

# Assignment 1

Mytraya Gattu, 180050032

14 August 2020

## 1 Problem 1

In what follows,  $r$  is the radius of the hemispherical caps,  $l$  the length and  $m$  of the cell. It is given that,

$$r = 0.5\mu\text{m}$$

$$l = 2\mu\text{m}$$

(a) Volume is given by

$$V = \pi r^2 l + \frac{4}{3} \pi r^3 = 2.094\mu\text{m}^3$$

Typical cell volume is  $0.6 - 0.7\mu\text{m}^3$

[https://en.wikipedia.org/wiki/Escherichia\\_coli](https://en.wikipedia.org/wiki/Escherichia_coli)

(b)  $m$  is the sum of the mass of water contained and that of the other substances. Since, density of water is

$$10^3 \frac{\text{kg}}{\text{m}^3} = 10^{-6} \frac{\text{pg}}{\mu\text{m}^3}$$

We have

$$m = 10^{-6} \cdot \left( \frac{2}{3} + 1.3 \cdot \frac{1}{3} \right) \cdot V_{\text{pg}} = 2.304 \cdot 10^{-3} \text{pg}$$

Typical cell weight is  $1 \cdot 10^{-3} \text{pg}$

[https://ecmdb.ca/e\\_coli\\_stats](https://ecmdb.ca/e_coli_stats)

It is evident from the above calculations, that the it is the assumed shape of *E. coli* which deviates greatly from what is true.

## 2 Problem 2

Let at the end of 9<sup>th</sup> cycle there be  $N$  nuclei at the surface. At the end of 13<sup>th</sup> cycle, therefore there are  $16N$  nuclei, which is given as  $\approx 6000$ . Hence,

$$N \approx 375$$

Since, every embryo must start from a single fertilized nucleus, at the end of 9<sup>th</sup> cycle, there must be  $2^9 = 512$  nuclei in total. Therefore, the fraction of nuclei which migrated to the surface is

$$\frac{375}{512} \approx 0.7324$$

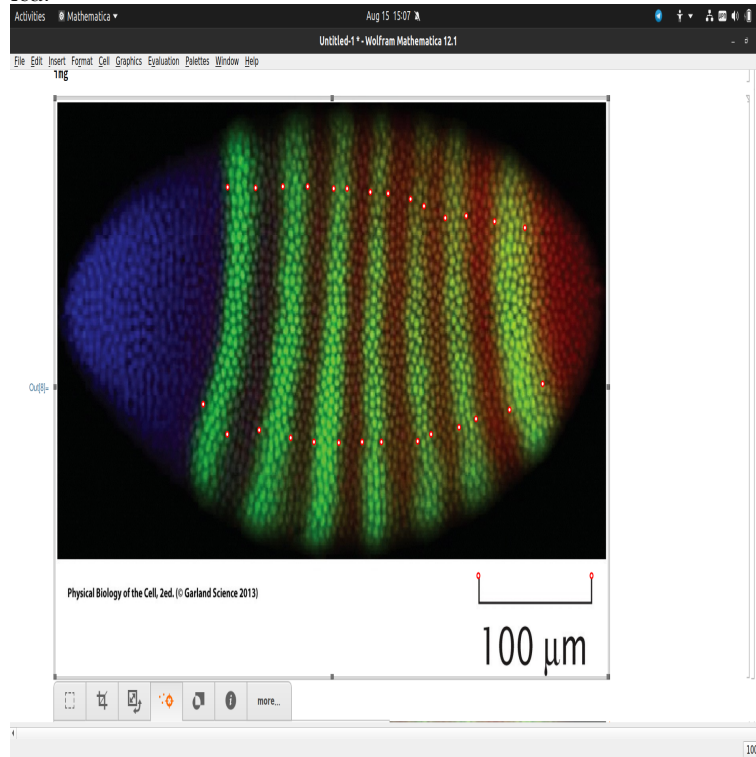
Following the notation of Problem 1, surface area of the spherocylinder is

$$A = 4\pi r^2 + 2\pi r l = 4.3982 \cdot 10^{-7} \text{m}^2$$

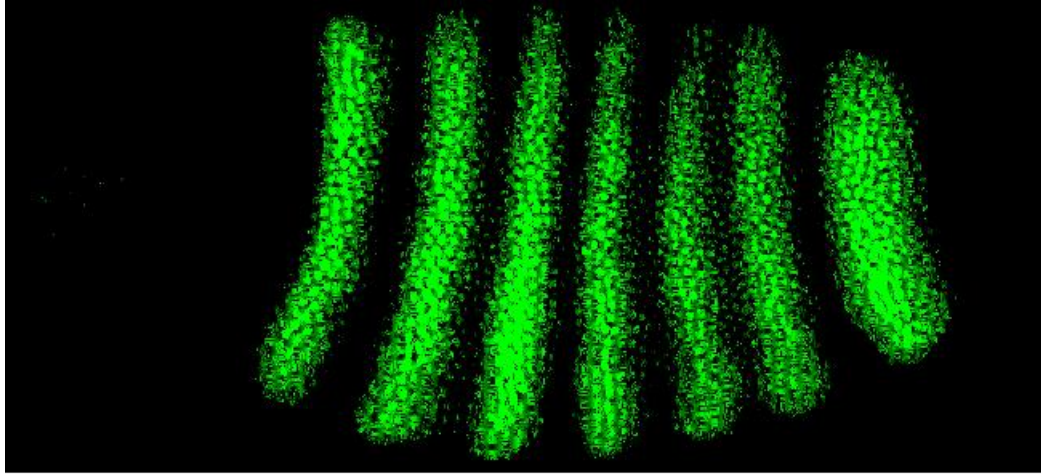
Therefore, the areal density is

$$1.3642 \cdot 10^{10} \frac{\text{nuclei}}{\text{m}^2}$$

To find the length of the green region, I used the co-ordinate tool in Mathematica.



Length of green region is  $\approx 20.36\mu\text{m}$ . Area is calculated, by isolating the green regions and finding the ratio between the number of pixels contained in these regions and the rest of the image.



