



# Identifying Binding Site Templates for Ligand Docking to AlphaFold Models

John Wang<sup>1</sup>, Masha Karelina<sup>1</sup>, Ron Dror<sup>1</sup>

<sup>1</sup>Dror Lab, Department of Computer Science, Stanford University

## Background and Introduction

- AlphaFold is an AI model capable of predicting 3D protein structure from a sequence of amino acids with near-experimental accuracy [1]. Easily accessible and accurate protein structure predictions are extremely useful for drug discovery. However, AlphaFold structures struggle in ligand docking performance benchmarks.

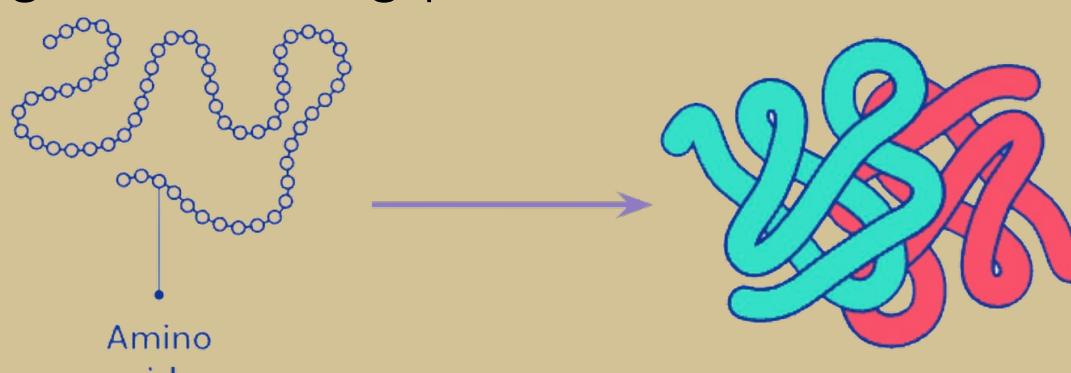


Figure 1: Given a sequence of amino acids, AlphaFold predicts the 3D structure of the sequence.

Our goal is to improve AlphaFold structures for ligand docking.

- Ligands are molecules such as drugs that attach to a protein's binding site.
- Binding site refinement improves the structural accuracy of a protein's binding site.
- One approach is to find binding sites templates from proteins with similar sequences or from similar families.
  - Guterres *et al.* successfully used a local structure alignment tool to improve docking performance [3].

