




Af2-analysis: a Python package for alphafold 2 analysis

Alaa Reguei¹ and Samuel Murail¹ 

¹ Université Paris Cité, Inserm U1133, CNRS UMR 8251, Paris, France

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright
and release the work under a
Creative Commons Attribution 4.0
International License ([CC BY 4.0](#)).

Summary

The publication of AlphaFold 2(Jumper et al., 2021) has significantly advanced the field of protein structure prediction. Predicting protein structures has long been a central challenge in structural bioinformatics, with the ultimate aim of elucidating the relationship between protein structure and function. Accurate protein structure prediction is crucial for various applications, including drug discovery, protein engineering, and the investigation of protein-protein interactions. AlphaFold 2, a deep learning-based approach, has demonstrated unprecedented accuracy in protein structure prediction, surpassing other contemporary methods. The availability of AlphaFold 2 has opened new avenues for protein structure analysis, allowing researchers to investigate the structure-function relationship of proteins with greater precision. Here, we present af2-analysis, a Python package that provides tools for the analysis of AlphaFold 2 results. af2-analysis is designed to facilitate the analysis of protein structures predicted by AlphaFold 2, providing functions for the comparison of predicted structures with experimental structures, the visualization of predicted structures, and the calculation of structural quality metrics.

Statement of need

Since the publication of AlphaFold 2(Jumper et al., 2021) in 2021, the scientific community can access a prediction accuracy of protein structures that was previously unattainable. The derivatives Colabfold(Mirdita et al., 2022) and AlphaFold Multimer(Evans et al., 2022) have been developed to predict the structure of protein complexes, once again defining a new standard for protein-protein and protein-peptide docking. However as shown by Björn Wallner, it is sometime necessary to generate thousands of model to get few good models(Wallner, 2023). The analysis of the results can then be a fastidious, moreover if alphafold quality metrics are good, supplementary metrics have been developed to evaluate the quality of the models as pdockq(Bryant et al., 2022), pdockq2(Zhu et al., 2023), and LIS score(Kim et al., 2024). All this metrics have to be calculated from different scripts. Another point to access is the diversity of the models, and as shown in AFsample(Wallner, 2023) it is sometime necessary to compute up to ten thousand models, and then to cluster them to select the best models. Af2-analysis is designed to facilitate the analysis of set of model structures and metrics. The library is build upon pandas library and will import an alphafold or colabfold prediction directory as a pandas dataframe. The library will provide functions to add additional metric to the dataframe, to compare the models with experimental structures, to visualize the models, to cluster the models, and to select the best models.

References

- Bryant, P., Pozzati, G., & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, 13(1), 1265.
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., ... Hassabis, D. (2022). Protein complex prediction with AlphaFold-multimer. *bioRxiv*. <https://doi.org/10.1101/2021.10.04.463034>
- 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, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kim, A.-R., Hu, Y., Comjean, A., Rodiger, J., Mohr, S. E., & Perrimon, N. (2024). Enhanced protein-protein interaction discovery via AlphaFold-multimer. *bioRxiv*. <https://doi.org/10.1101/2024.02.19.580970>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), 679–682.
- Wallner, B. (2023). AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics*, 39(9), btad573. <https://doi.org/10.1093/bioinformatics/btad573>
- Zhu, W., Shenoy, A., Kundrotas, P., & Elofsson, A. (2023). Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics*, 39(7), btad424. <https://doi.org/10.1093/bioinformatics/btad424>