**ECS 129**

**Project3**

**Protein Structure Prediction**

**Due:** Thursday, February 29, 2024

**Protein geometry**

Predicting the structure of a protein remains a formidable task. However, in the last three years, AlphaFold and its current version, AlphaFold2, have proven to be quite successful in solving this challenge, as demonstrated in successive protein structure prediction challenges (CASP). In this assignment, you will:

* Predict the structure of two protein sequences, using AlphaFold2
* Compare the predictions with the “gold standard”, i.e. the experimental structure of those proteins (see below)
* Discuss the results.

**The two protein sequences**

*Sequence 1:*

> Fimbrial adhesin|Proteus mirabilis (strain HI4320) (529507)

SIFSYITESTGTPSNATYTYVIERWDPETSGILNPCYGWPVCYVTVNHKHTVNGTGGNPAFQIARIEKLRTLAEVRDVVLKNRSFPIEGQTTHRGPSLNSNQECVGLFYQPNSSGISPRGKLLPGSLCGIAPPP

***Sequence 2:***

>CST complex subunit CTC1|Homo sapiens (9606)

AISQAIIRLLVEDGTAEAVVTCRNHHVAAALGLCPREWASLLD

**Predicting the structures of the two protein sequences.**

You will use AlphaFold2, available at:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

(see <https://www.youtube.com/watch?v=mTjYvIU3KCY> for how to use it)

**Comparing with the gold standard**

Sequence 1 corresponds to the protein structure identified as 6Y4F chain A from the PDB ([www.rcsb.org](http://www.rcsb.org))

Sequence 2 corresponds to a fraction of the protein structure identified as 6W6W chain B from the PDB.

For convenience, I have provided the corresponding PDB files on the web page (those PDB files contain the part of the experimental structure corresponding to the sequences given above).

*There is no need to send a lengthy write-up, but it should definitely include an introduction, results and analysis, and a conclusion.*

**Good Luck !**