Bioinformatics, 39(1), 2023, btac749
https://doi.org/10.1093/bioinformatics/btac749
Advance Access Publication Date: 22 November 2022
Applications Note

Structural bioinformatics

AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer

Dingquan Yu[1,2], Grzegorz Chojnowski[1], Maria Rosenthal [3] and Jan Kosinski [1,2,4,*]

[1]European Molecular Biology Laboratory Hamburg, Hamburg 22607, Germany, [2]Centre for Structural Systems Biology (CSSB), Hamburg 22607, Germany, [3]Bernhard Nocht Institute for Tropical Medicine, Hamburg 20359, Germany and [4]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Received on August 5, 2022; revised on October 13, 2022; editorial decision on November 12, 2022; accepted on November 21, 2022

Abstract

Summary: The artificial intelligence-based structure prediction program AlphaFold-Multimer enabled structural modelling of protein complexes with unprecedented accuracy. Increasingly, AlphaFold-Multimer is also used to discover new protein–protein interactions (PPIs). Here, we present AlphaPulldown, a Python package that streamlines PPI screens and high-throughput modelling of higher-order oligomers using AlphaFold-Multimer. It provides a convenient command-line interface, a variety of confidence scores and a graphical analysis tool.

Availability and implementation: *AlphaPulldown* is freely available at https://www.embl-hamburg.de/AlphaPulldown.

Contact: jan.kosinski@embl.de

Supplementary information: Supplementary note is available at *Bioinformatics* online.

1 Introduction

AlphaFold2 (Jumper *et al.*, 2021) and AlphaFold-Multimer (Evans *et al.*, 2022) have enabled structural modelling of monomeric proteins and protein complexes with accuracy comparable to experimental structures. Various modifications of AlphaFold2 have been developed to facilitate specific applications, such as ColabFold (Mirdita *et al.*, 2022), which accelerates the program and exposes useful parameters. AlphaFold-Multimer can also be applied to screen large datasets of proteins for new protein–protein interactions (PPIs) (Bryant *et al.*, 2022a; Humphreys *et al.*, 2021) and to model combinations of proteins and their fragments when modelling complexes (Mosalaganti *et al.*, 2022). To streamline such applications, we developed AlphaPulldown (Fig. 1), a Python package that (i) provides a convenient command-line interface to run four typical scenarios in PPI screens and modelling of large complexes, (ii) reduces the computing time by separating CPU- and GPU-based calculations, (iii) allows selecting protein regions for modelling while retaining the original residue indexes and (iv) provides a unique analysis pipeline that assesses the predicted interfaces with multiple scores and generates a Jupyter notebook for interactive analysis.

2 Software description

2.1 Separation of stages

The original AlphaFold pipeline is composed of two main stages. The first uses CPUs to calculate multiple sequence alignments (MSAs) and to search for templates from PDB as input features for the second stage, which performs the actual structure prediction using GPUs. To speed up the pipeline, similar to other packages, AlphaPulldown separates these two stages, enabling the calculation of features for many sequences in parallel using multiple CPUs and prediction of structures using GPU.

2.2 Calculation of features

At this stage, AlphaPulldown searches each protein sequence against sequence databases to pre-calculate MSA and identify template structures, using either the original AlphaFold pipeline or more computationally efficient MMSeqs2 (Steinegger and Söding, 2017) from ColabFold (Mirdita *et al.*, 2022). It will also create features informing sequence by-organism pairing, which in the original AlphaFold-Multimer is performed only at the prediction state. AlphaPulldown stores these features in Python pickle files that can be reused for multiple PPI predictions.

2.3 Prediction modes

Once MSAs and structural template features are calculated, the user can choose from the following four modes of the main prediction.

2.3.1 Pull-down mode

Inspired by pulldown assays, this mode takes one or more proteins as 'baits' and a list of other proteins as 'candidates'. AlphaPulldown will automatically run AlphaFold-Multimer prediction between each bait and candidate. To demonstrate this mode, we show two examples. First, we screened for interactions of human eIF4G2 with

