Lab 2 Biostatistics report
The Machinary of life

aneupanetims2016

September 18, 2020

1 Introduction

Proteins of blood cells recognize and destroy the invading cells and viruses.Every living things on Earth uses the similar set of molecules to eat, to breath, to move and to reproduce. Many molecular machines are virtually identical in all living cells.We will look machine molecules,unusual molecular world , how they combine with living cells.

2 Body

We take a data of four thousands protein, half associated with the keyword antibody and the other half not related to the keyword.Each protein is a string of letters representing each amino acid in the sequence. We name the set of antibody protein by X where as not antibody by Y. Both have length 2000.Out of all we choose, the protein of first position. It is denoted by X[0] in python. We use online python colab to understand the name of protein in the given set. While looking to the, it come put huge long name which difficult to understand.Since the name of Amino acid are notified by alphabetical name (to ref Biostatistics Lab 2), we change the letter string of the protein into a vector form using python code.

```
def process_strings(c):
    '''Takes in a list of sequences 'c' and turns each one
       into a list of numbers.'''

    X = []

    for  m, seq in enumerate(c):
        x = []
        for letter in seq:
            x.append(max(ord(letter)-97, 0))

        X.append(x)

    return X
```

Python Sample 1: Mount Google Drive

A vector come out having numbers Where the number represents the alphabet letter such as 0 means A, 1 means B and so on. We make a histogram of the protein Which is as follows We find

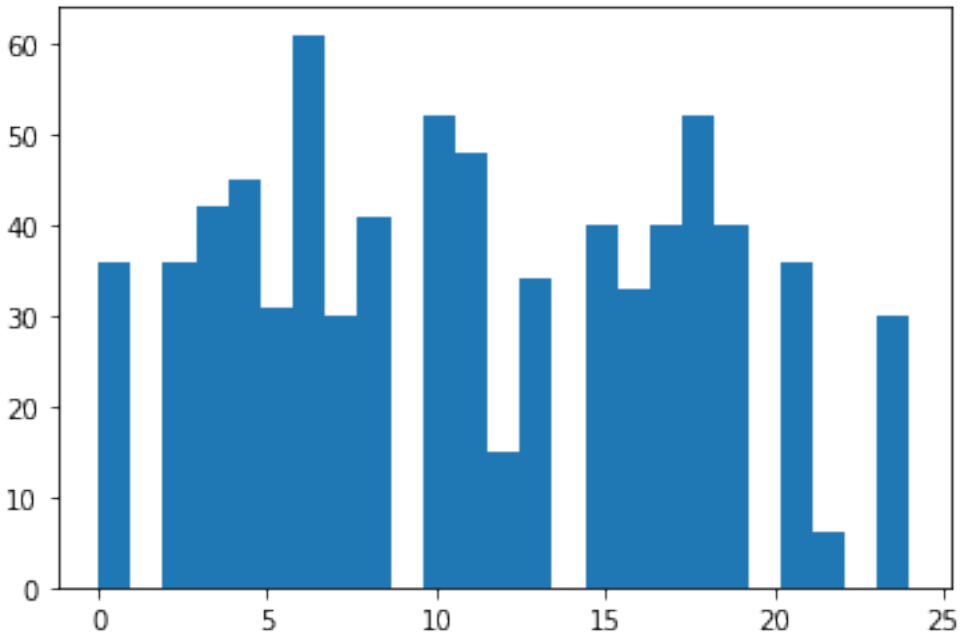

Figure 1: This is a matplotlib Histogram plot for a antibody protein X[0]

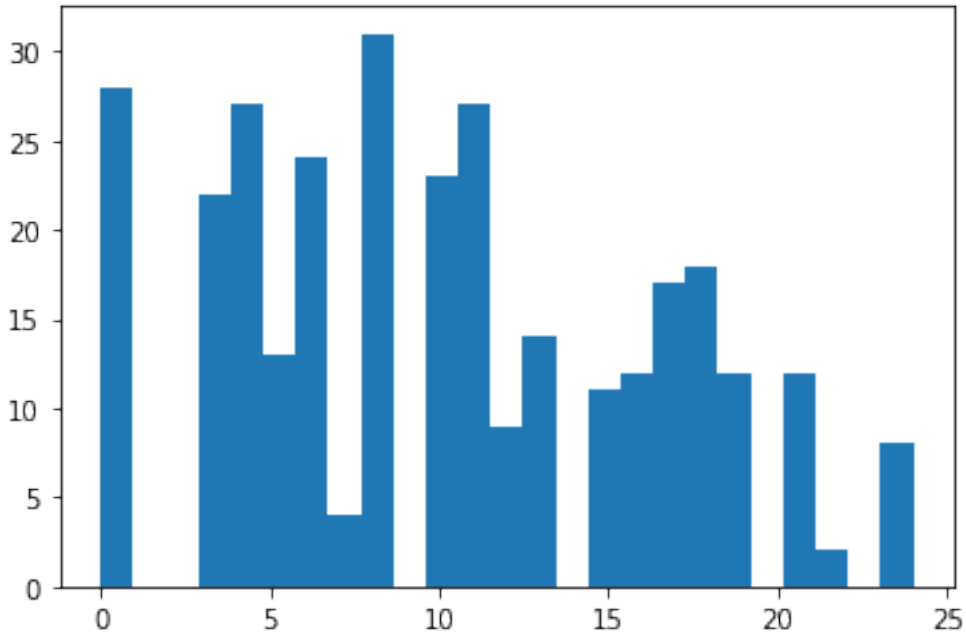

Figure 2: This is a matplotlib Histogram plot for a not antibody protein Y[0]

