How to run AlphaFold2-multimer (AFM):

For AFM computing, we will provide scripts and instructions to run and install AFM as we did in our article to facilitate result replication. However, replicating AFM results in a reasonable timeframe is infeasible due to the extensive setup and computation on HPC machines, a process that spanned several months.

To run AFM locally on HPC system, we installed LocalColabFold (version 1.3.0) an installer script to make ColabFold functionality available on users' local machines. Instructions for installation are provided here: https://github.com/YoshitakaMo/localcolabfold

In order to integrate the calculation of the interface score and piTM score, as defined by Gao et al (https://www.nature.com/articles/s41467-022-29394-2) and as reported in our article, three scripts underwent necessary modifications for inclusion in the AFM simulations: *model.py, confidence.py* and *batch.py*. You can find them in data/scripts.
Replace the file at "/path/to/colabfold_batch/colabfold-conda/lib/python3.7/site-packages/alphafold/model.model.py" with "model.py."
Replace the file at "/path/to/colabfold_batch/colabfold-conda/lib/python3.7/site-packages/alphafold/common/confidence.py" with "confidence.py."
Replace the file at "/path/to/colabfold_batch/colabfold-conda/lib/python3.7/site-packages/colabfold/batch.py" with "batch.py."

The main script to run AFM simulations is *run_colabfold_AF2Mv2_unpaired_paired.sh.* This can be used as a template for your own purposes. The advantage of this script is that it recursively submits new jobs after a certain time (set by *sleeptime* variable in line 31).
The script sets the following setting for AFM simulations:
- --num-recycle 3
- --templates
- --pair-mode unpaired+paired
- --model-type AlphaFold2-multimer-v2

To test another AFM setting as we did in our article, select the pair mode ('unpaired+paired', 'unpaired', or 'paired') and the model type ('AlphaFold2-multimer-v1' or 'AlphaFold2-multimer-v2').

The input sequence for AFM ($SEQUENCE_FILE$PAIR_TO_TEST) comes from the 'FASTA_files' directory that contains the fasta sequences of the protein pairs in the positive set (Positive_set_Yeast.txt) and negative set (Negative_set_Yeast.txt) of Yeast DIP network. To generate this directory, run *create_fasta_input_AF2*. Total execution time: ~**18 min.**