Using the areal density, number of cells in the green region(Even-skipped) is  $\approx 198$ . Similarly, number of cells in the red region(Caudal) is  $\approx 139$  and number of cells in the blue region(Bicoid) is  $\approx 58$ . (Counting as a single surface)

### 3 Problem 3

#### 3.1 E. coli

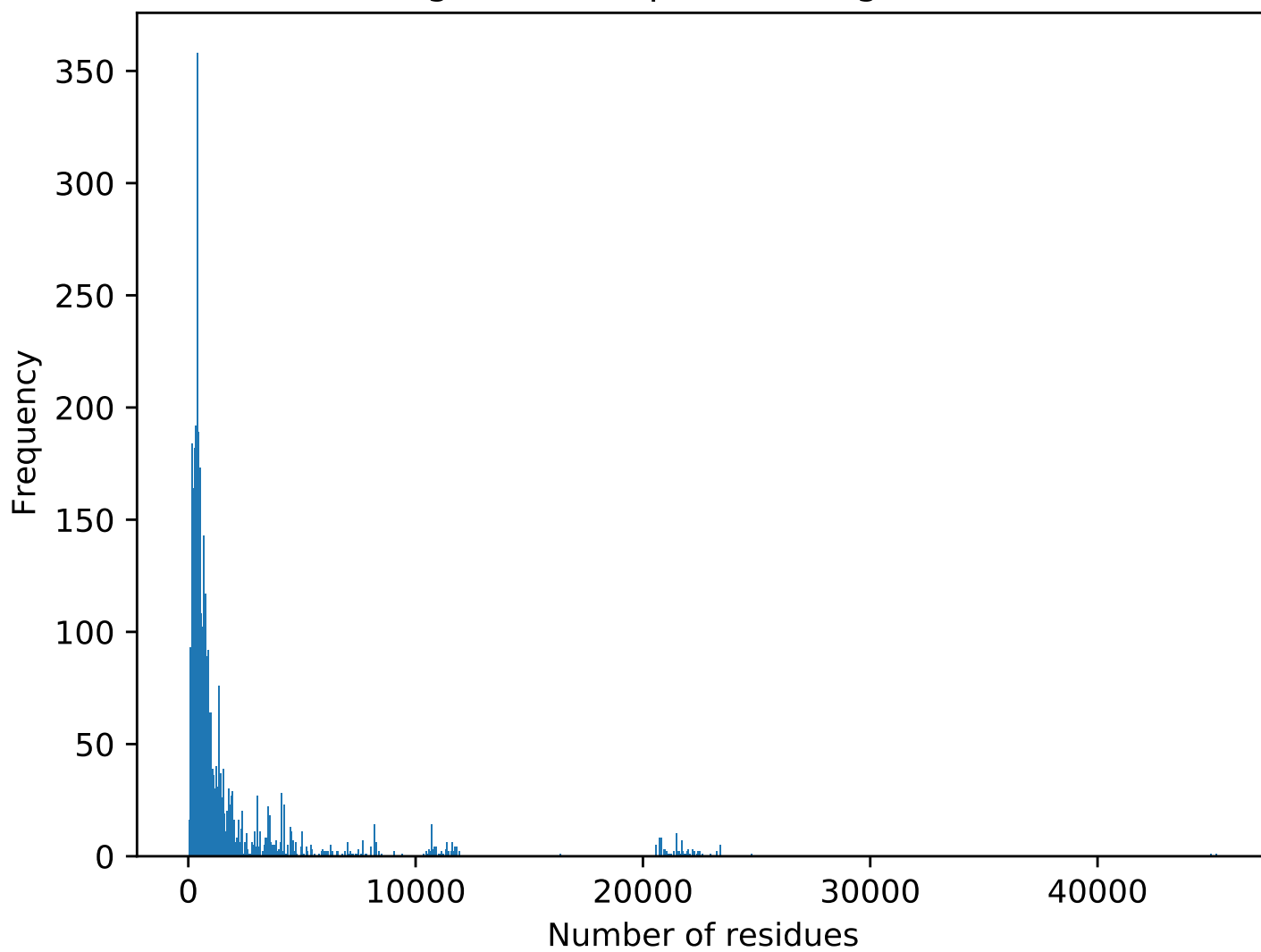
Mean of sequence length: 1892.00

