# Bioinformatics Project Proposal

## Najib Ishaq

### Project Name

CASP (Critical Assessment of protein Structure Prediction)

This is a biennial competition in Bioinformatics where teams are tasked with predicting the structure (among other things) of proteins given the ammino-acid sequences.

### Outside Help

- Course professors for what they know about protein folding.
- My friend Tom Howard, should he have time, because we work very well together.
- Strangers on the internet when I inevitably post on online forums asking questions about various snags I hit.

### Plan for the Project

I will tackle structure prediction using various neural networks. My plan so far is as follows:

1. Use convolutional layers, autoencoders, and other methods for feature extraction from the raw ammino acid sequence.
2. Feed the features to various types of networks to predict structure.

As a baseline, I will use a convolutional neural network. I will then try various combinations to enhance my performance. I will use the same metrics that are used by the judges for the CASP competition.

My minimum goal for the semester will be to produce my baseline model and at least three iterations of improvement upon it.

In addition to this minimum, I think I will be able to use Recurrent Neural Networks to read the sequences and produce attention matrices that help look at long-term dependencies of the structure on the sequence.

As a lofty goal, something almost certainly outside my reach for this semester, I will try to replicate what Google's DeepMind did for the competition in 2018.

### Team Members

- Me.

- Just me.

**Team Leader**

- I am the leader by default.

**Schedule**

- One of the (occasional) downsides of being me is that I am forever saddled with my own company. Sometimes I amuse myself and at other times I frustrate myself. I will be in the lounge on floor 1 of Tyler on Wednesdays from 2pm to 5pm.
- Please note that I too may drop in on any of you to ask questions about relevant topics, talk about life, and possibly share dark chocolate.

**Resources Needed**

- I will read through several papers in the field. I have already started looked at the paper DeepMind released on AlphaZero. I am currently reading Dr. Daniels' thesis.
- I have my own GPU to run some serious computations but will likely need access to some better machines on campus. I already have accounts on Dr. Daniels' machines will discuss specifics with him when I feel I have something that needs the extra computational power.
- Any other resources that you think will be useful to me.