UPPSALA
UNIVERSITET

Partitioning error of transcription factors at cell division

M. Intakhar Ahmad

Degree project in Biology, 2011 (15 ECTS),
Biology Education Centre and Department of Engineering Sciences,
Uppsala University, Sweden.
Supervisor: Johan Elf & Andreas Grönlund

Summary

Transcription factors (TFs) are molecules involved in regulating gene expression in cells. They are usually proteins. Many transcription factors in bacteria are present in less than 10 copies per cell. Inaccuracy in partitioning of these molecules at cell division could lead to a large stochastic variation in copy number between daughter cells with severe consequences for the fidelity in gene regulation and other cellular processes. In this study, we have focused on the fluctuation behavior of transcription factors in the light of previously published theoretical models for noise in cell division. A computer simulation is made that closely resembles a recent (Elf lab- Uppsala, 2011) experimental model system which was done to inspect LacI molecule distribution at cell division in *Escherichia coli* in two different chemical conditions. The model includes a coupled process of stochastic gene expression and partitioning of protein molecules. Preliminary experimental data are used to improve the accuracy of the simulation. With three different distribution principles of molecules over a very long cell lineage reveals the extent of noise related to the particular distribution principle. We have simulated the equal distribution, binomial distribution and the distribution cased by chromosome hitch-hiking mechanism. We found that our experimental data fit well with the newly proposed model of chromosome hitch-hiking.

Contents

Summery		2
Contents		3
1.	**Introduction**	6
2	**Model and Methods**	7
	2.1 Automatic detection of Cell division and generation time	7
	2.2 The *lac* Operon: our model genetic circuit	7
	2.3 Rate of LacI molecule production per cell per generation.	7
	2.4 Time points of new LacI molecule production	8
	2.5 Principles of LacI molecule distribution at cell division	8
	2.6 Representation of LacI molecule distribution variation between two daughter cells	9
3	**Results**	10
	3.1 Selection of cells of interest	10
	3.2 Attributes of the selected cells	10
4	**Discussion**	12
5	**Conclusion**	13
	Acknowledgement	13
	Reference	15
	Appendix	16

Table of figures and tables

2.1	Schematic representation of LacI molecule production in *E. coli* Cells	7
3.1	Calculated parameters for simulation	11
3.2	Graphical representation of statistical significance of experimental data compared to the simulation	12

List of appendix

Serial	Name & file type	Description
1.	GenTimeRaw.mat	Generation time of 151 cells ,calculated experimentally
2.	WithIPTG.mat	Selected cells and the distribution for IPTG condition
3	WithOutIPTG.mat	Selected cells and the distribution for non- IPTG condition
4.	AvgKcount.m	Counts the average molecule production per cell per generation
5.	Kscript.m	Calculates the molecule production rate ,before and after *lacI* replication form the average molecule production rate.
6.	ExperimentCal.m	Calculates the variation parameter(V) in a molecule distribution
7.	MysimulationSim.m	The simulation script for making a loop of many insilico experiment
8.	Mysimulation.m	The main structure of the simulation script
9.	TestDist.m	To test the distribution logic
10.	cdfCal.m	CDF calculation for a normal distribution
11.	Gauss_distribution.m	A function to make normal curve form the simulated data
12.	drawNormal.m	Used to draw normal distribution
13.	SimCon1.m	Simulated data ,condition equal(without IPTG)
14.	SimCon2.m	Simulated data ,condition binomial(without IPTG)
15.	SimCon3.m	Simulated data ,condition hitch-hiking(without IPTG)
16.	SimCon1withIPTG.m	Simulated data ,condition equal(with IPTG)
17.	SimCon2withIPTG.m	Simulated data ,condition binomial(withIPTG)
18.	SimCon3withIPTG.m	Simulated data ,condition hitch-hiking(with IPTG)

List of abbreviation

E. coli	*Escherichia coli*
LacI	Lac opreon Inhibitor or repressor
lacP	Lac Promotor
lacO	Lac operator
IPTG	Isopropyl-β-D-thio-galactoside
CDF	Cumulative Distribution Functions