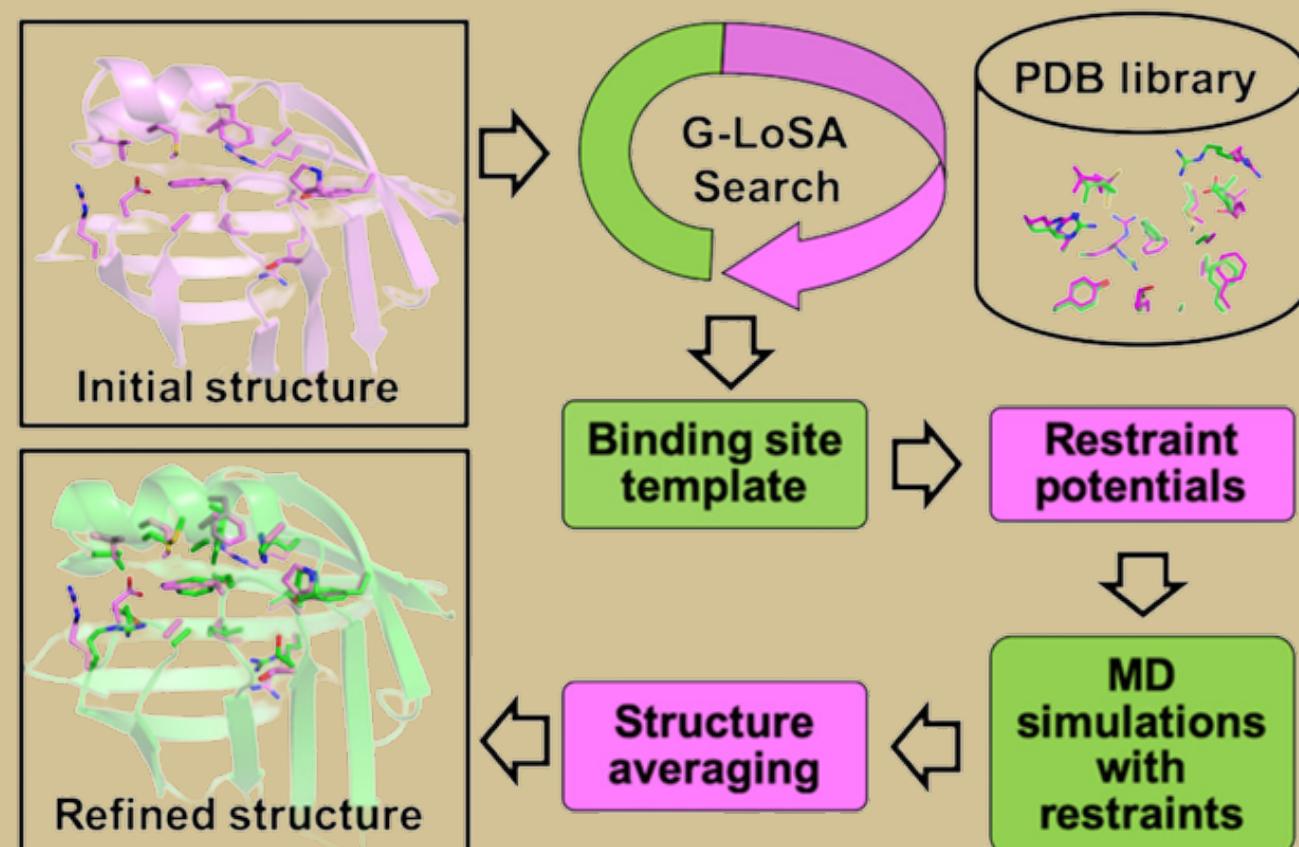


Figure 2: G-LoSA is a local structure alignment tool used to find binding site templates. Templates were used to run molecular dynamics (MD) simulations and improve structures for docking.

- Our project aims to test this method by characterizing the binding site templates we find for our AlphaFold structures (20 G protein-coupled receptors or GPCRs) and exploring whether finding these templates would be unique to the local structure alignment method.

## Workflow

- Why local structure alignment?
  - Proteins with dissimilar overall (global) structure can have similar binding site (local) structure [2].

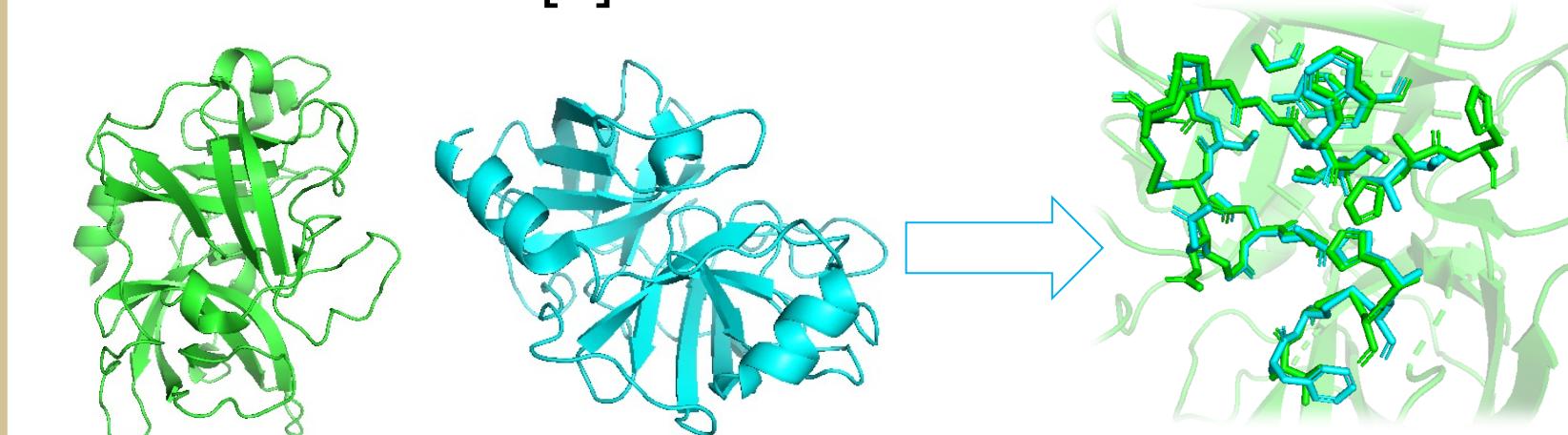


Figure 3: Alignment of trypsin and urokinase, two proteins with dissimilar function but highly similar binding sites.

