**Capstone Project: Progress Report**

**Background**

Our project is inspired by a recent publication (Collins et. al), which benchmarked AlphaFold protein structures as “needing improvement before they can help identify drug mechanisms of action”. We hope to push this field by asking the question of whether or not a flexible structure will have more success than a static one at predicting these protein-protein interactions. This has implications in many fields including drug manufacturing, vaccinations, and countless other biological technologies. We are using the trRosetta algorithm(Yang) in order to create our flexible structures, since they output 5 structures which will model more flexible movements of the protein. We will use AutoDock Vina as our docking software to test protein-protein interactions, inspired by the MIT Study (Collins).

**Progress**

We have made significant progress on our project since the last progress report. In terms of the workflow, trRosetta is fully functional on Cypress. This means we are now in a position to begin toying with our outputs, changing the settings on how many PDB files we would like to see. We downloaded two docking softwares onto Cypress. These are AutoDock Vina and dock6. Dock6 is functional, yet we decided to switch to working with AutoDock Vina out of ease and similarity to the MIT study. This requires a bit more additional work until fully functional, but it is almost ready for use. We have also bookmarked the proteins which we would like to use, making a list of 12 experimental controls we will start with, following with the list of 296 essential E. Coli proteins that will be compared to 218 known antibacterial compounds and 100 known inactive compounds. These are the same proteins from the MIT study, and they provided us with a table of binding affinity scores between all proteins. We are currently in a place where we know how the programs run, which programs we would like to run, and all that is left is to use SLURM (the batch method on cypress) to run these proteins.

**Open Questions**

We still have several open questions standing. First, AutoDock Vina has the capability to do “basic docking” or “flexible docking”. We will explore both of these options with hope that the flexible docking option will be different and interesting compared to the basic docking mechanism.

Which aggregation function will we use to concatenate the binding affinities of all 5 “versions” of the proteins we choose? Should we use mean, median, mode, max? This is a question we have time until we absolutely need to decide, but it is something we would like to start thinking about already.

**Timeline**

We are very enthusiastic about our progress and how much we have learned this semester. We both plan on working on the project over Winter Break, hopefully returning and starting the semester on a strong note by being able to run our workflow from start to finish. By the end of January, we hope to perfect the workflow and begin running our extended list of proteins using the SLURM on Cypress. We don’t know exactly how long this will take, but we will devote February and March to analysis of our results. This means deciding on our aggregation function, possibly converting the data to a pandas dataframe so we may play around with it and make some nice graphs to visualize statistical significance. Finally, in April we will conclude our analysis and be ready to share our project.

**Workload Distribution**

We have been working very well together, since we each have different strengths. Yali took the front role in choosing proteins and examining the study. Rafi dove head first into trRosetta, dealing with many technical issues along the way. We work together and meet up several times each week, and this partnership has been very fair.

**Faculty Mentor Meeting Plan**

We have been meeting with Maus weekly, which has been extremely helpful in navigating the technical issues of Cypress that we are unaware of. We are very grateful to have Maus as our faculty advisor, and we are extremely excited for another semester of working with him.

**Citations**

Wong, F., Krishnan, A., Zheng, E. J., Stärk, H., Manson, A., Earl, A., Jaakkola, T., & Collins, J. (2022, September 1). *Benchmarking alpha fold‐enabled molecular docking predictions for ...* Molecular Systems Biology. Retrieved September 13, 2022, from https://www.embopress.org/doi/full/10.15252/msb.202211081

Yang, Jianyi. “TrRosetta.” *Protein Structure Prediction by TrRosetta*, https://yanglab.nankai.edu.cn/trRosetta/.

“AutoDock Vina: Molecular Docking Program.” *AutoDock Vina: Molecular Docking Program - Autodock Vina 1.2.0 Documentation*, https://autodock-vina.readthedocs.io/.

Database, AlphaFold Protein Structure. “Alphafold Protein Structure Database.” *AlphaFold Protein Structure Database*, https://alphafold.ebi.ac.uk/.

“What Are Proteins and What Do They Do?: Medlineplus Genetics.” *MedlinePlus*, U.S. National Library of Medicine, https://medlineplus.gov/genetics/understanding/howgeneswork/protein/.