

Turning biology into mathematics

Lab 02

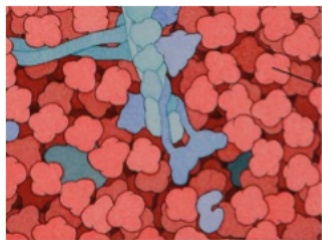
Niranga Mahesh Udumulla¹

Florida Atlantic University

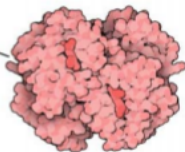
Biostatistics, Fall 2020

Introduction

In this presentation, we will explore this common birthright of molecular machines. We will start with a look at the machines themselves and the unusual molecular world in which they operate. Then, we will explore how they are combined in living cells. Finally, we will look at a few special topics related to our own molecules and cells.



1,000,000 X

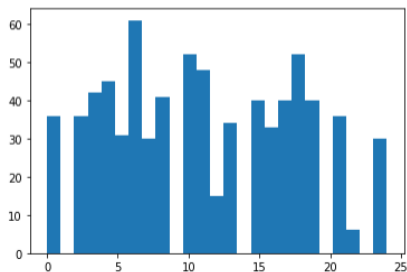


5,000,000 X

Procedure

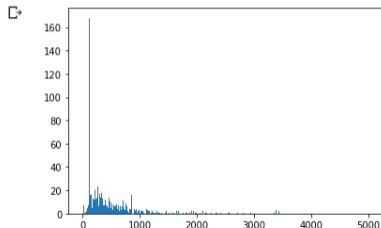
There is a database name uniprot and it has primary structure sequence of amino acids..After that we performed different types of statistical analysis to interpret our data set.Overall, 2000 protein samples have been loaded for our analysis.

Histogram for first 25 observations of the first protein



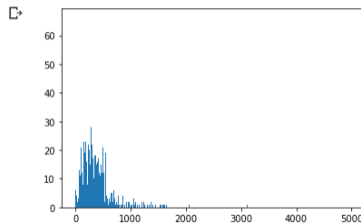
Histogram for first observations of the all protein sequence 1

```
[ ] 1 plt.hist(X_lengths,bins=1000,range=(0,5000));
```



Histogram for first observations of the all protein sequence 2

```
[ ] 1 plt.hist(Y_lengths,bins=1000,range=(0,5000));
```



Conclusion

Unprot data set has 2 data sets and each data set contains thousands of proteins and each protein category has their own values. According to the first sequence, loaded has 2000 proteins and each protein contains different types of values and most of proteins have size below 1000. There are very rare case occur in X and Y data sets which is the values in a protein is greater than 3000.