Our workflow aligns structures locally using G-LoSA, then uses several scripts to filter results out.

- 1) A center-of-mass (COM) script filters out results where the template ligand does not fit in the AlphaFold binding site.
- 2) A sequence identity script aligns the AlphaFold structure to the whole protein of each binding site hit using blastp and returns the identity and alignment length.

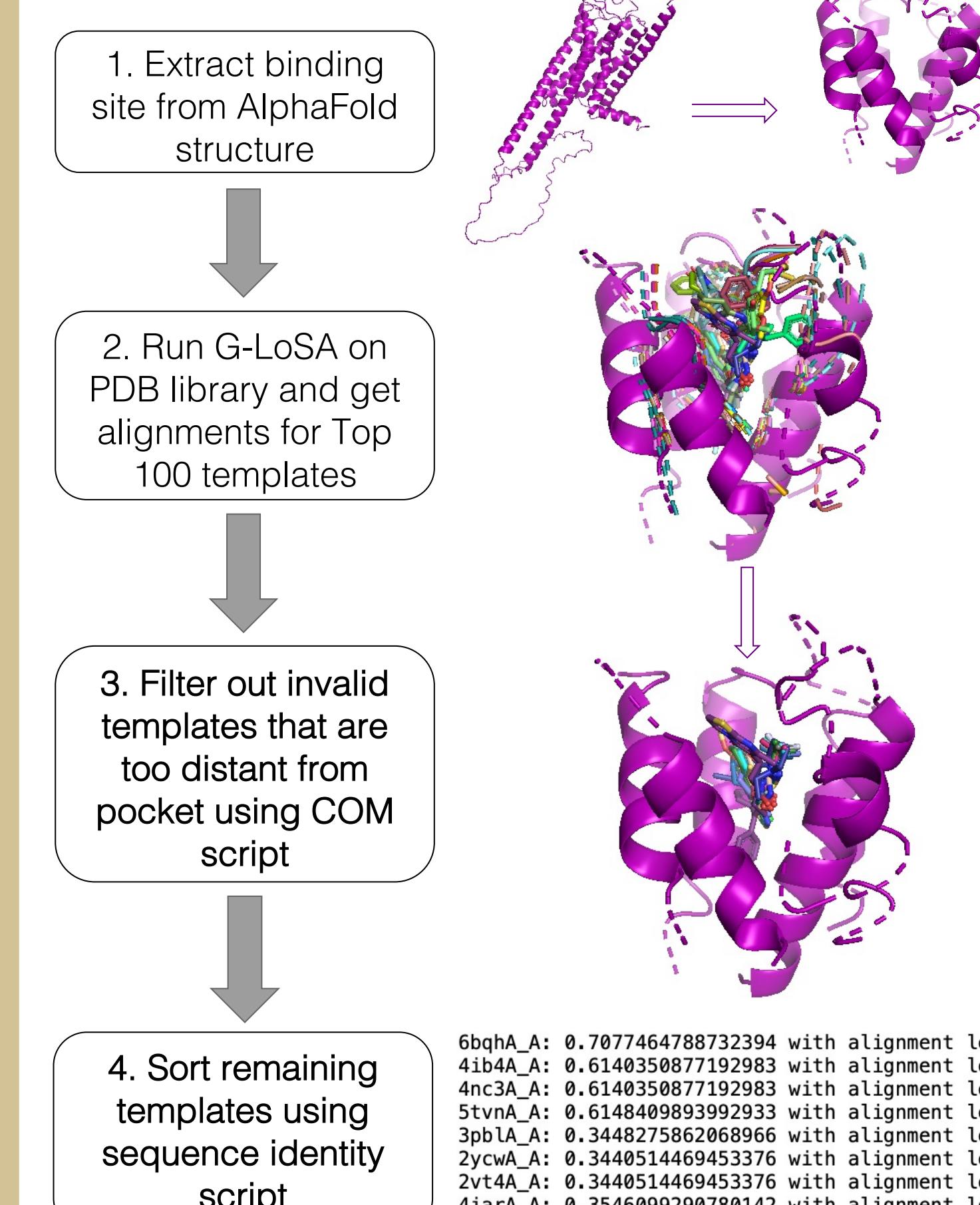
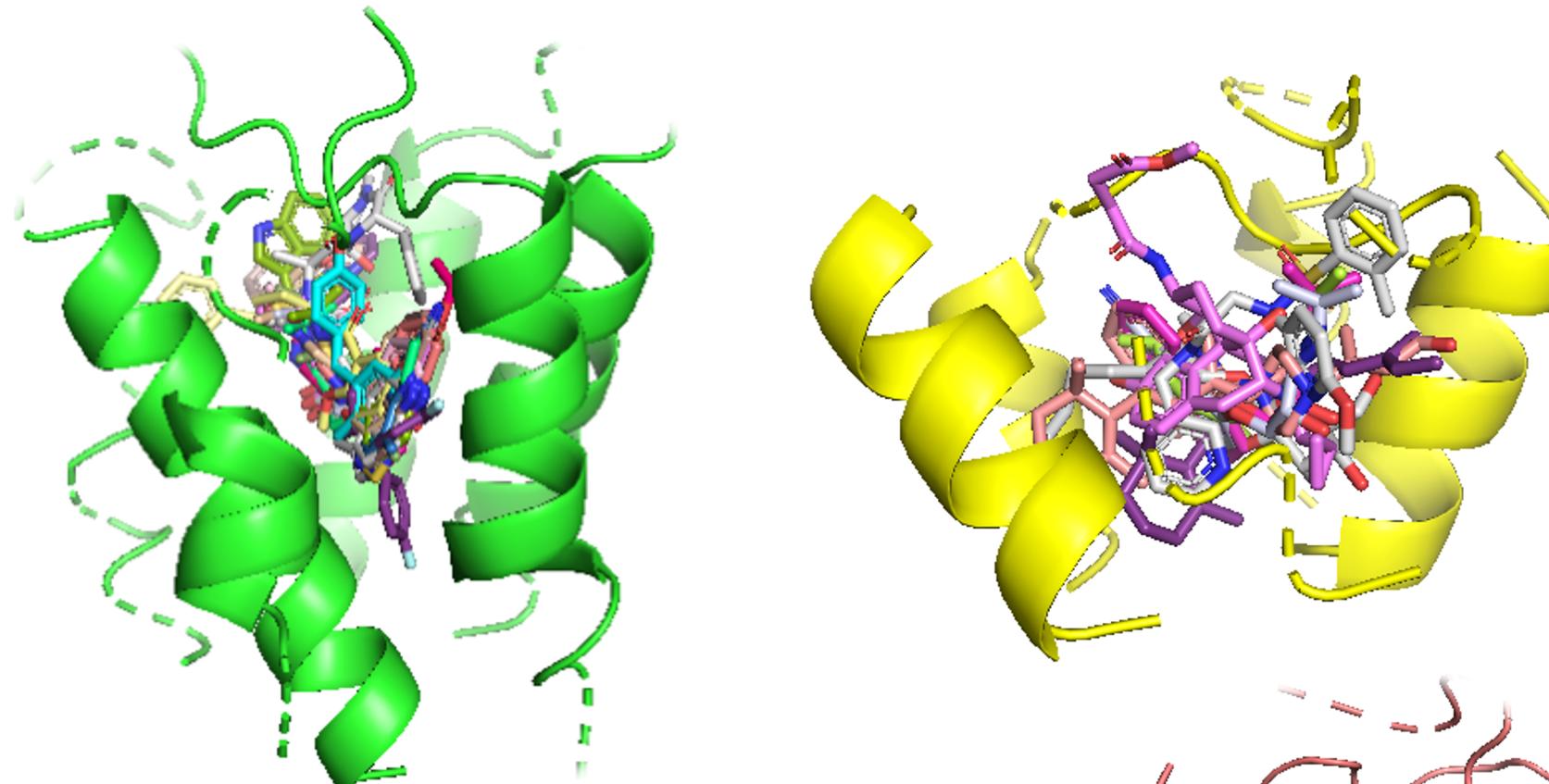


Figure 4: Our workflow for finding and processing binding site templates.

