

The MATLAB file in this work is:

- *a7ssa.m*: a code to simulate the three-stage model of gene expression by the Gillespie stochastic simulation algorithm (GSSA).

Figure 0.1 shows the protein distribution for three different values of the parameters (the parameters are the same as ones in the paper). By comparison with the figures in the paper, I can say that the simulation works and I am in the right track!

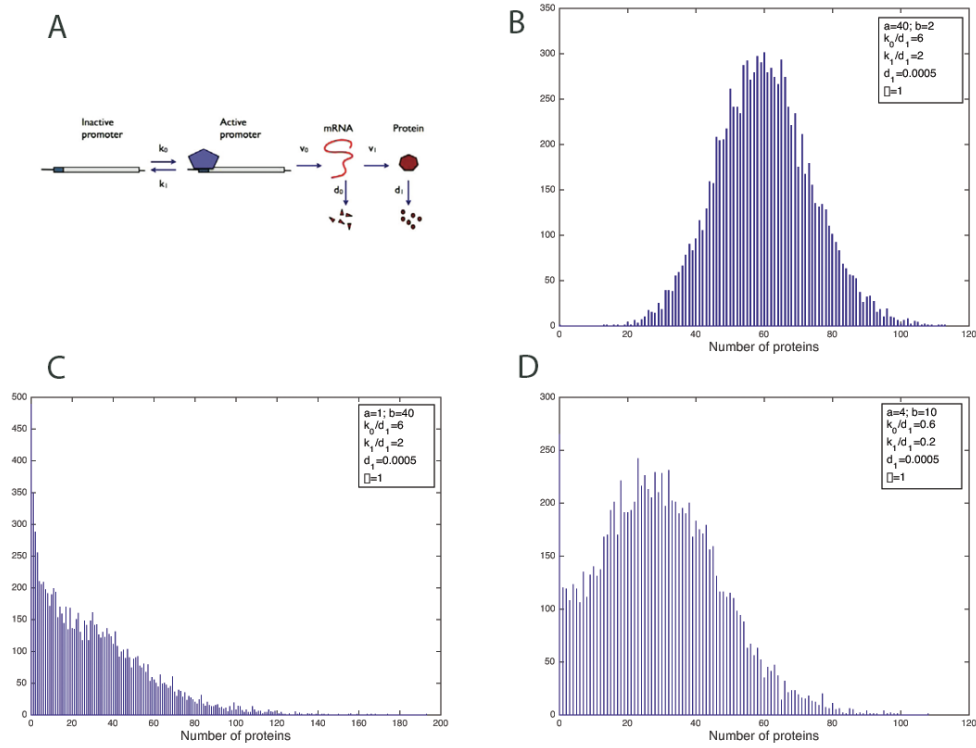


Figure 0.1: protein distributions for different values of parameters. Here I just tried to get the same results obtained in the paper. Schematic model of the three stage gene expression is from the paper.

1- To answer the first question, I added a few lines to the code (*a7ssaCV.m*) to consider the mean, the variance and the coefficient of variation (CV) of the model. I found that to have a small, tightly-controlled number of proteins, the lifetime of the protein should be greater than lifetime of the mRNA ($d_1 < d_0$). In other words, the mRNAs should be

degraded faster than the proteins. To check that I run the simulation for different values of $\gamma = \frac{d_0}{d_1}$. It is clear that when the γ gets bigger, the variance and the coefficient of variation (CV) becomes smaller (figure 0.2).

