Class 11 Alpha Fold

Eric Wang A17678188

Background

In this hands-on sesison we will utilize AlphaFold to predict protein structure from sequences Without the aid of such approaches, it can take years of expesive laboratory work to determine i f the strucature of one protein. With AAlphaFold we can now accurately compute a typical protein structure in as little as then minures'

The PDB database (the main repository of experimental structures) Only has ~**250 thousand** sequences! only 0.125 % of knonw sequences that a knonw structure - Only 0.125% of konwn sequences have a known structure- this is callefd the "structure knowledge gap."

```
250000/200000000*100
```

```
[1] 0.125
```

Structure are much harder to detmermine than sequences. THey are expensive on average ~ 1 illion each. They take on average 3-5 years to solve.

EBI AlphaFold Database

The EBI has a database of pre-computed AlphaFold (AF) modles called AFDB. ## Running AlphaFold

We can download and run locally (on our own computers) but we need a GPU. Or we can use "cloud" computing to run this on smoeone elses computers :-)

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

We previously found there was no AFDB entry for our HIV sequence:

```
>HIV-Pr-Dimer
PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLK
EALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPT
PVNIIGRNLLTQIGCTLNF
```