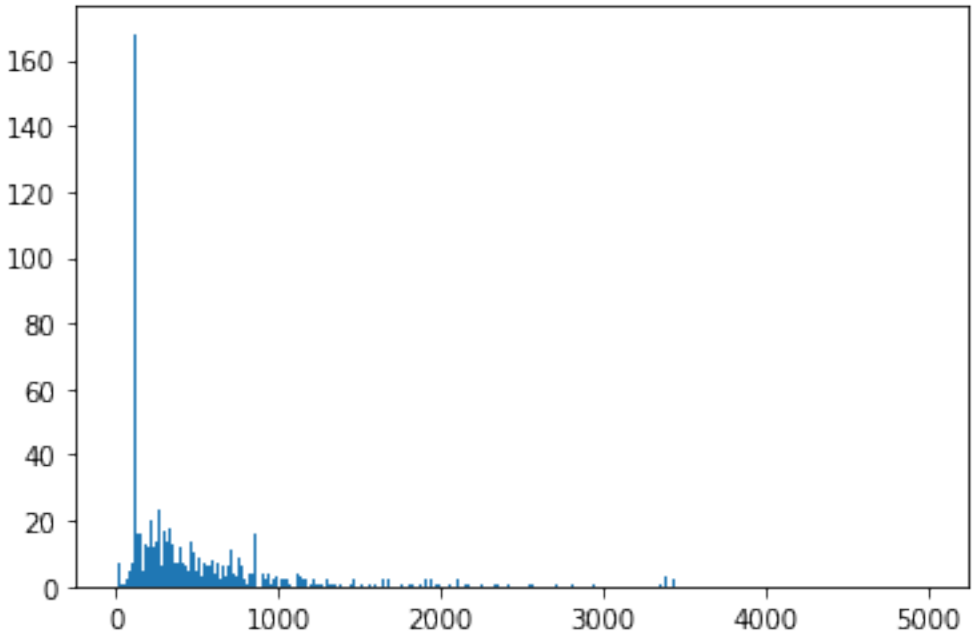

Figure 3: This is a matplotlib Histogram plot for a antibody protein X

out mean amino acid and how the amino acid deviated in the protein.Which are given in [1]. We find out length of all proteins of the given data set. We find out not only protein of having max lenght but also min length and the histogram of those are as follows:

3 Conclusion

We study about protein which is used by body to fight with Bacteria Viruses. Protein is made up of combination of amino acids and we study a protein and amino acids in it and length, deviation as well as compare the length of all protein in the given data set.

References

[CL2] https://github.com/aneupanetims2016/Fall-2020/blob/master/Copy$_o f_B iostatistics_L ab_2.ipynb$

[1] $https://github.com/aneupanetims2016/Fall-2020/blob/master/Life_M achinery_N otes.pdf$

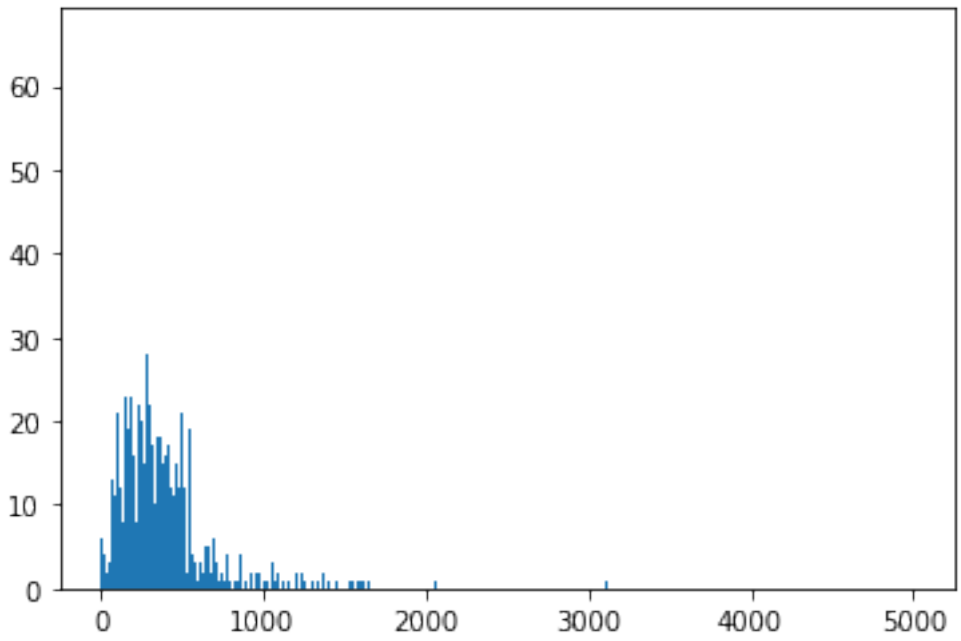

Figure 4: This is a matplotlib Histogram plot for a not antibody protein Y