Standard deviation of sequence length: 3872.73

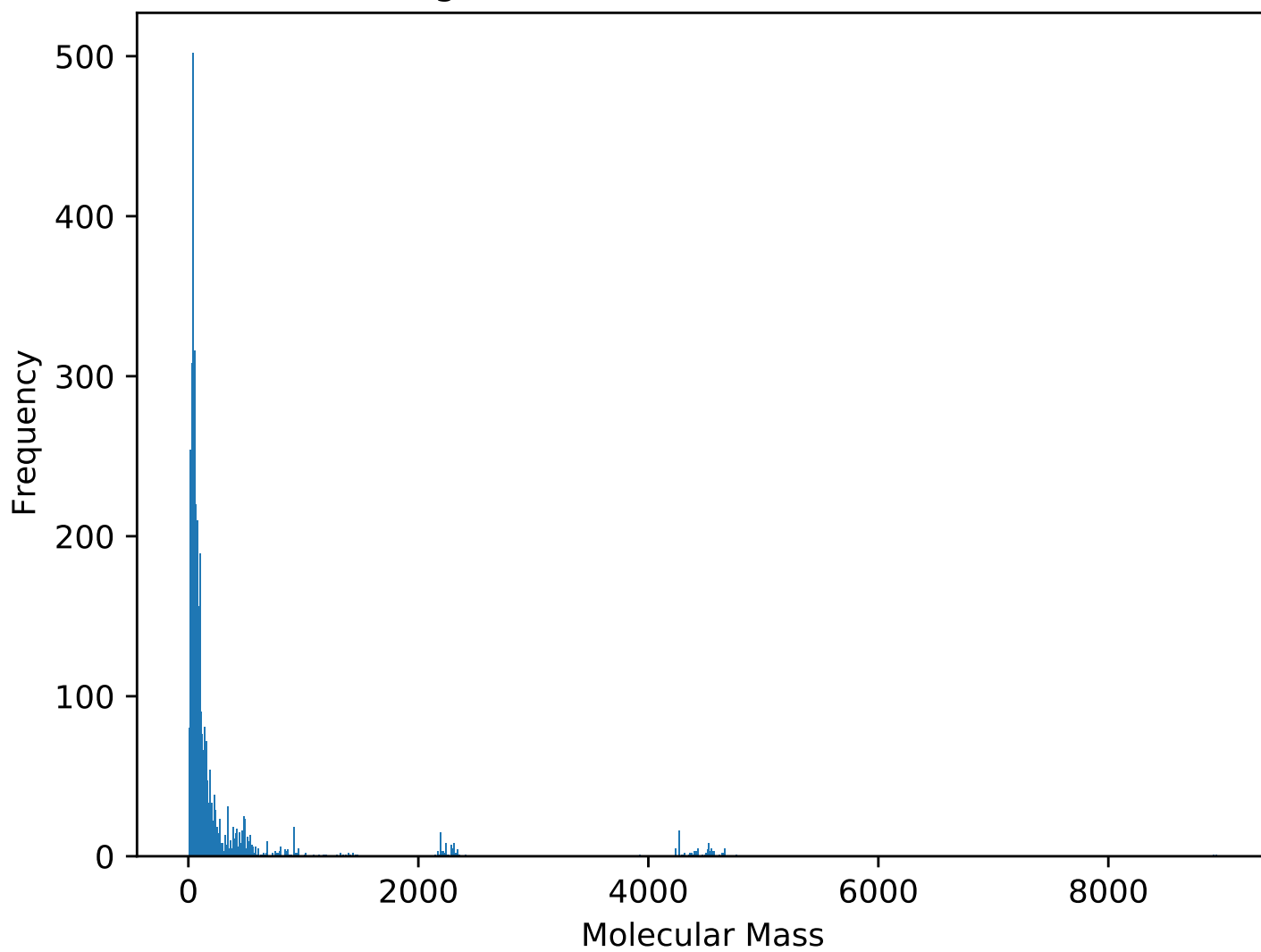
Mean of molecular mass: 289.03

Standard deviation of molecular mass: 775.76

Histogram of Sequence Lengths- E. coli



Histogram of Molecular Masses- *E. coli*



### 3.2 *S. cerevisiae*

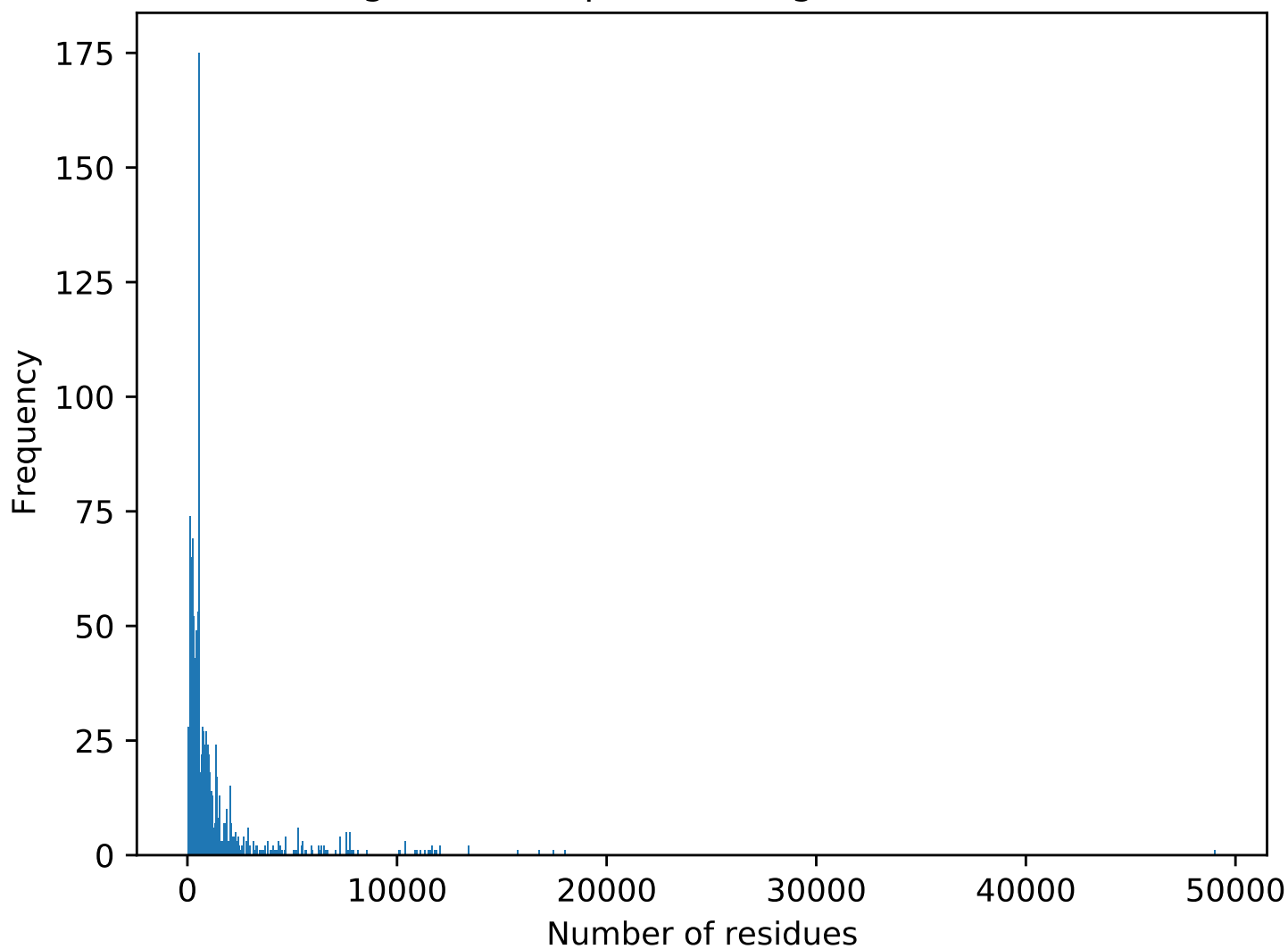
Mean of sequence length: 1287.8953846153847