## Results



- Visualized AlphaFold structures with many (green), some (yellow), and almost no (red) GPCR hits.
- In general, more GPCR hits indicates more well-aligned binding site templates

Figure 5: AlphaFold binding sites shown with their filtered templates (by COM). Green – Dopamine receptor D1, Yellow – Melatonin receptor 1A, Red – Prostaglandin E2 receptor

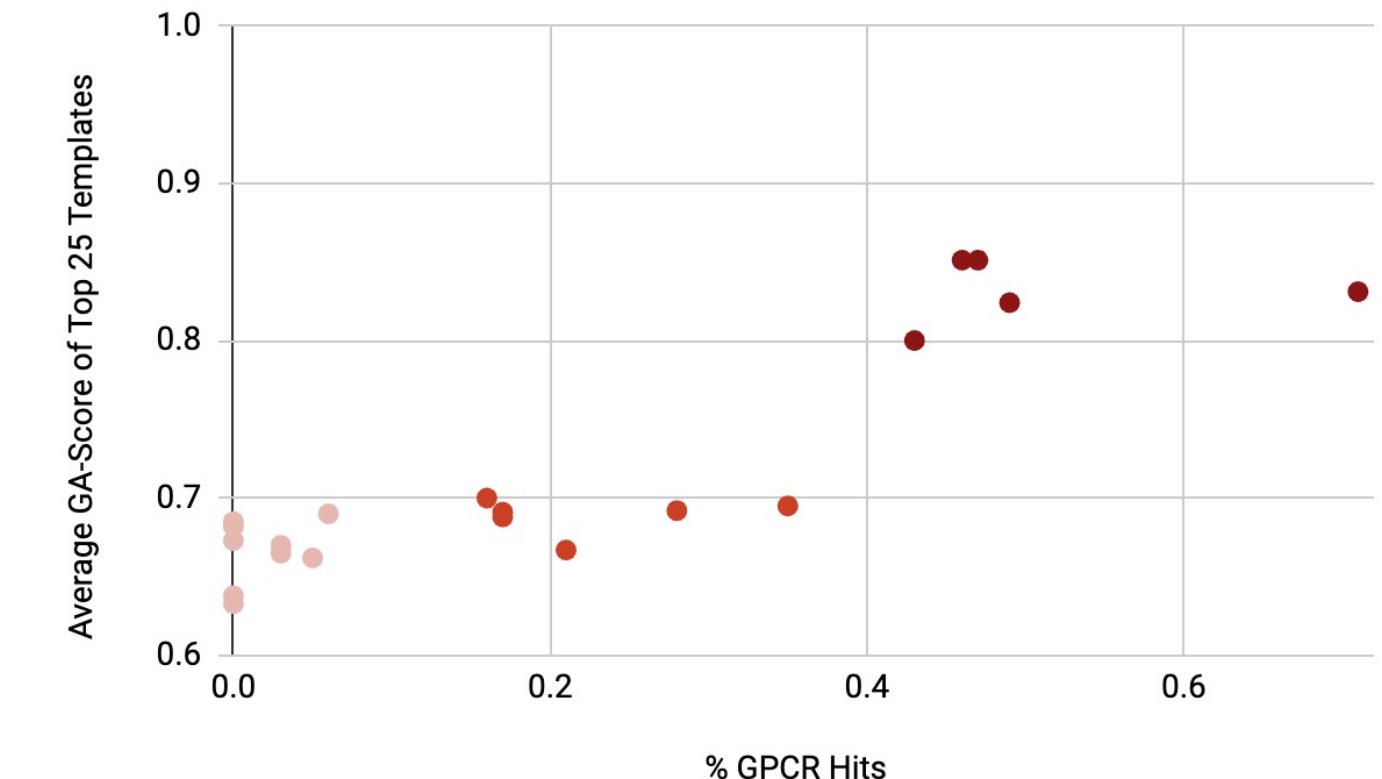
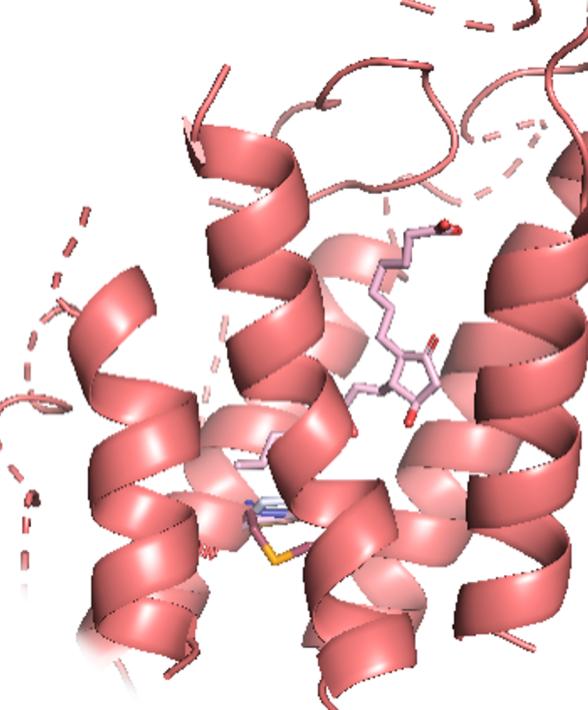


