**Project Title:** Predicting Contact-maps from protein sequence using variational autoencoders

**Name(s):** Anowarul Kabir

**Project description:** Knowing tertiary structure is critical for protein sequence to understand its functionality and characteristics. To simplify the problem of mapping sequence to tertiary structure generation, it would be a fine idea to predict contact maps or distance matrix from sequences, since contact-map representation is the middle ground of sequence and tertiary structure. It can be further extend from contact map to 3d coordinate generation problem, if time permits. Analogically, The problem can be similar as text to image generation. So the input is low dimensional and output is high dimensional and more robust. As we learned from the class, this type of problems can be attacked using variational autoencoder.

**Project goal:**
1. To understand and analyze the characteristics of variational autoencoders in the specific problem domain.
2. Both inputs and outputs can be of different shape, so to understand how to handle the situation for different input and output shapes

**Project timeline:**
1. Distant matrix generation from known protein 3d structures and create train, test and validation set.
2. Design custom or use known variational autoencoder model for translating sequence representation to distance matrix representation.

**Data:** The data is not fixed yet.

**Resources/Key references:**