**Project Progress 20.4.17**

**Progress so far:**

1. We wrote a python program that is getting as input dna sequences files in "fasta" format, and translate the sequences into postion matrix. after the translation, the matrix will be saved in a file in "pickle" format.

2.We wrote a python program that gets a PWM matrix, and position matrix of a dna sequence, and returns a vector of probabilties. The probabilities in the vector are the probabilities for the biological molecule represented by the PWM matrix to bind onto the DNA sequence.

**Future plans:**

**Future framework:**

1.One important note is that out current model can be valid not only for nucleosomes, but that it can be valid for any biological molecule that are binding dna, like transcription factors, for example. so we want our model to be as abstract posibile, so we will be able to use for a lot of different molecules.

2.Until now, we have assumed that the probabilities of the molecule to bind onto a specific locations are indepndent from each other. We want to make our model more complex by include the influence of every k- molecules on each other. This is also called markov order. We want the user to be able to enter the rquiered markov order, and calculate the probabilities according to that order.

3.We also wants to include the the influence of two binding molecules on each other. if we have two binding molecule, and two binding sites with high probebility score that are very close to each other, so if one of the molecule will bind onto ont site, it will also cover the second, and the next molecule wouldn't be able to bind onto those sites.

We want to include that in our model. This part of the model will touch on the nucleosome side secificlly, because there is a very high concentration of nucleosome on the DNA. In other molecules, like transcription factors, for example, there is lower concentration, so this calculation wouldn't be relevant for them.

**Imeddiate plans:**

1. We want to change the part that creat a postion matrix from a dna sequence. we want to be able to creat a position matrix according to the markov order that will be given as input by the user, with all the permutations of k dna sequence

2.So far, we've managed to creat a position matrix from dna sequences. Now, we want to be able to creat a PWM from a list of binding sites (probably in fasta format) that will be given by the user. we will have two ways of creating the pwm. The method in which the program will use will be picked by the user.

The methods are:

a.Getting a pwm from a file that will include a pwm matrix in advance.

b.Getting a list of dna sequences in fasta format, and creat a pwm from this sequences. we will do it with the next method:

-for every sequence, creat a postion matrix, so we will have n matrixes

-sum all those matrixes.

-multiply the sum by 1/n. we can see it as doing "average" on every postion.

we also want to creat a pwm according to markov order that will be entered by the user. There will be a little compliction with that, but by now we will do it in the intiutive way.