Figure 6: AlphaFold structures plotted with their GA-score vs percentage of hits that are GPCRs. Structures are roughly clustered into 3 groups by % GPCR hits.

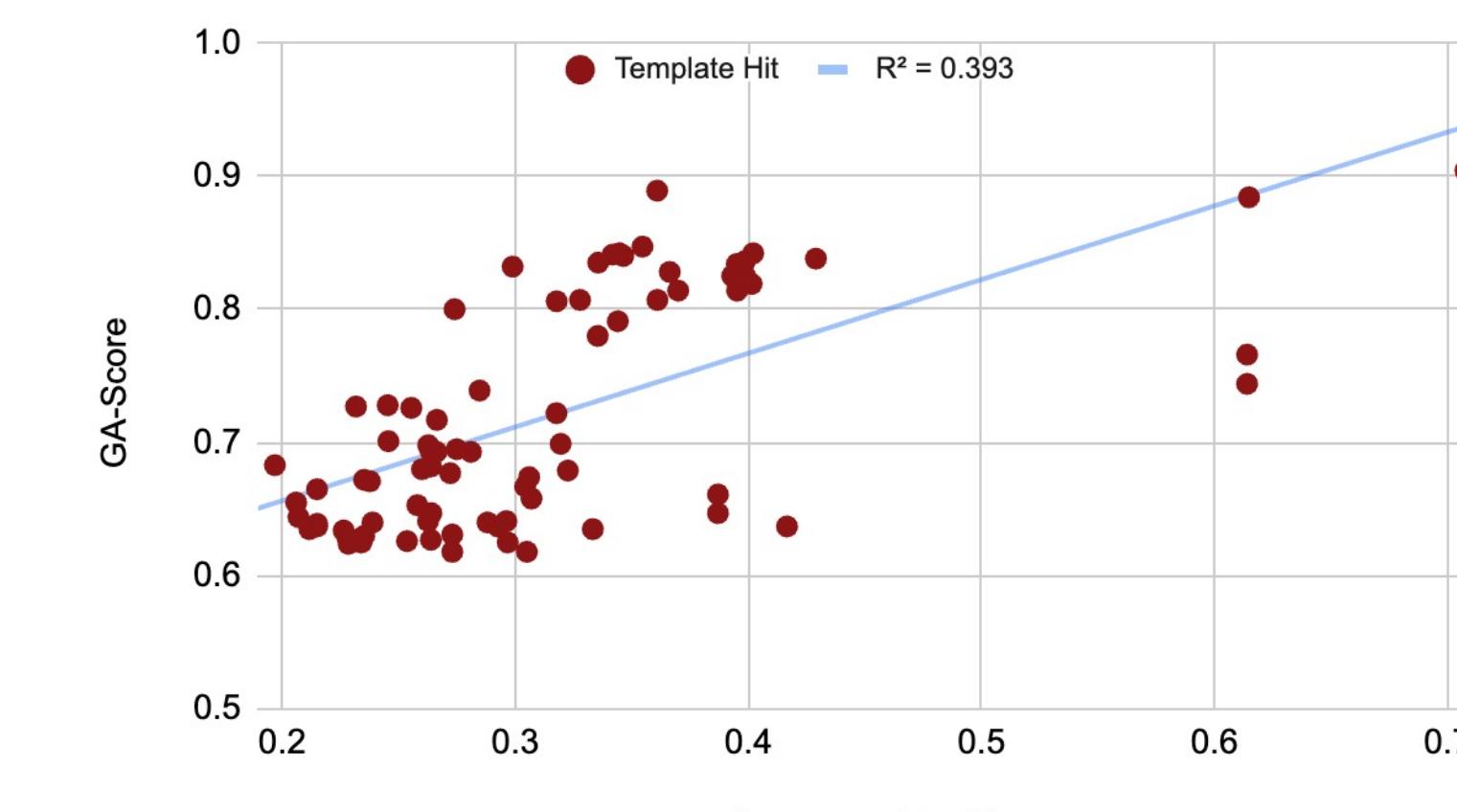
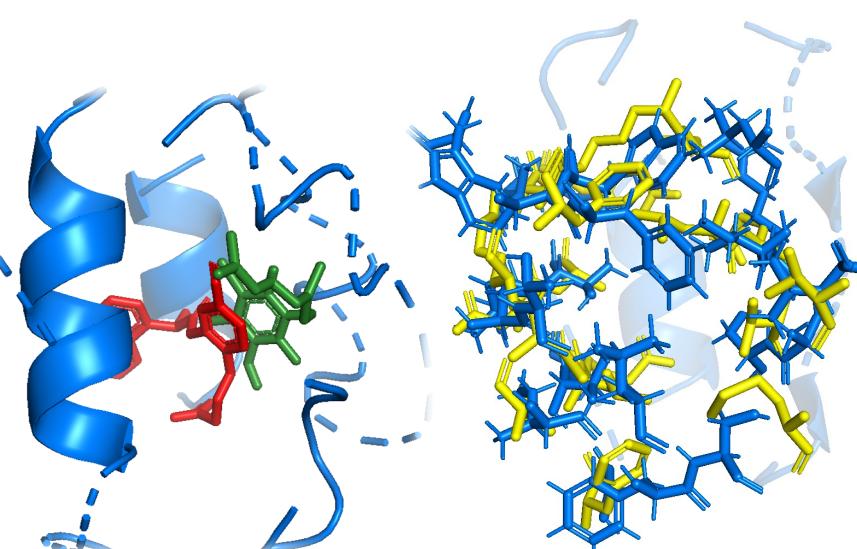


