Technical Test - Protein Design

Problem Setting

Let there be a protein Prot_A which binds to two other proteins, Prot_B and Prot_C, such that the binding interfaces of Prot_A with Prot_B and Prot_A with Prot_C overlap. Prot_B and Prot_C could themselves be multimers (e.g. Antibodies) but Prot_A is a monomer. Assume the structures of these complexes have been resolved.

For this technical test, we would like you to propose a plan for an automated, multi-step analytical pipeline of the interface residues of Prot_A for any set of {Prot_A,Prot_B,Prot_C}. The purpose of this pipeline would be to provide insights regarding the mechanism that govern protein-protein interfaces. Using these insights, or not, perform one interface design experiment where Prot_A would abrogate binding with Prot_B while maintaining or increasing the binding affinity for Prot_C.

Then, we would like you to demonstrate individual steps of your choice by implementing them for the specific use case:

- Prot_A = SARS-CoV-2 spike protein
- Prot_B = Antibody P5A-3C8
- Prot_C = ACE2

We suggest using the following PDB entries: 6M0J and 7Z0X, but feel free to pick others that match the specific use case presented above. If you pick other PDBs please let us know why.

Deliverable

Your deliverable should be divided into the following parts:

1. Automated Analytics Pipeline
    a. A slide deck presenting your analytical pipeline.You should be able to explain the proposed methods in detail as well as their limitations. Comparing different methods, from different areas of research will be a bonus.

2. SARS-CoV-2 Proof of Concept
    a. The implementation of the chosen steps;
    b. A report of the resulting observations in the form of either a PDF document or a Jupyter notebook.

Please note that:

- If you choose a pdf document, you must also provide the code used for the project.
- If you provide a Jupyter notebook, please re-run it from scratch and save the notebook

with the outputs of the cells.

- Given the time constraints, your code does not have to be generic, please focus on the proposed use case and choose the steps you are comfortable with.

Evaluation

We will pay attention to the code quality and the documentation. You will also be evaluated on your capacity to communicate the results of your work both verbally and writing to a technical audience.

Compute

In case you need more computing power than locally available on your computer, please consider the following free resources:

- Google Colab: access to one GPU or one TPU, time limit of 12 hours (kernels are shut down after 12 hours)
- Kaggle notebooks: access to one GPU (NVIDIA P100), time limit of 6 hours

Hope you have fun!
Please feel free to contact us if you have any questions, we'll be happy to help