Standard deviation of sequence length: 2444.4862593462344

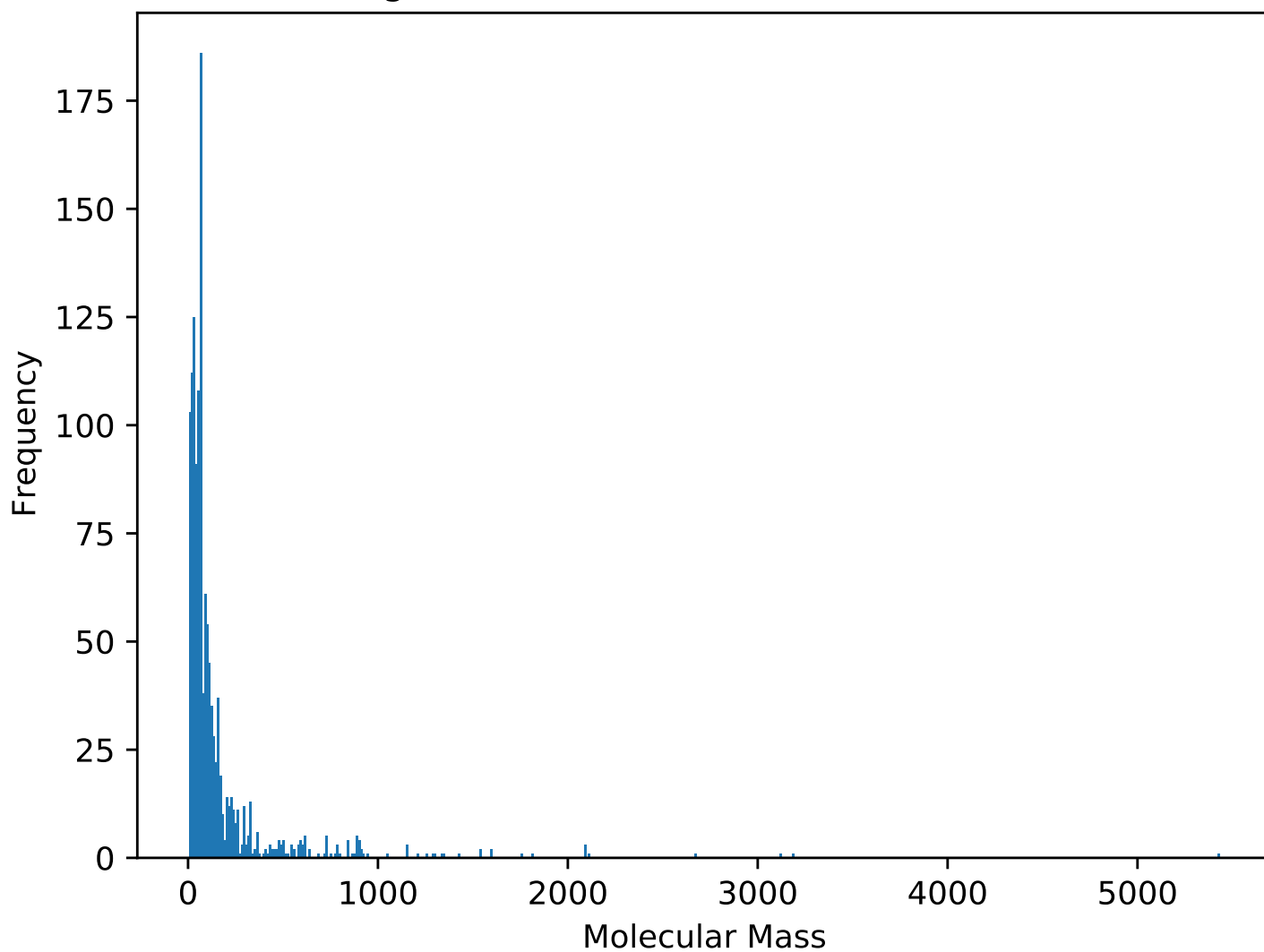
Mean of molecular mass: 154.77802307692306