Figure 7: Individual binding site hits plotted by their GA-score and protein sequence identity, randomly sampled from the set of all template hits. A very weak correlation is seen between the two metrics.

## Summary

- AlphaFold structures that have higher average scores typically have more GPCR hits and more templates that fit the AlphaFold binding pocket.
- This follows the reasoning that similar proteins generally also have similar functional sites such as binding pockets.
- It is noteworthy that some AlphaFold structures with no GPCR hits still return templates that have ligands fitting the AlphaFold pocket. This is the case for the human prostaglandin E receptor shown:

Figure 8: The binding site of an oxidase found in a kinetoplastid (yellow) aligned with the binding site of the human prostaglandin E receptor (blue). Two dissimilar proteins with similar binding pockets according to G-LoSA. Alignment of ligands (kinetoplastid, green and GPCR, red) is also shown.



- Hits that are not GPCRs but still seem to be good templates are of interest because they fit the idea that proteins with dissimilar functions and families can have similar local structure (binding sites).

## Future Work

- Streamline current workflow and look at other methods for (1) filtering good templates and (2) algorithms for finding templates.
- Further analysis on how to best rank templates. Metrics include sequence identity, chemical similarity of ligand, GA-score, etc.
- Exploration on how to use templates to modify AlphaFold models based on the template include:
  - MD simulations with restraints, such as in [3]
  - Minimization around top template ligands
- Ligand Docking benchmarks should be used to determine the final usefulness of templates / modification methods

## References

- [1] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
  - [2] Lee HS, Im W. G-LoSA: An efficient computational tool for local structure-centric biological studies and drug design. *Protein Sci.* 2016 Apr;25(4):865–76. doi: 10.1002/pro.2890. Epub 2016 Mar 6. PMID: 26813336; PMCID: PMC4941214.
  - [3] Hugo Guterres, Hui Sun Lee, and Wonpil Im. Ligand-Binding-Site Structure Refinement Using Molecular Dynamics with Restraints Derived from Predicted Binding Site Templates. *Journal of Chemical Theory and Computation* 2019 15 (11), 6524-6535. DOI: 10.1021/acs.jctc.9b00751
- Acknowledgements: Thanks to Masha Karelina for her mentorship, Prof. Dror and other members of the Dror Lab, and CURIS for funding and support.