

STATISTICAL MECHANICS OF GENE EXPRESSION

1. E. coli as a chemical reactor



r : rate of transcription (~ 0.3 mRNAs per minute in E. coli)

γ : rate of degradation **per molecule** (~ 0.3 per minute per molecule in E. coli)

mRNA birth: $\Delta n_b = r \Delta t$: average increase in the # of mRNAs in time Δt

mRNA death: $\Delta n_d = \gamma n \Delta t$: average decrease in the # of mRNA molecules

↓ Proof by example:

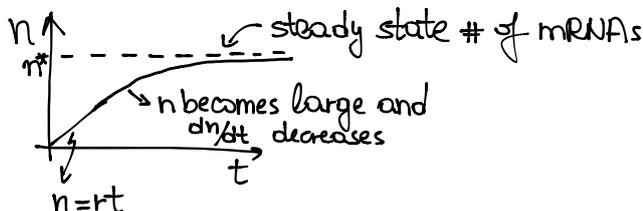


$\gamma \Delta t$ is the probability that any one of the 8 molecules decay in time Δt ; say $\gamma \Delta t = 1/2$.

Then $1/2$ of the molecules will decay, i.e. $4 = 8 \cdot 1/2$ of them!

$$\Delta n = r \Delta t - \gamma n \Delta t$$

$$\frac{\Delta n}{\Delta t} = r - \gamma n \quad \text{rate equation}$$



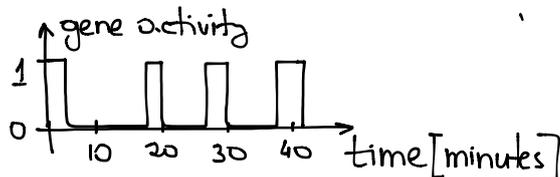
In steady state $\frac{dn}{dt} = 0 \Rightarrow r - \gamma n^* = 0$

$$n^* = r/\gamma = r\tau \quad \text{where } \tau = 1/\gamma \text{ is the mRNA life-time}$$

(in E. coli for a typical mRNA $n^* \sim 1$ but there's considerable variation)

2. Transcriptional regulation

Golding 2005 data suggests:



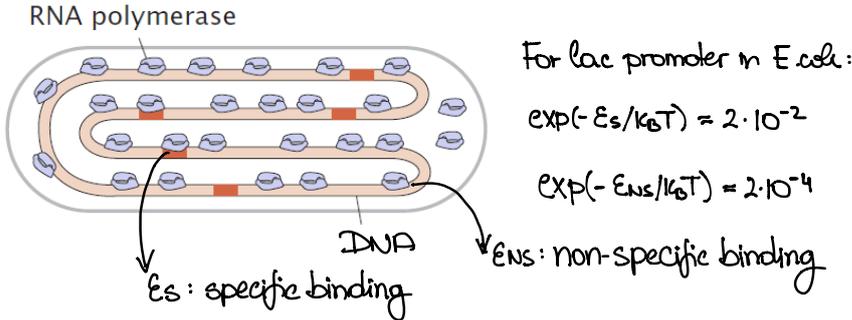
p_{active} : fraction of the time that the gene is actively transcribed

$$\Delta n_b = r \Delta t p_{\text{active}} \Rightarrow n^* = \frac{r}{\gamma} p_{\text{active}}$$

↳ statistical mechanics provides a framework for computing this quantity.

We assume that p_{active} is the probability of RNAP bound to promoter DNA, which we compute using statistical mechanics. There is an implicit assumption that RNAP binding to DNA is in equilibrium.

3. Probability of RNAP bound (active)



STATE	WEIGHT	TRANSCRIPTION
	1	inactive
	$\frac{P}{N_{ns}} e^{-\Delta E/k_B T}$	active

$P_{active} = \frac{P/N_{ns} e^{-\Delta E/k_B T}}{1 + P/N_{ns} e^{-\Delta E/k_B T}}$ This formula makes predictions how changing the # of RNAPs or the sequence of the promoter changes the steady state # of mRNA transcripts of the gene.

$n^* = \frac{r}{\gamma} P_{active}$ is the steady state # of mRNAs in E. coli

For lac genes: $P_{active} = \frac{3 \cdot 10^3 / 5 \cdot 10^6 \cdot 10^2}{1 + 3 \cdot 10^3 / 5 \cdot 10^6 \cdot 10^2} = 5 \cdot 10^{-2}$ (need activator??)