Standard deviation of molecular mass: 313.30732277898954

Histogram of Sequence Lengths- *S. cerevisiae*



Histogram of Molecular Masses- *S. cerevisiae*





### **3.3 Humans**

Mean of sequence length: 630.7416856492027

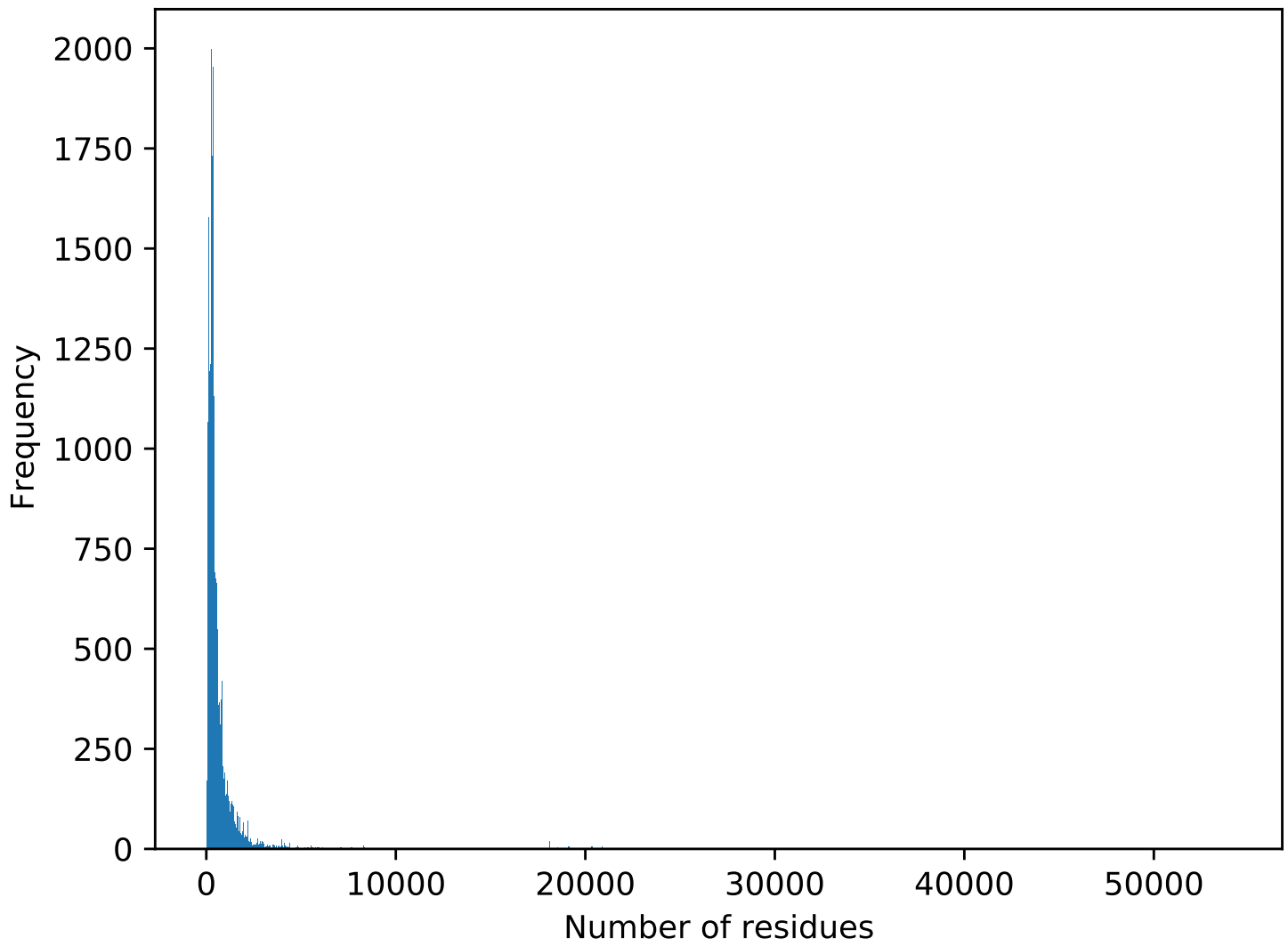
Standard deviation of sequence length: 1263.078244813831

Mean of molecular mass: 74.20564692482915

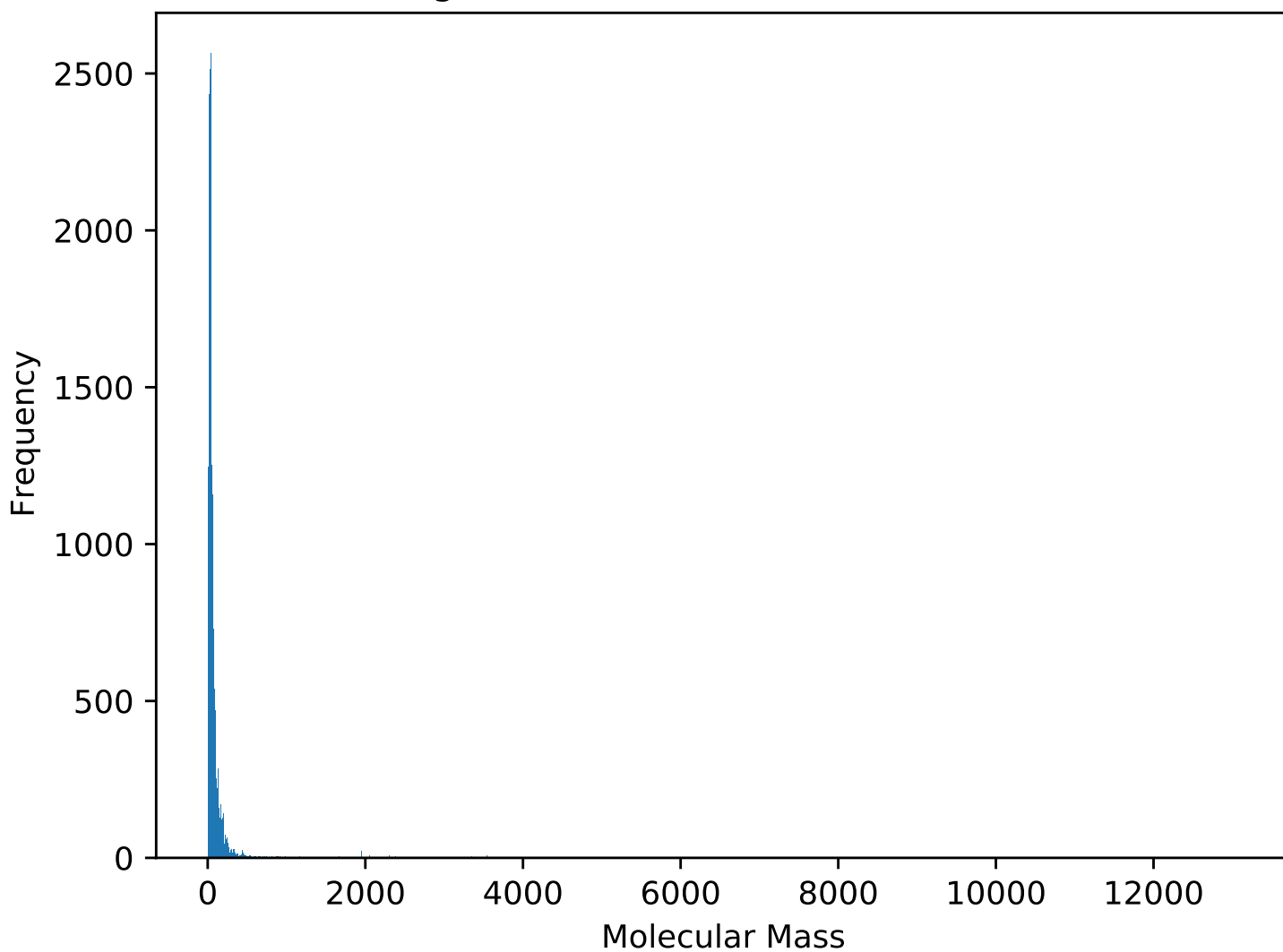
Standard deviation of molecular mass:

181.92998156053864

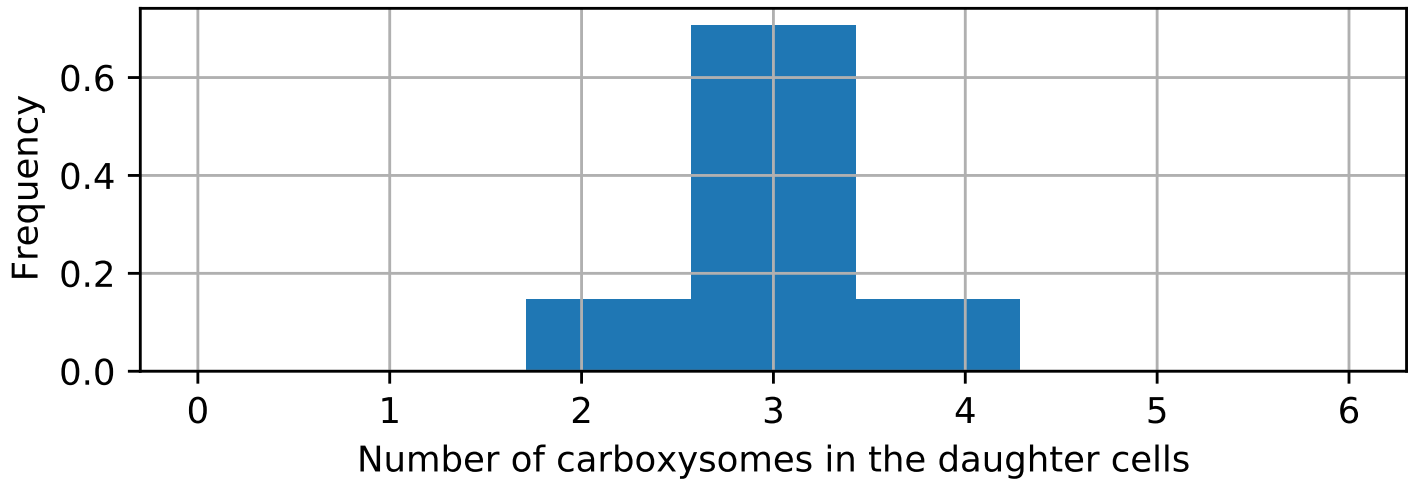
Histogram of Sequence Lengths- Humans



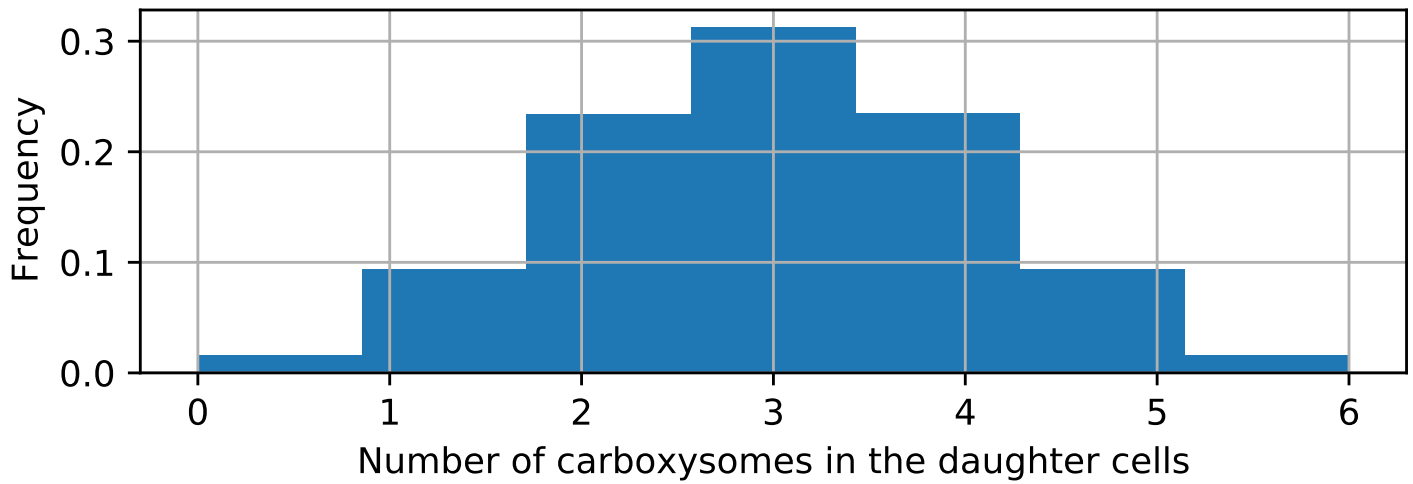
Histogram of Molecular Masses- Humans



Experimental



Theoretical



(In the previous page is the histogram.)

