

# Af2-analysis: a Python package for AlphaFold 2 analysis

Alaa Reguei<sup>1</sup> and Samuel Murail<sup>1</sup>

<sup>1</sup> 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Published: unpublished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. The Prediction of protein structures has long constituted a central challenge in the field of structural bioinformatics, with the ultimate objective of elucidating the relationship between protein structure and function. The accurate prediction of protein structures is of great importance for a number of applications, including drug discovery, protein engineering, and the investigation of protein-protein interactions. AlphaFold 2, which employs a deep learning-based approach, has demonstrated unprecedented accuracy in protein structure prediction, outperforming other contemporary methods. In this paper, we present af2-analysis, a Python package that provides tools for the analysis of AlphaFold 2 results. af2-analysis has been designed to facilitate the analysis of protein structures predicted by AlphaFold 2. It provides functions for the comparison of predicted structures with experimental structures, the visualisation of predicted structures, and the calculation of structural quality metrics.

## Statement of need

The publication of AlphaFold 2(Jumper et al., 2021) in 2021, has enabled the scientific community to achieve a previously unattainable level of accuracy in predicting protein structures. Derivatives of AlphaFold 2, namely Colabfold(Mirdita et al., 2022) and AlphaFold Multimer(Evans et al., 2022) have been developed to predict the structure of protein complexes, thereby establishing a new standard for protein-protein and protein-peptide docking. However as demonstrated by Björn Wallner, the generation of thousands of models is sometimes necessary to obtain a few high-quality models(Wallner, 2023). Subsequent analysis of the results may prove to be a laborious and meticulous process. Furthermore if quality metrics produced by AlphaFold are good, additional metrics have been developed to assess the quality of the models. This include pdockq(Bryant et al., 2022), pdockq2(Zhu et al., 2023), and LIS score(Kim et al., 2024). All of the aforementioned metrics must be calculated from different scripts. A further point to consider is the diversity of the models. As illustrated in AFsample(Wallner, 2023), there are occasions when it is necessary to compute up to ten thousand models, and then cluster them in order to select the most appropriate models. The Af2-analysis has been developed with the objective of facilitating the analysis of sets of model structures and associated metrics. The library has been constructed on the pandas library and is capable of importing an alphafold or colabfold prediction directory as a pandas DataFrame. The library provides a range of functions that enable addition of further metrics to the DataFrame, the comparison of models with experimental structures, the visualisation of the models, the clustering of models, and the selection of the best models.

## References

- Bryant, P., Pozzati, G., & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, 13(1), 1265. <https://doi.org/10.1038/s41467-022-28865-w>
- 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. <https://doi.org/10.1038/s41592-022-01488-1>
- 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>