Challenge 5: Predict protein complex structures with AlphaFold

Background: Peptide-MHC-TCR complex prediction is a challenging problem even in the era of AlphaFold. In this challenge you will implement structure prediction protocols and compare the quality of the reconstructed models.

Goal: Here we're interested in comparing different AlphaFold models on tertiary complex prediction tasks, specifically peptide-MHC-TCR complexes. We demoed ColabFold in class and now you will use it to run AlphaFold. Download the file 5tje_trim.fasta and 5tje_trim.pdb. The ".pdb" file is the resolved crystal structure of a mouse TCR in complex with a viral peptide presented by a mouse MHC. For brevity, the structure has been pruned to keep the parts we are interested in.

Tools:
https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
Pymol and pymol wiki: https://pymolwiki.org/index.php/Main_Page
Or any code and package you find online such as Biopython, OpenFold suite, etc.

Approach: Predict the structure of the tertiary complex composed of peptide, mhc, tcr_A, and tcr_B with ColabFold. We will predict models with and without templates using monomer and multimer_v3 (total 4 models for each combination). Set up ColabFold accordingly and at each step download your results. Once you have made all the predictions you need, look at your predicted metrics and guess which structure is the best predicted model. Justify your reasoning.
Now open pymol and import your predictions for each condition (there will be multiple predictions for each condition, feel free to use all the models or the top ranked one). We are interested in assessing the quality of the dock and the CDR loop modeling of the TCR. The sequence of CDR3a is CAALYGNEKITF and CDR3b is CASSDAGGRNTLYF. Superimpose the predicted structures to the native structure (use the most reasonable chains to base the superimposition on). Now compute the RMSD of the prediction to the native for the peptide, CDR3a, and CDR3b. Report your metrics and rank the models based on these metrics. Were your assumptions correct? Plot the RMSD values against predicted AlphaFold metrics. Explain what you see. Save your pymol sessions at the end and submit it with all the commands you executed in pymol along with your write-up, your plots and colabfold plots.
Note: If you are more advanced and feel comfortable doing the superimposition and RMSD calculations in python, feel free to do so.
Pro tip: Pymol does not have a real "undo" button. So make sure to save your pymol sessions after each successful step!