4. Statistical mechanics of repression

STATE	WEIGHT	TRANSCRIPTION
	1	inactive
	$\frac{P}{N_{ns}} e^{-\frac{\Delta E_P}{k_B T}}$	active
	$\frac{R}{N_{ns}} e^{-\frac{\Delta E_R}{k_B T}}$	inactive

$P_{active} = r \frac{P/N_{ns} e^{-\frac{\Delta E_P}{k_B T}}}{1 + P/N_{ns} e^{-\frac{\Delta E_P}{k_B T}} + R/N_{ns} e^{-\frac{\Delta E_R}{k_B T}}}$

Experiments often report Repression = $\frac{\text{Amount of gene expression (R=0)}}{\text{Amount of gene expression (R)}} > 1$ (e.g. Ochler '94 data)

$$\begin{aligned} \text{Repression} &= \frac{P_{active}(R=0)}{P_{active}(R)} \\ &= \frac{\frac{P/N_{ns} e^{-\Delta E_P/k_B T}}{1 + P/N_{ns} e^{-\Delta E_P/k_B T}}}{\frac{P/N_{ns} e^{-\Delta E_P/k_B T}}{1 + P/N_{ns} e^{-\Delta E_P/k_B T} + R/N_{ns} e^{-\Delta E_R/k_B T}}} \end{aligned}$$

$$= 1 + \frac{R/N_{NS} e^{-\Delta E_R/k_B T}}{1 + P/N_{NS} e^{-\Delta E_P/k_B T}}$$

typically this is $\ll 1$ as shown above

$$\text{Repression} = 1 + \frac{R}{N_{NS}} e^{-\Delta E_R/k_B T}$$

This calculation does not take into account that lac repressor is a tetramer and can bind DNA @ two sites



This has the effect of there effectively 2x as many repressors. Each dimer is capable of repressing independently. Therefore in experiments with tetramers:

$$\text{Repression} = 1 + \frac{2R}{N_{NS}} e^{-\Delta E_R/k_B T} \quad \text{Lac repressor tetramers}$$

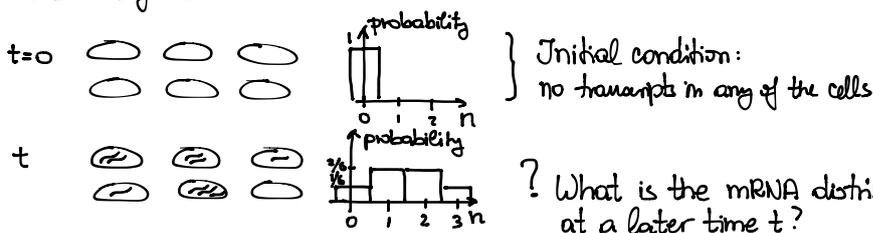
STOCHASTIC TRANSCRIPTION

Unregulated (constitutive) promoter: \xrightarrow{r} mRNA $\xrightarrow{\gamma}$ ϕ

Two meanings of the rate r : 1. $r \cdot t$ = average # of transcripts synthesized in time t

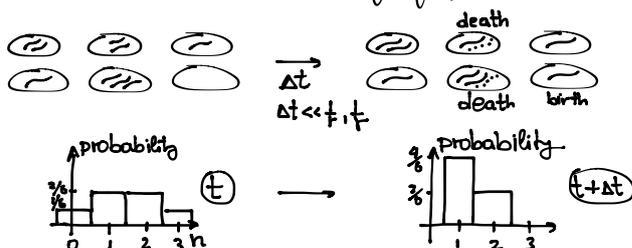
2. $r \cdot \Delta t$ ($r \cdot \Delta t \ll 1$) = probability of a transcript synthesized in the time interval $(t, t + \Delta t)$. Note that the probability does not depend on t (the production process is memoryless).

Given that transcription is off of a constitutive promoter we would like to compute the cell-to-cell distribution of mRNAs.



To solve this problem we make use of the chemical master equation first introduced by Delbruck 1947

The idea is to consider the change of $P(n, t)$ over a very small time interval Δt :



Change in the probability of a cell (i.e. fraction of cells in a genetically identical population) having n mRNA molecules:

$p(n,t+\Delta t) - p(n,t) = \text{birth} - \text{death}$; here "birth" (or "death") refer to creation (annihilation) of cells with n mRNAs and NOT (like before) the creation and annihilation of mRNAs

$$p(n,t+\Delta t) - p(n,t) = \underbrace{p(n-1,t) r \Delta t}_{\text{probability that a cell has } n-1 \text{ mRNAs and gains 1 mRNA in time } \Delta t} + \underbrace{p(n+1,t) \gamma \Delta t}_{\text{probability that a cell has } n+1 \text{ mRNAs and then one of them decays}} - \underbrace{(p(n,t) r \Delta t + p(n,t) \gamma n \Delta t)}_{\text{probability that a cell has } n \text{ mRNAs and either an additional one is made or one decays. Either way it will no longer have } n \text{ mRNAs!}}$$

$$\frac{dp(n,t)}{dt} = r p(n-1,t) + (n+1)\gamma p(n+1,t) - (r + n\gamma) p(n,t)$$

