

How to generate/replicate Figure 4:

Figure 4. Interaction disorder index and computation time of CH and AFM.

1. Running the code

To run the code, execute the function *run_figure_4* with one of these options:

- Option 1: 1 to generate item with existing results. Usage: *run_figure_4(1)*.
- Option 2: 2 to recreate item from original data, involving all required computations. Usage: *run_figure_4(2)*.

Here is an overview of the execution of each option. The execution times reported below are measured executing the code in Windows 10 Pro with 256 GB RAM, and AMD Ryzen Threadripper PRO 3995WX 64-Cores CPU with 2.70 GHz. The software environment is MATLAB 2019a.

Option 1: Total execution time **~8 min**

Runs *plot_figure_4* to generate item with existing results located in data folder. In data folder, there is:

- *original_data*: contains the raw data of Yeast DIP downloaded from its source.
- *matrix*: contains the adjacency matrix of the Yeast DIP network and a cell array whose first column contains list of positive protein pairs (1% of the links) and second column list of negative protein pairs (1% of the links) to test for each simulation (10).
- *network_similarities*: contains CH similarity scores of Yeast DIP network.
- *FASTA_files*: contains the fasta sequences of protein pairs in the positive set and negative set.
- *INTS_outputs*: contains the AFM-IS (interface score) for the protein pairs in GSP and GSN sets.
- *PITMS_outputs*: contains the AFM-piTMS for the protein pairs in GSP and GSN sets.
- *Yeast_AFM_output*: contains the AFM results of AFM-IS and AFM-piTMS and the timing file for each AFM-modelled protein pair for the best AFM setting we have selected, meaning AlphaFold2-multimer-v2 unpaired+paired.

- Iupred3: contains individual text files, each containing IUPred3 scores for every amino acid within each protein of a respective pair predicted by AFM.
- script: contains *plot_figure_4*, the main function to plot Figure 4.

Option 2: Total execution time **~10 min**

All the results of the following scripts are stored in the directory *data_replicated*. Below are the different steps to implement the computation:

- In order to download the data, follow the instructions reported in the file *instructions to download the data* located inside the folder “data”.
- *create_Scere_DIP_net* : creates the adjacency matrix for Yeast DIP network. Output in “matrix”. Total execution time: **~13 s**.
- *create_list_pairs_network_perturbed_1percent_Yeast*: conducts 10 simulations, each time generating a list of positive and negative protein interactions. These interactions are randomly selected from the positive and negative sets, respectively, and their quantity is set to 1% of the total links in the network. The script then removes the positively interacting protein pairs from the network. As output, the script returns 10 lists and 10 perturbed networks. This step is necessary to calculate later the average AFM computation time for one link in a small set of 1% of the links predicted by AFM. Outputs in “matrix”. Total execution time: **~2 min**
- *run_CHA_Yeast_DIP_net*: computes the CH similarity scores of the Yeast DIP network. Outputs in “network_similarities/CH_L2_L3/results”. Total execution time: **~1 min 30**
- *protein_pairs_ints_data_processing* and *protein_pairs_pitms_data_processing*: for the selected AFM setting, they calculate the average interface score (INTS) and piTM score values respectively for protein pairs in positive and negative sets, and saves the results to text files. Outputs in “INTS_outputs” and “PITMS_outputs”. Total execution time: **~1 min**.
- *iupred3_AF2_Yeast*: generates individual text files, each containing IUPred3 scores for every amino acid within each protein of a respective pair predicted by AFM. Outputs in

“results_AF2_Yeast_DIP_Pos” and “results_AF2_Yeast_DIP_Neg”. Total execution time: **~4 min.**

- *plot_figure_4*: Plot Figure 4. Total execution time: **~1 min.**

2. Required modules for Python codes

The Python version used is 3.9.13 and the list of required modules is:

- pathlib
- itertools
- numpy
- glob
- csv
- re
- sys
- json
- pickle
- itertools
- Bio
- Iupred3_lib