acknowledgements

The authors thank Alix Regnier for reviewing the code.

## Author contributions

Quentin Rouger (Conceptualization [equal], Software [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Visualization [lead], Writing – original draft [lead], Writing – review & editing [lead]), Emmanuel Giudice (Supervision [supporting], Writing – review & editing [supporting]), Damien F. Meyer (Conceptualization [equal], Investigation [equal], Methodology [equal], Supervision [lead], Writing – original draft [equal], Writing – review and editing [equal]), Kévin Macé (Conceptualization [equal], Project administration [lead], Funding acquisition [lead], Supervision [lead], Writing – review & editing [equal]).

## Supplementary data

Supplementary data are available at *Bioinformatics Advances* online.

Conflict of interest: None declared.

## Funding

This work was supported by France 2030 under the Agence Nationale de la Recherche [ANR-22-PAMR-0005]; the ANR Tremplin-ERC VIRULENSSE; and the Centre National de la Recherche Scientifique (CNRS).

## Data availability

The data underlying this article are available in *Github* at https://github.com/Qrouger/PPIFold.

## References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019;**37**:420–3.

Baek M, DiMaio F, Anishchenko I *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;**373**:871–6.

Bryant P, Pozzati G, Elofsson A. Improved prediction of protein–protein interactions using AlphaFold2. *Nat Commun* 2022;**13**:1265.

Evans R, O'Neil M, Pritzel A *et al.* Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021, https://doi.org/10.1101/2021.10.04.463034, preprint: not peer reviewed.

Gao M, Nakajima An D, Parks JM *et al.* AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun* 2022;**13**:1744.

Humphreys IR, Zhang J, Baek M *et al.* Protein interactions in human pathogens revealed through deep learning. *Nat Microbiol* 2024;**9**:2642–52.

Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9.

Malhotra S, Joseph AP, Thiyagalingam J *et al.* Assessment of protein–protein interfaces in cryo-EM derived assemblies. *Nat Commun* 2021;**12**:3399.

Yu D, Chojnowski G, Rosenthal M *et al.* AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics* 2023;**39**:btac749.

Bioinformatics Advances, 2025, 00, 1–4
https://doi.org/10.1093/bioadv/vbaf090
Application Note