## 4 Problem 4

The probability of having 0 and 6 is

$$p^3 q^3$$

. The probability of having 1 and 5 is

$$\binom{3}{1} p^4 q^2 + \binom{3}{1} p^2 q^4$$

The probability of having 2 and 4 in one cell is

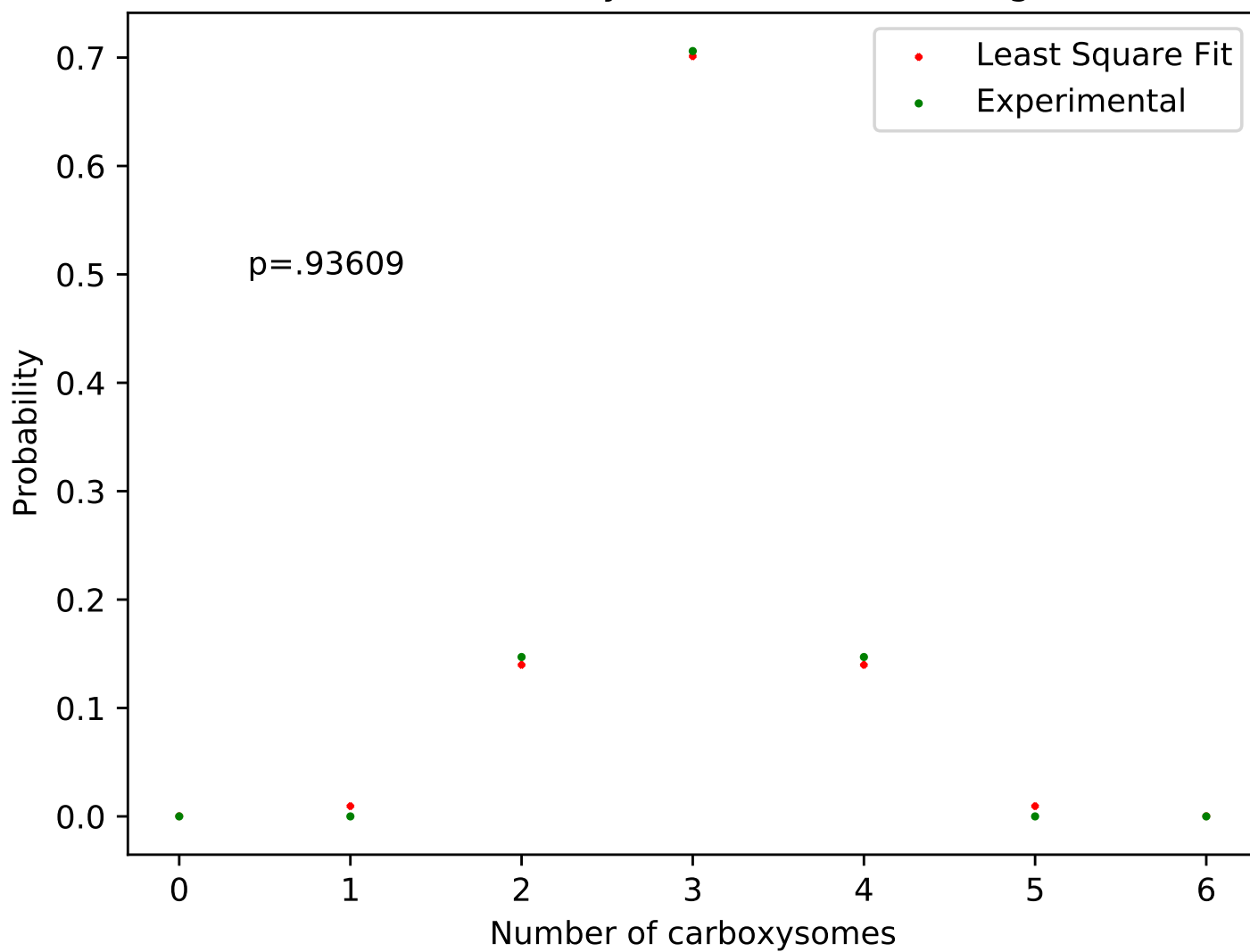
$$\binom{3}{2} p^5 q^1 + \binom{3}{2} p^1 q^5 + \binom{3}{1} \binom{3}{1} p^3 q^3$$

The probability of having 3 in one cell is

$$p^6 + q^6 + \binom{3}{1} \binom{3}{2} p^2 q^4 + \binom{3}{2} \binom{3}{1} p^4 q^2$$

By substituting  $p = 0.5 + x$ , and minimizing using the least-squares method with respect to the experimental distribution, obtained value for  $p$  is 0.936093. Attached, is a scatter plot showing the deviation. Given that, there are only 6 molecules, and the error is about  $0.007 \pm 0.001$  (for the non-zero values), I believe it's a good model.

Distribution of carboxysomes between daughter cels



## 5 Problem 5

- (a) There are  $4^3$  possible ways of choosing the bases to form a triplet, each with the same probability. Therefore,

$$p_s = \frac{3}{64}$$

- (b) There are 64 possible codons. Three of these are not allowed (The stop codons), to form an ORF. Therefore, the number of acceptable possibilities is  $(64 - 3)^N$ .

Required probability is  $(\frac{61}{64})^N$ .

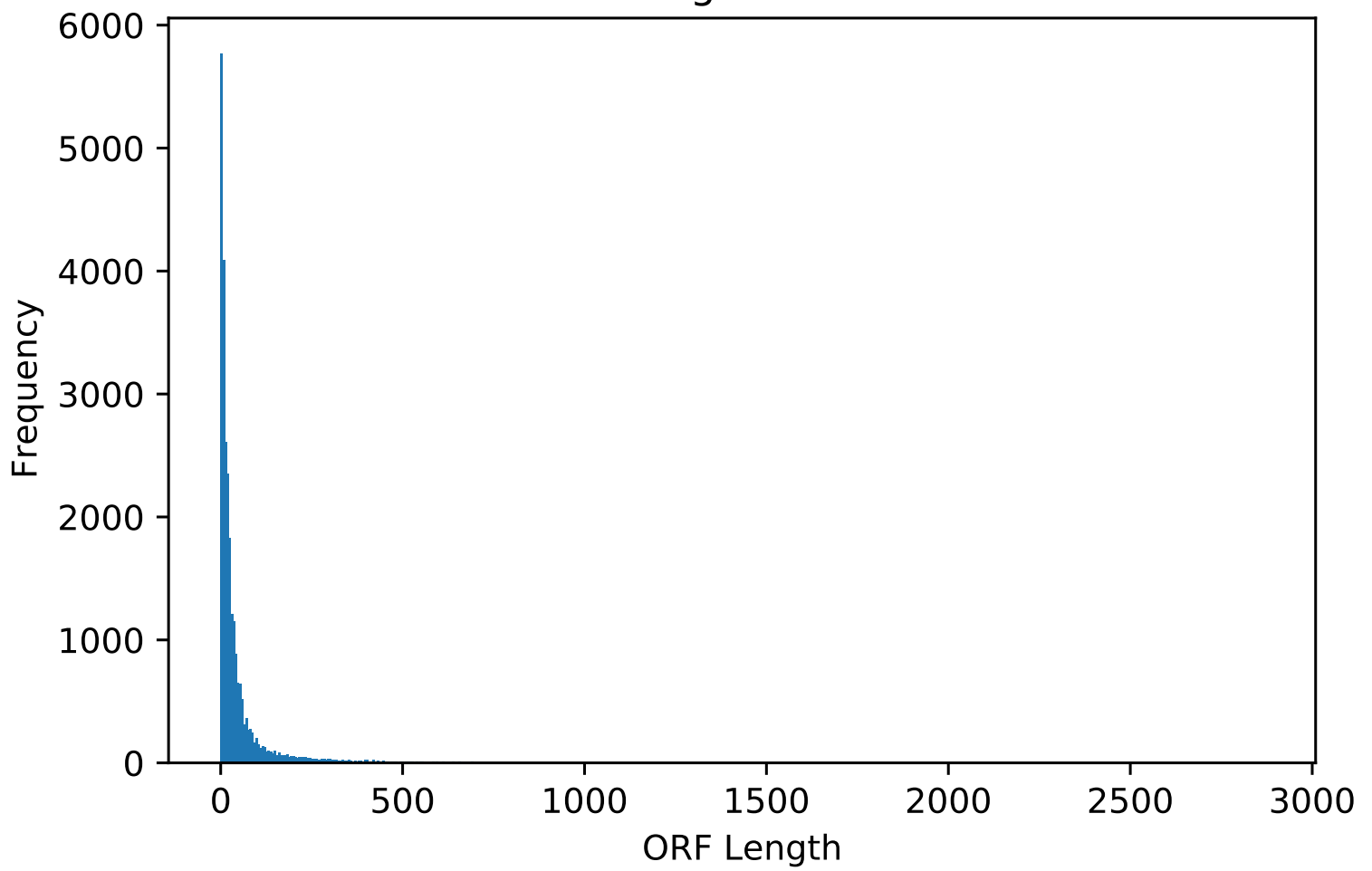
- (c) Since the DNA is circular, selecting only one junction at which to start/stop reading the codon sequence suffices.

