Class11AF

Ksenia Mekhonoshina

Background

In this hands-on session we will utilize AlphaFold to predict protein structure from sequence (Jumper et al. 2021).

Without the aid of such approaches, it can take years of expensive laboratory work to determine the structure of just one protein. With AlphaFold we can now accurately compute a typical protein structure in as little as ten minutes.

The PDB database (the main repository of experimental structures) only has *~250k* structures (obs. from last lab). The main seq database has over *200mil* seq.

Thats is a 0.125% coverage. Meaning only 0.125% of known sequences have a known structure - this is called a "structure knowledge gap".

Structures are much harder to determine than seq: they are expensive (~1mil$) and they take 3-5 years to solve.

EBI AlphaFold Database

The European Bioinformatics Institute (EBI) has a database of pre-computed AlphaFold (AF) models called AFDB. This is growing all the time and can be useful to check before running AF ourselves.

Running AlphaFold

We can download and run locally (on our own computers), but we neeed a GPU. Or, we can use "cloud" computing to run this on someone elses computers ;-).

We will use ColabFold https://github.com/sokrypton/ColabFold

>HIV-Pr
PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQