**Project summary**

Biological background: Nucleosome is a basis unit of DNA packaging in eukaryote cells.

Nucleosome form the fundamental repeating units of eukaryotic chromatin, which is used to pack the large eukaryotic genomes into the nucleus while still ensuring appropriate access to it.

The nucleosome consists 147 base pares (bp) of DNA.

The nucleosomes organization in the DNA is influenced by two factors:

1. The DNA sequences (There are specific locations on the DNA that the nucleosome tends to be sitting on, and vice versa).
2. Proteins in the cell.

Secondary goals: Build a model that will give us the probability of a specific nucleosome organization on a DNA molecule **in vitro.**

We are interested in the probability **in vitro** because we want to isolate the influence of the DNA sequence on the nucleosome organization in the cell. Also, in doing so, we are hoping the we could isolate the influence of the **proteins** on the nucleosome organization in eukaryote cells.

Plan:

1. Raw binding model: create a model f, so for every sequence of 147 bp (base pairs) the model will give us the binding probability of the sequence by a nucleosome.

Function description: we will have a pwm (probability weight matrix) of a nucleosome, and a position weight matrix of the specific sequence.

Then, we can multiply the two matrix. The scalar multiplication of the upper diagonals will give us the probabilities of the nucleosome to attach itself onto the sequence start from the i position (i- the number of the diagonal).

1. Occupancy for every sequence: with the model that we have built in stage 1, we will create a model that will give us the probability of being occupied by a nucleosome for every DNA sequence.

Function description: with the model that we built in stage 1, we will calculate the probabilities for every nucleosome possible organization on the cell DNA.

Then, for getting the probability of a specific DNA sequence to be attached to a nucleosome, we will sum up the probabilities for the nucleosome occupancies were the sequence is covered by a nucleosome.

This sum will the probability that the model will return.