Chemical master equation for unregulated transcription

If we wait long enough the distribution of mRNAs will reach a steady state. This is what is typically measured. In steady state:

$$\frac{dp(n,t)}{dt} = 0 \Rightarrow r p^*(n-1) + (n+1)\gamma p^*(n+1) - (r + n\gamma) p^*(n) = 0; p^*(n) \text{ is the steady state distribution}$$

$$[r p^*(n-1) - n\gamma p^*(n)] - [r p^*(n) - (n+1)\gamma p^*(n+1)] = 0$$

↓ setting these to 0 is a solution (and the only solution)

$$\Rightarrow r p^*(n-1) = n\gamma p^*(n) \rightarrow \text{"detailed balance"}$$

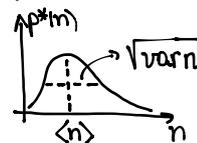
$$p^*(n) = \frac{r}{\gamma} \frac{1}{n} p^*(n-1) = \frac{r}{\gamma} \frac{1}{n} \left(\frac{r}{\gamma} \frac{1}{n-1} p^*(n-2) \right) = \frac{r}{\gamma} \frac{1}{n} \frac{r}{\gamma} \frac{1}{n-1} \left(\frac{r}{\gamma} \frac{1}{n-2} p^*(n-3) \right) = \dots = \left(\frac{r}{\gamma} \right)^n \frac{1}{n!} p^*(0)$$

We can compute the missing $p^*(0)$ using the normalization condition: $p^*(0) + p^*(1) + \dots + p^*(n) + \dots = 1$

$$\sum_{n=0}^{\infty} p^*(n) = 1 \Rightarrow \underbrace{\left(\sum_{n=0}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{n!} \right)}_{e^{r/\gamma}} p^*(0) = 1$$

$$\Rightarrow e^{r/\gamma} p^*(0) = 1 \text{ and therefore } p^*(0) = e^{-r/\gamma}$$

Finally: $p^*(n) = \left(\frac{r}{\gamma} \right)^n \frac{1}{n!} e^{-r/\gamma}$ Poisson distribution



Using this distribution we can compute the mean and variance of the mRNA

Mean of n :

$$\begin{aligned}
 \langle n \rangle &= \sum_{n=0}^{\infty} p^*(n) n \\
 &= \sum_{n=0}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{n!} e^{-r/\gamma} n \\
 &= \sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{(n-1)!} e^{-r/\gamma} \\
 &= \frac{r}{\gamma} \sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^{n-1} \frac{1}{(n-1)!} e^{-r/\gamma} = \frac{r}{\gamma} !
 \end{aligned}$$

So the mean # of mRNAs in a cell is r/γ , which is what we computed previously using the rate equation approach.

Variance of n :

Its also useful to compute the variance of n , which tells us about the width of the distribution

$$\begin{aligned}
 \text{var } n &= \langle (n - \langle n \rangle)^2 \rangle = \langle n^2 - 2n\langle n \rangle + \langle n \rangle^2 \rangle \quad ; \quad \langle n \rangle = \frac{r}{\gamma} \\
 &= \langle n^2 \rangle - \langle 2n\frac{r}{\gamma} \rangle + \left(\frac{r}{\gamma} \right)^2 \\
 &= \langle n^2 \rangle - 2\frac{r}{\gamma} \frac{r}{\gamma} + \left(\frac{r}{\gamma} \right)^2 \\
 &= \langle n^2 \rangle - \left(\frac{r}{\gamma} \right)^2 \quad (\text{In general } \text{var } n = \langle n^2 \rangle - \langle n \rangle^2!)
 \end{aligned}$$

$$\begin{aligned}
 \langle n^2 \rangle &= \sum_{n=0}^{\infty} n^2 p^*(n) \\
 &= \sum_{n=0}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{n!} e^{-r/\gamma} n^2 \\
 &= \sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{(n-1)!} e^{-r/\gamma} \frac{n}{(n-1) + 1} \\
 &= \sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{(n-1)!} \frac{e^{-r/\gamma}}{(n-2)!} (n-1) + \sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^n \frac{1}{(n-1)!} e^{-r/\gamma} \\
 &= \left(\frac{r}{\gamma} \right)^2 \underbrace{\sum_{n=2}^{\infty} \left(\frac{r}{\gamma} \right)^{n-2} \frac{1}{(n-2)!} e^{-r/\gamma}}_{=1} + \frac{r}{\gamma} \underbrace{\sum_{n=1}^{\infty} \left(\frac{r}{\gamma} \right)^{n-1} \frac{1}{(n-1)!} e^{-r/\gamma}}_{=1} \\
 &= \left(\frac{r}{\gamma} \right)^2 + \frac{r}{\gamma}
 \end{aligned}$$

Finally: $\langle \text{var } n \rangle = \langle n^2 \rangle - \langle n \rangle^2$

$$= \left(\frac{r}{\gamma} \right)^2 + \frac{r}{\gamma} - \left(\frac{r}{\gamma} \right)^2$$

$$= \frac{r}{\gamma} = \langle n \rangle \quad \text{For a Poisson distribution the variance is equal to the mean!}$$