Gene expression and its regulation

Wednesday, November 15, 2017
BME/CHE/PHY 558, Physical & Quantitative Biology
Rutgers University: Chemical Thermodynamics
Lecturer: Gábor Balázsi
The road from genotype to phenotype

Gene expression =

The multi-step process leading to protein synthesis based on genetic sequence
Challenge: What gives rise to function? Complete the lower half of the picture!

Sequence → Structure → ??? → Function
The more complete picture

Sequence \rightarrow Structure

Gene regulation \rightarrow Protein amount

Function

- Sequence
- Structure
- Gene regulation
- Protein amount
- Function
Protein amount and sequence affect the properties of cells

- The amounts and sequences of proteins determine the properties of cells (and organisms):
  - What the cell takes up and pumps out
  - The cell’s shape, ability to move, division rate
  - This is why we have liver, skin, blood cells, etc.
Harmful change: protein sequence

**Normal sequence:** HBB gene

...ATA GGA CTT CTT...
...UAU CCU GAA GAA...
...thr pro glu glu...

**Sequence change = mutation:** HBB gene

...ATA GGA CAT CTT...
...UAU CCU GUA GAA...
...thr pro **val** glu...

Normal red blood cell

Sickle cell anemia red blood cell
Harmful change: protein amount

Normal protein levels:

Normal breast tissue: ER protein

Normal breast tissue: PTEN protein

Abnormal protein levels:

Cancerous breast tissue: too much ER protein

Cancerous breast tissue: too little PTEN protein
The biology of gene expression
Genes code for proteins: The Central Dogma of Molecular Biology

- DNA = double chain: A, C, T, G
- RNA = single chain: A, C, U, G
- Proteins, made of 20 amino acids
From DNA to protein: sequence and amount

DNA is copied into RNA molecules

RNA is copied into protein molecules

RNA degrades

Protein degrades
Biomachines make mRNA and proteins

RNA polymerases make mRNA-s from DNA.

Ribosomes make proteins from mRNA.
Transcription by RNA polymerase

\[ \text{DNA} \rightarrow \text{mRNA} \]
Translation by ribosome

Protein ←

RIBOSOME

mRNA ←
Transcription across the domains of Life
Transcription in bacteria

Bacterial genes are typically ON (available for transcription) by default.

[Diagram showing a section of DNA with promoter regions labeled as TTGACG and TATAAT, a sigma factor, and a transcription start site marked by +1. Below the diagram, there is a YouTube video link: https://youtu.be/1b-bRVgqof0]
Translation in bacteria

Translation is coupled to transcription in bacteria.
Eukaryotic transcription

Eukaryotic genes are typically OFF (not available for transcription) by default.
Eukaryotic translation

Transcription (in nucleus) and translation (in cytoplasm) are uncoupled in eukaryotes.
Gene expression in Bacteria vs. Eukaryotes

**BACTERIA**
- DNA
- mRNA
- Growing amino acid chain

**EUKARYOTES**
- DNA
- Nucleus
- Transcription and processing
- Cytoplasm
- mRNA
- Transport

5' → 3'
Gene expression cartoon

Laa lalaaa

Yeah! I know it is weird. But this is how they expressed themselves.

http://biocomicals.blogspot.com
Simplest model: Synthesis + Degradation

The simplest models of gene expression: synthesis & degradation of protein \((P)\) from mRNA \((M)\).

\[
\begin{align*}
\text{protein synthesis} & \quad k = 1 \\
M & \xrightarrow{k} P + M \\
\text{protein degradation} & \quad g = 0.1 \\
P & \xrightarrow{g} \emptyset
\end{align*}
\]

ODE model:

\[
\frac{dP}{dt} = k - gP
\]

Solution:

\[
P(t) = \frac{k}{g} \left(1 - e^{-gt}\right)
\]

Steady state:

\[
P_\infty = \frac{k}{g} = \frac{\text{synthesis}}{\text{degradation}}
\]

Steady state: \(P_\infty = 10\)
Degradation = first-order decay

$P \xrightarrow{g} \emptyset$

How does the amount of $A$ change over time?

\[
\frac{d[P]}{dt} = -g[P]
\]

$[P] =$ amount of $P$ (in molecules or moles)

$g =$ rate coefficient = 

= probability of $P$ degrading per unit time

$P(t) = P_0 e^{-gt}$

Exponential decay

\[
[0, 100] \quad [0, 10]
\]

$g = 1$
The coin elimination game

(1) Take 100 coins.
(2) Toss all. Remove Heads.
(3) Repeat until no coin is left.

\[ N(t = 0) = 100 \]

\[ \langle N(t = 1) \rangle = \frac{100}{2} = 50 \]

\[ \langle N(t = n) \rangle = \frac{100}{2^n} = 100e^{-t \ln(2)} \]

There can be fluctuations around the average.

\[ P[N = 99] = \frac{100!}{99!1!} p^{99} (1 - p) = \frac{100}{2^{100}} \approx 7.9 \times 10^{-29} \]

\[ P[N = 52] = \frac{100!}{52!48!} p^{52} (1 - p)^{48} \approx \frac{10^{29}}{2^{100}} \approx 0.0735 \]
Stochastic chemical kinetics: Degradation

\[ A \xrightarrow{g} \emptyset \quad A_0 = 100 \]

Ord. diff. equation: average of [A]

\[
\frac{d[A]}{dt} = -g[A]
\]

\[ A(t) = A_0 e^{-gt} = 100 e^{-t} \]

Master Equation: probability of \( N_A \) at \( t \)

\[
\frac{dP(N_A)}{dt} = \frac{g(N_A + 1)P(N_A + 1, t) - gN_A P(N_A, t)}{dt}
\]
Solving the Master Equation: Degradation

\[ A \xrightarrow{g} \emptyset \quad A_0 = 100 \]