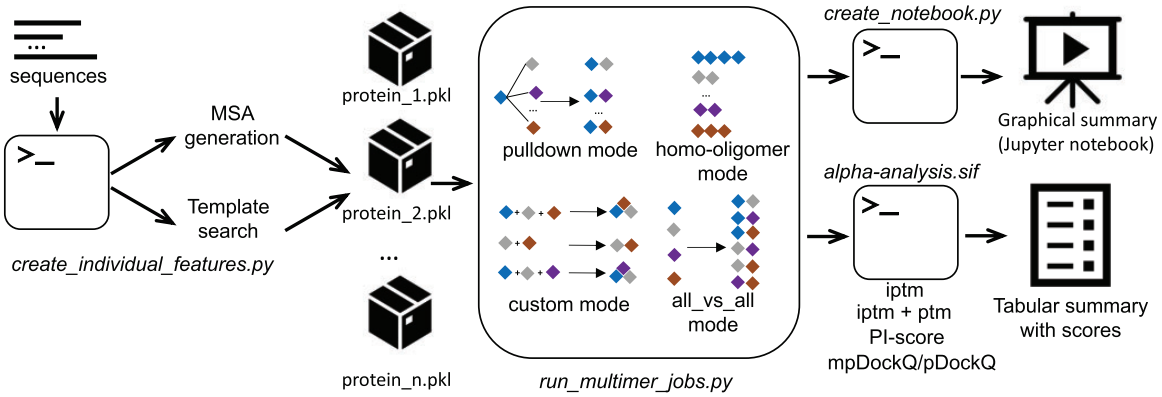

Fig. 1. The workflow of AlphaPulldown. The first stage of AlphaPulldown runs on CPUs and is handled by the create_individual_features.py script, which takes protein sequences as input, generates MSAs, and finds structural templates for each protein. MSAs and structural features are stored in Python 'pickle' files. The next stage takes these 'pickle' files as input and predicts protein complex structures on GPUs. The user can choose from the following modes: (1) pulldown, (2) homo-oligomer, (3) custom and (4) all versus all. Once all predictions are finished, AlphaPulldown generates a Jupyter notebook that provides a graphical summary and uses an analysis pipeline built in a Singularity image, alpha-analysis.sif, to produce a table with various model quality scores and interface properties

proteins of the human translation pathway (Supplementary Note S1). The results confirmed a known interaction, revealed new interactions, and, based on the additional interface quality scores provided by AlphaPulldown, identified potential false positive and negative AlphaFold predictions (Supplementary Table S1 and Fig. S1). Second, we modelled the interaction between the L and Z proteins of the Lassa virus, which could not be predicted using full-length sequences but only when screening Z against a series of fragments of L (Supplementary Note S2, Table S2 and Fig. S2). This example shows that fragmenting large proteins may help AlphaFold find correct interaction interfaces. AlphaPulldown provides a convenient interface to specify any combination of residue ranges without needing to recalculate MSAs or template features. Moreover, AlphaPulldown keeps the residue numbering of original full-length sequences in the models.

2.3.2 All-versus-all mode

Based on a single list of proteins, AlphaPulldown will automatically generate all pairwise combinations and predict their structures. This mode can be useful to predict interaction networks and provide input for modelling large complexes with tools such as MoLPC (Bryant *et al.*, 2022b).

2.3.3 Homo-oligomer mode

This mode simplifies the process of modelling homo-oligomers and testing alternative homo-oligomeric states. The user only needs to provide an input file with desired oligomeric states, and AlphaPulldown will automatically run modelling for each state.

2.3.4 Custom mode

This mode allows the user to provide any combination of any number of proteins and their fragments as an input, not limited to pairwise predictions. This mode can be used, for example, to screen a pre-defined list of interactions from other sources such as crosslinking studies or model entire complexes. As mentioned above, even if fragments are used, the MSA and template features do not need to be recalculated and the resulting models keep the residue numbering of the full-length sequences.

3 Additional features

For all models with an inter-chain predicted alignment error below the user-defined cut-off, a Jupyter notebook will be generated to provide the user with a clear visual overview of these models as well as their ipTM and pLDDT scores (Supplementary Fig. S3). The analysis pipeline will also return a CSV table containing the ipTM and ipTM+pTM scores reported by AlphaFold, a pDockQ score for dimers (Bryant *et al.*, 2022a), mpDockQ score for multimers (Bryant *et al.*, 2022b), protein interface-score (PI-score) (Malhotra *et al.*, 2021), and physical and geometrical properties calculated by the PI-score pipeline, such as the interface surface area or the number of hydrogen bonds (Supplementary Tables S1 and S2).

4 Conclusion

AlphaPulldown streamlines AlphaFold-multimer for PPI screens and modelling of large complexes fragment-by-fragment in a manner similar to that we applied to the human nuclear pore complex (Mosalaganti *et al.*, 2022). It both facilitates the prediction process and provides a toolbox for a more thorough analysis of model confidence. We anticipate it will help the structural biology community unlock the full potential of artificial intelligence-based structure prediction. Possible directions for further development of AlphaPulldown may include support for modelling algorithms other than AlphaFold, acceleration of the modelling step and new PPI scoring functions.

Acknowledgements

We thank Agnieszka Obarska-Kosinska for testing the software and useful suggestions.

Funding

This work was supported by the German Research Foundation (DFG) [KO 5979/2-1].

Conflict of Interest: none declared.

Data availability

All the code and example data are available at https://www.embl-hamburg.de/AlphaPulldown and https://github.com/KosinskiLab/AlphaPulldown.

References

Bryant,P. *et al.* (2022a) Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.*, **13**, 13, 1–11.

Bryant,P. *et al.* (2022b) Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search. *Nat. Commun.*, **13**, 6028.

Evans,R. *et al.* (2022) Protein complex prediction with AlphaFold-Multimer. https://github.com/deepmind/alphafold.

Humphreys,I. *et al.* (2021) Computed structures of core eukaryotic protein complexes. *Science (1979)*, **374**. https://www.science.org/doi/10.1126/science.abm9506.

Jumper,J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.

Malhotra,S. *et al.* (2021) Assessment of protein–protein interfaces in cryo-EM derived assemblies. *Nat. Commun.*, **12**, 1–12.

Mirdita,M. *et al.* (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.

Mosalaganti,S. *et al.* (2022) AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science (1979)*, **376**. https://www.science.org/doi/10.1126/science.abm4805.

Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.