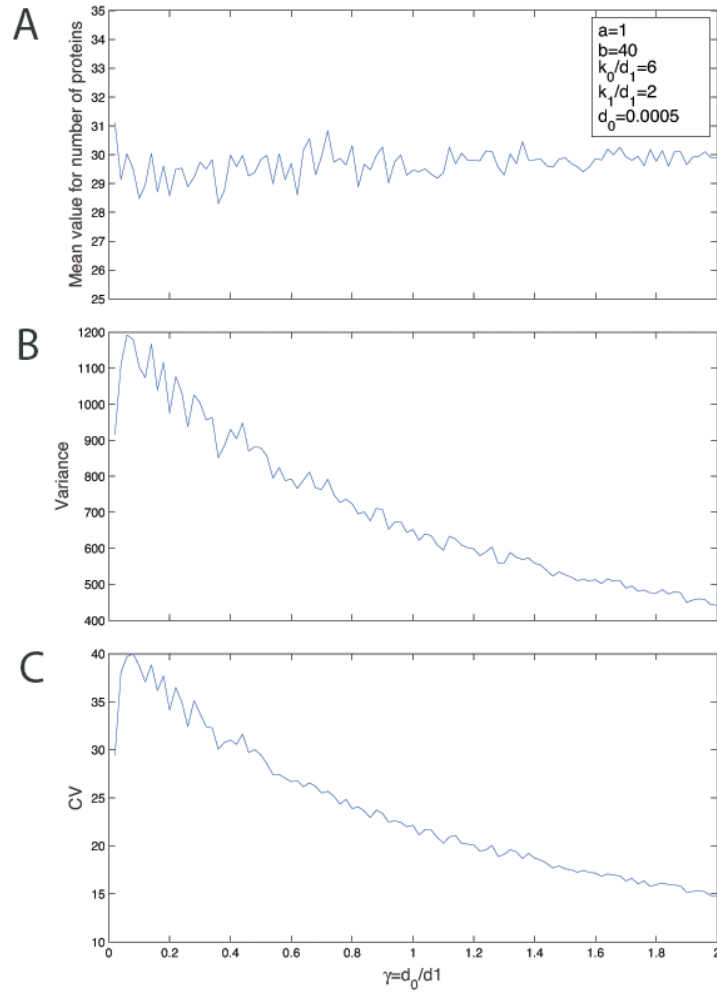


Figure 0.2: The mean number of proteins (A), the variance (B), the coefficient of variation (C) against γ .

2- As mentioned in the paper, the bi-modality arises from the **slow** transitions between active and inactive states. So, to have a bimodal distribution, one condition is that k_0 and k_1 should have small values. In the paper they are $0.6d_1$ and $0.2d_1$ ($d_1 = 0.0005$).

I considered histograms in different values of all involved parameters (the rate constants). In figure 0.3, we have $b > a$ and in figure 0.4, we have $a > b$. It seems that another condition to have two separated modes is: $b > a$.

As shown in figure 0.3, we can just have a bimodal distribution with one peak at **zero** number of proteins and one at non-zero number of proteins.

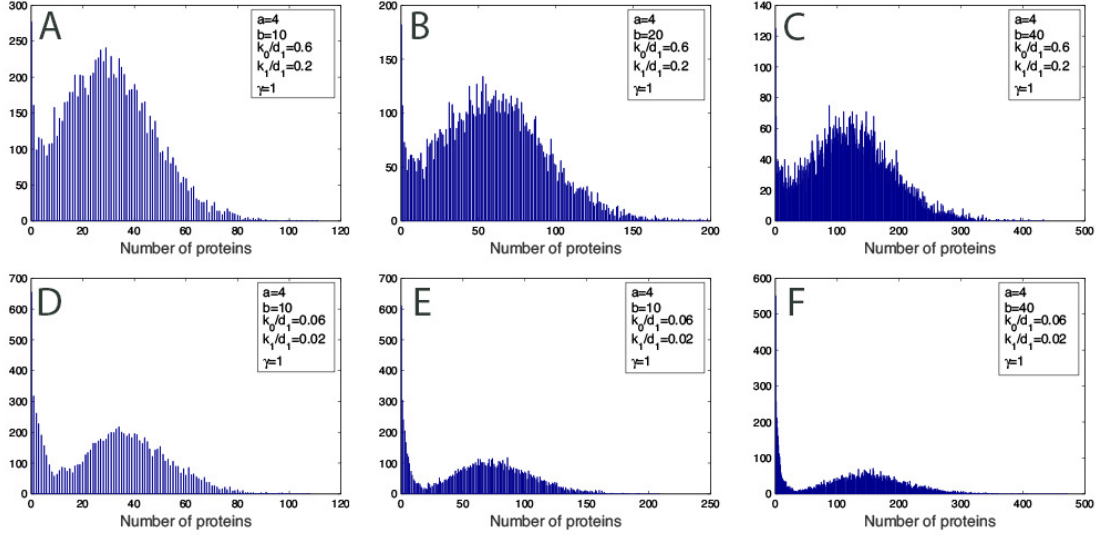


Figure 0.3: distribution of protein numbers for different values of parameters. A, B, C: $k_0 = 0.6d_1$ and $k_1 = 0.2d_1$ but the value of b changes. D, E and F: $k_0 = 0.06d_1$ and $k_1 = 0.02d_1$ but the value of b changes again. Here I just want to show how distribution deforms for different values of b and k_0 and k_1 .

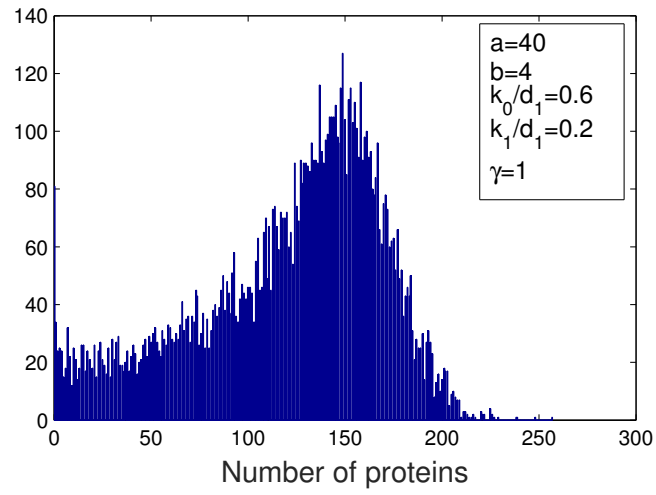


Figure 0.4: Protein distribution when $a > b$