1. Introduction

Gene expression is a fundamental biological process which organizes specific sets of genes (genomic information) according to physiological or environmental conditions of the organism. Transcription factors (mostly protein in nature) govern the gene expression by binding or unbinding to the promoter of the particular gene, usually a DNA sequence "upstream" from the coding region of the gene. Gene regulation by transcription factors need to be highly specific to manage the cellar behavior properly [1]. Stochastic fluctuation in expression levels is inhabitable due to the low copy number of transcription factors which can bring about significant heterogeneity even in a isogenic populations of cells [2]. The extent of noise in the molecule number of a species is generally captured by the following mathematical term call "Fano" factor (F) [3].

$$F = \frac{\sigma_W^2}{\mu_W},$$

Here, σ_W^2 =variance, μ_W = mean of a random process in some time window W

The same fluctuation principal is also applicable to cell division and the error in equal resource partitioning between two daughter cells can be responsible for persistent fluctuation in all other partition dependent cellular processes [4].

Transcription factors are mostly bound to the chromosome, this property might make them different from most other protein molecules in terms of distribution at cell division. Bound molecules might travel to the new born cells residing on the newly replicated chromosome (termed as “Chromosome Hitch Hiking”). If it is so, the general believe of binomial partitioning of transcription factor could be incorrect.

With the recent experimental development in Elf lab at Uppsala University has managed to count the numbers of LacI molecule (the transcription factor for *lac* operon in *E. coli*) that goes to each daughter cell at cell-division and preliminary results indicates that the partitioning accuracy is better than what can be expected from random partitioning between the daughter cells. The experimentally generated data makes a scope for a computer simulation of the particular cellular event for a very long lineage of cells to widen our view in understanding the extent of fluctuation with probable consequences.

Gillespie's stochastic simulation algorithm (SSA) [5] is very popular in modeling occurrence of stochastic or random events which follows a Markov process with continuous time and discrete-state space. LacI molecule production is a random event which can be simulated with this algorithm. Rate of production can be determined from the experimental data. Three different recourse distribution principles can be taken into consideration. The molecules will divide equally, binomially or follow the hitch hiking

model of distribution. The hitch hiking model depends on the number of binding sites for the transcription factor on the chromosome.

2. Model and Methods

2.1 Automatic detection of Cell division and generation time

Five minutes long time-lapse movies of monolayer of growing *E. coli* are made with Phase contrast microscopy. With the help of microbeTracker [6] aided with some in-house MATLAB scripts, each and every cell throughout all the frames is tracked assigning a specific number for each cell and segmented form one another. A cell division function called “Lambda function” is measured for each cell [Ullman G, personal communication] to distinguish divided and undivided cells. We have also computed that the generation time (T_g) of cells, which is around 24 minutes (the median value of the experimental data of 151 cells) in the provided conditions.

2.2 The *lac* Operon*:* our model genetic circuit

The lactose operon is the perfect genetic circuit for our study. It is the most studied and well characterized genetic circuit so far. We can easily control the different dynamic states of the repressor molecule (LacI) of the circuit by using IPTG (Isopropyl-β-D-thio-galactoside) [7]. LacI has low copy number (about 10 molecules per cell). Moreover, our strain, JE12 has a special feature; the repression has been increased by inserting an O1 in the place of O3 because O1 has better binding strength [8]. It helps in better counting of LacI molecule having less copy number then the wild type. For our simulation the replication point of *lac* gene is crucial for the rate of LacI production. The replication point (T_r) is calculated theoretically by the following equation. Published features of the cell cycle have been taken into consideration [9].

$$Tr = 2.T_g - Rep_i + C\left(\frac{L1}{L/2}\right)$$

Here, Generation time (T_g) = 24minutes ,Replication initiation in mother cell before cell division(Rep_i) = 60minute(a full cell cycle), Replication time for the entire chromosome(C) = 40minutes, Distance between *lacI* and *oriC*(L1) = 1082411bp, Length of the whole genome(L) = 463967bp.

It is turned out that the replication point of *lacI* gene is about 6.8 minuets after birth.

2.3 Rate of lacI molecule production per cell per generation

Fluorescence images of the cells provide us the copy number of LacI molecules per cell. Lambda function is used to determine the cell age and the division feature. The

production rate doubles after the replication of *lacI* gene. The average molecule production per generation is needed to be divided into two parts.

Rate before transcription of LacI in the daughter cell, K1 = $\frac{Km}{\frac{Tr}{Tg}+2.\frac{Tg-Tr}{Tg}}$

Here, average rate of lacI production is Km, replication point of *lacI* gene (Tr) = 6.8 minutes after birth and generation time (Tg) =24 minutes

Rate after transcription of LacI in the daughter cell, K2 = 2K1.