What is the probability of having \( N_A \) at time \( t \)?

\[
\frac{dP(A)}{dt} = g(A + 1)P(A + 1, t) - gAP(A, t) \]

For simplicity, we denote: \( N_A = A \)

For 1 particle, the probability it has not decayed by time \( t \): 

\[ P(\text{intact}) = e^{-gt} \]

For all particles:

\[ P(A; t) = \binom{A_0}{A} e^{-Agt} [1 - e^{-gt}]^{A_0-A} \]

Monte-Carlo (stochastic) simulations can also estimate \( P(A; t) \)
Cells are microscopic reaction chambers

Each cell carries molecules that react. Some molecules are in low numbers. Cells are biological dice.

*E. coli* bacteria
Stochastic protein synthesis + degradation

Synthesis, \( k=1 \) \( \emptyset \xrightarrow{k} P \)

Degradation, \( g=0.1 \) \( P \xrightarrow{g} \emptyset \)

ODE model:
\[
\frac{dP}{dt} = k - gP
\]

solution:
\[
P(t) = \frac{k}{g} \left(1 - e^{-gt}\right)
\]

\( P_n \) = Probability of \( N_A=n \) at time \( t \)

\( k_n = k \)

\( g_n = gn \)

Master equation:
\[
\frac{dP_n}{dt} = -(k_n + g_n)P_n + k_{n-1}P_{n-1} + g_{n+1}P_{n+1}
\]
The Fokker-Planck Equation

The goal is to solve:

$$\frac{dP_n}{dt} = -(k_n + g_n)P_n + k_{n-1}P_{n-1} + g_{n+1}P_{n+1}$$

Taylor expansion of two functions:

$$k_{n-1}P_{n-1} \approx [k_nP_n - \frac{\partial(k_nP_n)}{\partial n} + \frac{1}{2} \frac{\partial^2 (k_nP_n)}{\partial n^2}]$$

$$g_{n+1}P_{n+1} \approx [k_nP_n + \frac{\partial(g_nP_n)}{\partial n} + \frac{1}{2} \frac{\partial^2 (g_nP_n)}{\partial n^2}]$$

$$\frac{dP_n}{dt} \approx -\frac{\partial(k_nP_n)}{\partial n} + \frac{1}{2} \frac{\partial^2 (k_nP_n)}{\partial n^2} + \frac{\partial(g_nP_n)}{\partial n} + \frac{1}{2} \frac{\partial^2 (g_nP_n)}{\partial n^2}$$

$$\frac{dP(n)}{dt} \approx -\frac{\partial}{\partial n} \left\{ (k - gn)P - \frac{1}{2} \frac{\partial}{\partial n}[(k + gn)P] \right\}$$

Fokker-Planck Equation: Diffusion = Random Walk in “concentration” space.
Biochemical forces: synthesis, degradation

\[ \varphi(n) = -\frac{2}{\beta} \int \frac{k - gn}{k + gn} \, dn' \]

\[ F(n) = -\frac{\partial \varphi(n)}{\partial n} = \frac{2}{\beta} \frac{k - gn}{k + gn} \]

\[ F(n) = F_s - F_D = \frac{2}{\beta} \frac{k}{k + gn} - \frac{2}{\beta} \frac{gn}{k + gn} \]

At equilibrium:

\[ F(n) = 0 \Rightarrow k = gn \]
Stochastic gene expression in yeast cells
Gillespie simulations and the Coefficient of Variation (CV)

Very often it is difficult or impossible to solve the Master Equation.

In that case, we run Monte-Carlo (Gillespie) simulations to predict stochasticity.


Coefficient of Variation:
\[
CV = \frac{\sigma}{\mu}
\]

*Noise (CV = \sigma/\mu)*:
- Quantify deviations from the population mean
Stochasticity depends on synthesis & degradation rates

**Synthesis rate**

\[ k = 0.1 \]
\[ k = 10 \]
\[ k = 1000 \]

\[ \gamma = \ln(2)/6 \]
\[ \gamma = \ln(2)/60 \]
\[ \gamma = \ln(2)/600 \]

**Degradation rate**


\[ \mu = \text{mean}, \ \sigma^2 = \text{variance}, \ CV = \sigma / \mu \]
Lowering the mean causes CV to increase

Synthesis rate

\( k=0.1 \)

\( \gamma = \ln(2)/6 \)

\( \mu = 0.865617 \)

\( \sigma^2 = 0.865617 \)

\( CV = 1.07482338 \)

\( k=10 \)

\( \gamma = \ln(2)/60 \)

\( \mu = 8.65617 \)

\( \sigma^2 = 8.65617 \)

\( CV = 0.339889 \)

\( k=1000 \)

\( \gamma = \ln(2)/600 \)

\( \mu = 86.5617 \)

\( \sigma^2 = 86.5617 \)

\( CV = 0.010748 \)

Degradation rate

\( \mu = \text{mean}, \ \sigma^2 = \text{variance}, \ CV = \sigma / \mu \)

The higher the mean, the lower the noise.

Poisson process:

\[
\mu = \frac{k}{\gamma}, \quad \sigma^2 = \frac{k}{\gamma}
\]

\[
CV = \frac{\sigma}{\mu} = \sqrt{\frac{\gamma}{k}}
\]

Gene expression bursts:
Non-Poissonian time series
Summary

• Genes code for proteins through gene expression
• Two factors define cell & organism properties:
  – Protein sequence
  – Protein amount
• Gene expression steps define protein amounts
• Gene expression is a stochastic process