EnsembleLab tutorial

1. Open the following link (using Google Chrome is highly recommended) https://colab.research.google.com/github/fpesceKU/EnsembleLab/blob/main/EnsembleLab.ipynb

2. Make sure to have **GPUs** enabled: go to "Runtime", select "Change runtime type", and select "GPU".

3. In "0. IDP sequence and data", enter a **NAME** for your protein and paste your amino acid **SEQUENCE** in one-letter code. If you want to analyse SAXS data, then select "**EXPERIMENT**": SAXS. A prompt will appear that allows to upload a file with SAXS data to session storage.

4. Following in "0. IDP sequence and data", you can specify environmental conditions for the simulation. If aiming at comparing simulations to experiments, we recommend **adjusting the settings to reproduce the conditions of the experiments** (temperature, pH, salt concentration).

5. Run the three "Preliminary operations" cells **one by one**, waiting for the execution of each cell to be complete (a green check mark will appear) before running the next. When the execution of "Preliminary operations: setting the environment (i)" is complete, the session will restart and Colab will report a message related to the session crashing. That is normal and required for all packages the notebook needs to work properly.

6. After all "Preliminary operations" are executed and complete, you can execute the following cells altogether (executing cell-by-cell or selecting "Run after" from the "Runtime" menu).

    6a. In cell "2.1 Launch MD simulation", the default option "AUTO" automatically sets the simulation time based on the sequence length. Alternatively, it is possible to provide a simulation time in ns.

    6b. In cell "3.3 Analyze reweighting", the default option "AUTO" for "THETA_LOCATOR" automatizes the choice of θ (free-parameter that sets the balance between experiment and simulation during reweighting; see dedicated literature for a more exhaustive explanation). If you wish more control over the reweighting, you can set the "THETA_LOCATOR" to "INTERACTIVE".

7. Cell "2.2 Calculate structural observables from simulation", will plot the distributions and averages of some structural parameters: radius of

gyration (R_g), hydrodynamic radius (R_h), end-to-end distance (D_{ee}), scaling exponent (ν) from the fitting of the scaling profile.

8. Cell "3.3 Analyze reweighting" analyzes the results of the reweighting executed in cell "3.2 Execute reweighting". If the "AUTO" option is set for the "THETA_LOCATOR", check that in the resulting plot of ϕ_{eff} vs. χ_r^2, the chosen θ (in red) is located at the **elbow of the curve**. If this is not the case, or if you would prefer to select a different theta, set "THETA_LOCATOR=INTERACTIVE". A different ϕ_{eff} vs. χ_r^2 will appear, reporting on the θ value for each point. A prompt will appear below the plot, where you can provide a value for θ. You will then obtain a plot similar to that from cell "2.2 Calculate structural observables from simulation", but where also the reweighted quantities are plotted.

9. Finally, cell "4. Download results" will trigger the download of a zip archive containing the data from the simulation and reweighting. A **README** file is included that explains the content of the zip archive.