2.4 Time points of new LacI molecule production

The *lacI* gene is transcribed and the molecules are produced. It does not depend on the preexisting number of LacI in the cell. So, Production of a new LacI molecule is a memory less event.

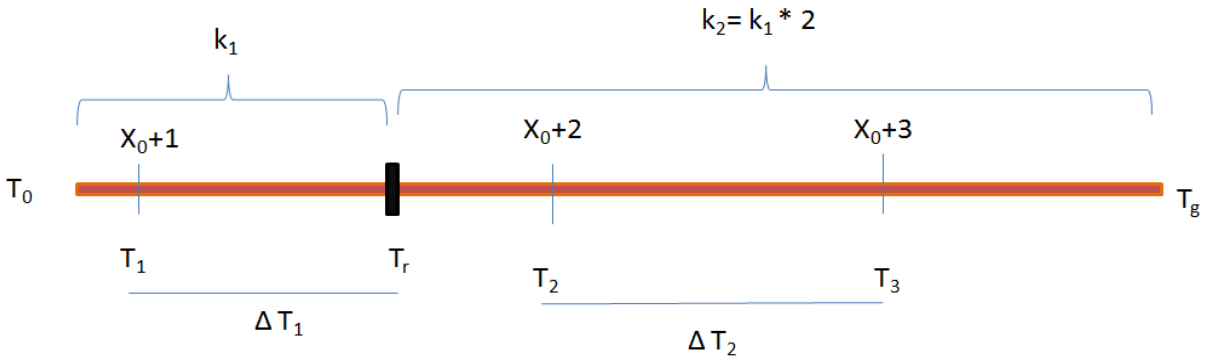


Figure 2.1 Schematic representation of LacI molecule production in *E.coli* Cells.

A cell is born at t_0, new lacI molecules(X) are produced as a random event with the interval of Δt and rate k_1. Production rate is doubled after new *lacI* replication Tr.

Memory less random time points (Δt) can be simulated by Gillespie algorithm [5] which obeys the following equation

$$\Delta t = -\frac{\ln(1-x)}{k}$$

x is probability mass (0> x <1) and k is the rate of production.

2.5 Principles of LacI molecule distribution at cell division

Three different distribution principles is taken into consideration for the simulation

a. Equal distribution: One half goes to one daughter cell and other half goes to another daughter cell (if we consider the division event is absolutely fair).

b. Binomial distribution: Each lacI molecule has 50% chance to any one of the two daughter cells (most publications hypothesize that [10])

c. Distribution following chromosome Hitch-hiking model: The Hitch-hiking model is dependent of the number of binding sites of the particular molecule on chromosome. Equal distribution is followed upto the number of available binds sites and the rest of the molecules distribute binomially. For example, in our case, in a growing *E. coli* cell there are 4 binding sites available on an average per cell (the number is dependent on the number of replication fork). Up to 4 molecules the distribution is equal and if there are more than 4 molecules, the rest of them follow binomial distribution. For *E. coli* cells the number of binding sites (N) in a growing cell can be calculated by the following equation.

$$N = n\,[\ln 2.\,(\tfrac{Tg}{C})]\,[2^{(C+D)/Tg} - 2^{D/Tg}]$$

Here, n=Number of binding sites per chromosome, Tg=generation time, C=C (replication time of total chromosome) period in *E. coli. D*=D (after replication to cell division) period in *E coli*.

2.6 Representation of LacI molecule distribution variation between two daughter cells.

In our simulation we represent the LacI molecule distribution variation between two daughter cells with a variable (V). We have V values for our experimental data with two different conditions (cells are growth with IPTG or without IPTG). We figured out, where is our experimental data seat on different simulation curves to understand the statistical significance of our data. The V value is calculated by the following equation

$$V = \frac{\sum_{i=1}^{n} \left[\frac{(D1-D2)^2}{M}\right]_i}{n}$$

Here, D1 = number of molecules in one daughter. D2= number of molecules in other daughter cell. M= number of molecules in mother cell what is distributed between two daughter cells. n=total Number of mother cells.

3. Results

3.1 Selection of cells of interest

We have experimental data for about 500 cells, both in IPTG and Without IPTG condition. We have calculated the lambda value for each cell. We have selected cells of our interest (just divided) depending on the lambda value (all cells having lambda above 13.5). We got 103 cells in withoutITG condition and 105 in withIPTG condition.