Since, there are 4 bases, this can be done in  $\binom{4}{2} = 6$  ways.

- (d)  $\lfloor \frac{5 \cdot 10^6}{3} \rfloor = 1666666$  and the remainder is 2. Therefore, there are 1666666 codons. Now, ORFs of length 600 are desired. We can have an utmost of  $\lfloor \frac{1666666}{602} \rfloor = 2768$  ORFs (Since, at the ends of an ORF are a start and a stop codon). Each ORF (along with the start and stop codons) can be chosen in  $3 \cdot 61^{600}$  ways.

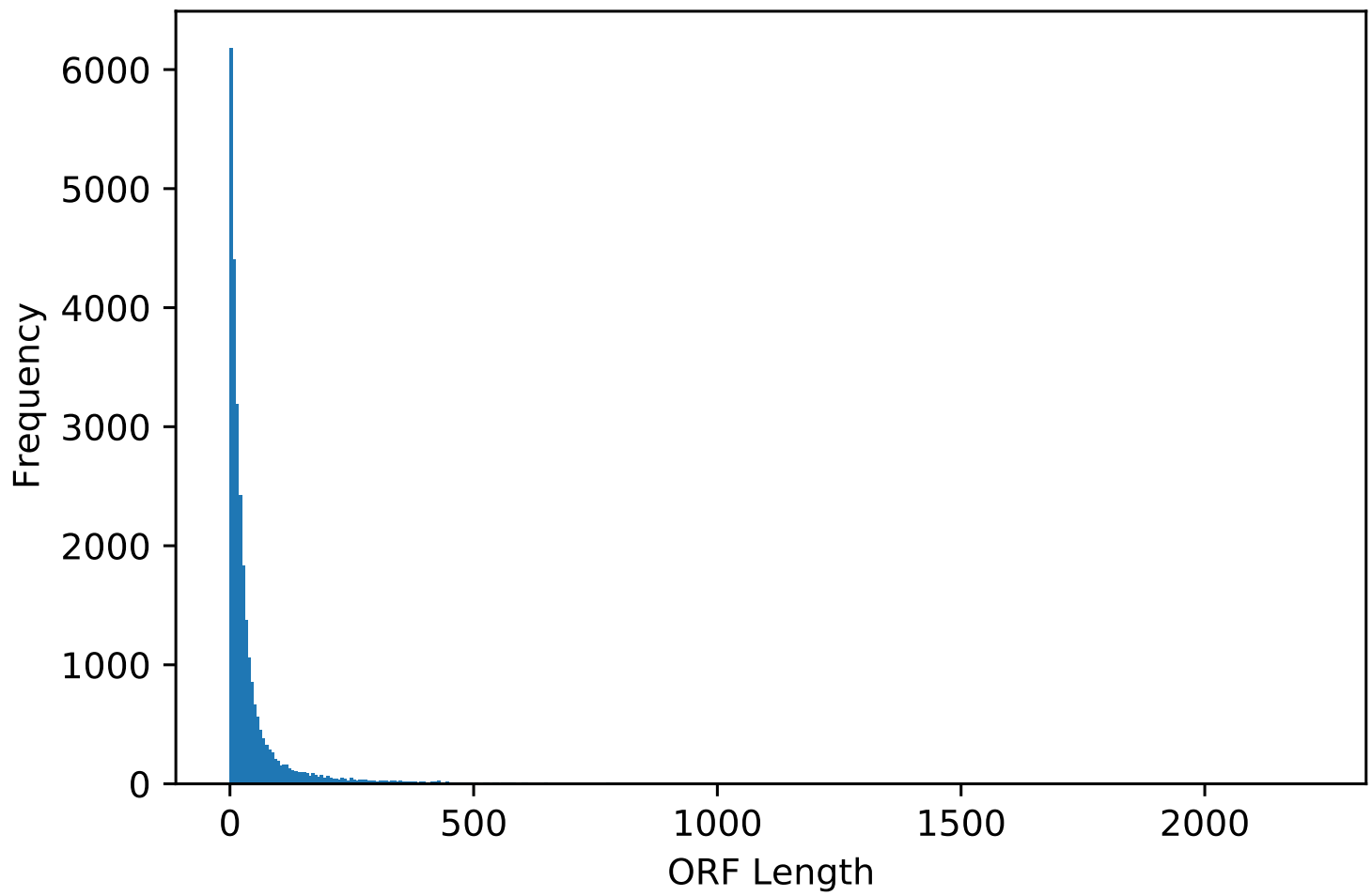
- (e)

Reading frame: +0





Reading frame: +1



Reading frame: +2