3.2 Attributes of the selected cells

Attributes of the cells, necessary for the simulation have been calculated. The result is summarized in the following table.

Average number of molecule production per cell per generation	Rate of molecule production(K)	Distribution variation(V)	Condition
2.4190	Before *lacI* replication(k1)= 0.0587	**V=0.3376**	**+IPTG**
	After t *lacI* replication(k2)= 0.1174		
1.7864	Before *lacI* replication(k1)= 0.0434	**0.3135**	**-IPTG**
	After t *lacI* replication(k2)= 0.0867		

Table 3.1: calculated parameters for simulation

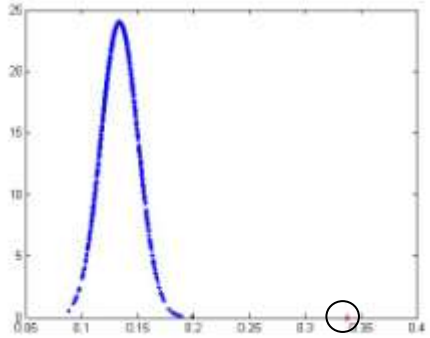
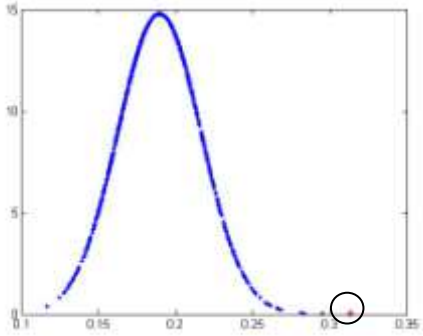
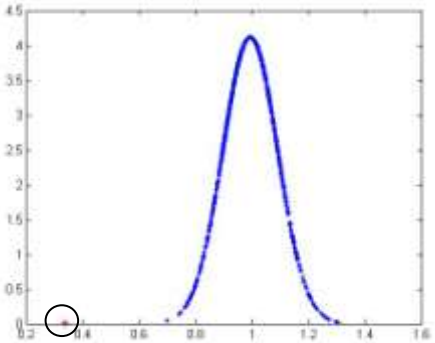
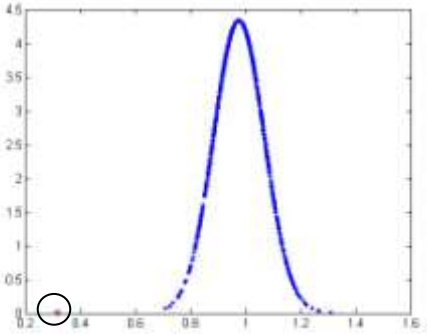
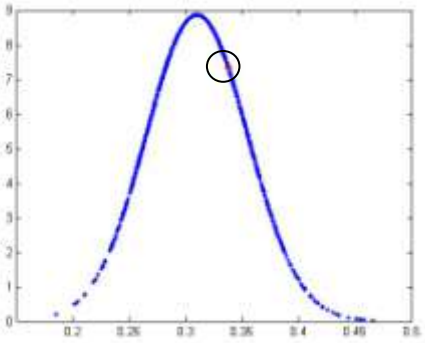
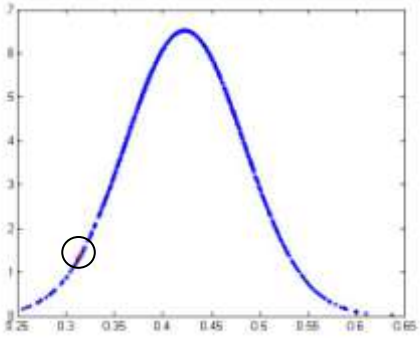
With IPTG condition	Without IPTG condition	Principle
 CDF=1(100%)	 CDF=1(100%)	Equal Distribution
 CDF= 0(0%)	 CDF= 0(0%)	Binomial Distribution
 CDF= 0.0587 (58%)	 CDF= 0.0013(1.3%)	Hitch-hiking Model

Figure 3.1: Graphical representation of statistical significance of experimental data compared to the simulation.

Here, the curves represent the distribution variation (V) of 1000 simulated experiment (equivalent of 100 hours real experiment for each simulated experiment, total=1000x100 hours).The values are normally distributed (blue line) and the red spot represent the placement of experimental value on the distribution curve.

4. Discussion

Stochastic is everywhere in biological systems. Biochemical reactions that involve small numbers of molecules destined to be noisy. Noise plays roles in various biological functions, including introduction of errors in replication process leading to mutation and evolution, noise drives discrepancy of cell-fate, noise can induce amplification or reduction of cellular signaling process, control a switching point in a pathway and produce probabilistic pathway selection. Principally, it helps to generate the qualitative individuality of cells. But if it crosses the threshold, it becomes detrimental. Looking at the coherence of a biological system it is expected that control of noise is under evolutionary selective pressure. For example, experiments suggest that suggests that inefficient translation is over come by frequent transcription [11]. Recent publications propose a new kind of noise that is molecule partitioning noise at cell division [4]. With some preliminary experimental data and computer simulation we tried to decipher the pattern and the extent of this noise.

In our experiment we had some limitation in terms of identifying single molecules with very high precision. In our experimental data set we allowed 10% inaccuracy. That is why; of experimental data hold more randomness than the true distribution. We had additional problem with the data generated in without IPTG condition. Then we have no IPTG in the medium LacI molecule are bound to the DNA. The binding distance is smaller than the resolving power of our optical system which actually introduced two types of error in our data. We underestimated the number of molecule in cells; fluctuation got more weight interns of variation. On the other hand, they looked more equality distributed because we cannot distinguish between one and two molecules. These suppositions make scene if we look at our simulation. The data without IPTG are exceptionally equality distributed. The data with IPTG also looks like they also have the tendency to distribute evenly. The plausible explanation could be; there is some non specific DNA binding occurs even if their binding site is blocked by IPTG or some other mechanism, unpredictable with our present knowledge. In future, we need to improve our magnification capability of our optical system. Our sample size should to bigger to come up with a strong conclusion.

The overall impression that we get from the simulation is the cell must has some mechanisms which help them to distribute the molecule more evenly than the random way. It is highly likely that the chromosome hitch-hiking could the reason or one of the many reasons.

5. Conclusion

The presence of stochasticity in gene expression and other regulatory processes have been confirmed in many publications. The possibility to have some experimental noise is the data is not very low. May be we need to repeat some experiment with better precision. Another fascinating thing has been published recently, which claims that bacteria can transfer molecules in a closed proximity from a very small membrane nano-tube serving as a technique for intra-cellular molecule exchange [12]. If it is so, we need to redefine the cellular individuality and these single cell variations.

Acknowledgements

A big thank to my supervisors, Johan Elf and Andreas Grönlund, for their overall support and constant guidance to complete this project. I am also grateful to Gustaf Ullman, Mats Wallden ,Anel Mahmutovic, Fredrik Persson and David Fange for their valuable comments and suggestions.

References

1. Alberts B. *Molecular Biology of the Cell*. New York: Garland Science, 2002. Print.

2. Kim KH, Sauro HM. **Measuring retroactivity from noise in gene regulatory networks**. *Biophys J.* **100**(5):1167-77 (2011).

3. Fano U. **Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions**. *Phys. Rev*. **72**(1):26–29 (1947).

4. Huh D, Paulsson J. **Non-genetic heterogeneity from stochastic partitioning at cell division**.

Nat. Genet. **43**(2):95-100 (2011).

5. Gillespie DT. **A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions**. *J Comput Phy*. **22**(4): 403-434(1976).

6. MicrobeTracker Suite. Available at http://emonet.biology.yale.edu/microbetracker/

7. Müller-Hill B. *The Lac Operon: a Short History of a Genetic Paradigm*. Berlin: Walter De Gruyter, 1996. Print.

8. Elf J, Li GW, Xie XS. **Probing transcription factor dynamics at the single-molecule level in a living cell**. *Science*. **316**(5828):1191-4 (2007).

9. Michelsen O, Teixeira de Mattos MJ, Jensen PR and Hansen FG. **Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r**. *Microbiol*.**149**:1001-1010 (2003).

10. Berg OG. **A model for the statistical fluctuations of protein numbers in a microbial population**. *J Theor Biol.* 71(4):587-603 (1978).

11. Raser JM and O'Shea EK. Noise **in Gene Expression: Origins, Consequences, and Control**. *Science .* **309**(5743):2010-2013 (2005).

12. Dubey GP and Ben-Yehuda S. **Intercellular Nanotubes Mediate Bacteria Communication**. *Cell* **144**(4):590-600 (2011).