Gene expression and its regulation

Wednesday, November 11, 2015
SBU:CHE/PHY558, Physical & Quantitative Biology
Lecturer: Gábor Balázsi
Gene expression

Gene expression = The multi-step process leading to protein synthesis based on genetic sequence
Genes code for proteins

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

DNA is copied into RNA

RNA is copied into protein

RNA degrades

Protein degrades
What makes mRNA and proteins?

RNA polymerases make mRNA-s. Ribosomes make proteins.
Protein amount and sequence affect the properties of cells

- The amounts and sequences of proteins determine the properties of cells (and thereby 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 changes: sequence and amount

1: **Sequence change = mutation**: HBB gene

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

DNA  mRNA  protein

Normal red blood cell

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

DNA  mRNA  protein

Sickle cell anemia red blood cell

2: **Change in amount**: Over- or under production of protein

Breast tissue:
ER protein

<table>
<thead>
<tr>
<th>Normal</th>
<th>Cancer</th>
</tr>
</thead>
</table>

Breast tissue:
PTEN protein

<table>
<thead>
<tr>
<th>Normal</th>
<th>Cancer</th>
</tr>
</thead>
</table>
Regulation of gene expression (Gene regulation)

- Gene expression can increase or decrease. How?
- What is on the DNA besides genes? Promoter regions.
  - Some proteins can bind to promoter regions.
  - Thereby certain proteins can help or inhibit the synthesis of other proteins (gene activity).

Genes don’t work alone. They form gene networks. They tell to each other to express more/less protein!
Measuring gene expression

Intracellular proteins are invisible. How to measure their quantity?

- The GFP protein shines green.
- Green cells contain more of the original protein as well.

Cells
Prokaryotic (bacterial) transcription

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

Transcription and translation are coupled in prokaryotes (bacteria).

https://youtu.be/1b-bRVgqof0
Prokaryotic translation
Eukaryotic transcription

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

https://youtu.be/JOBwqwxgLqc
Transcription and translation are uncoupled in eukaryotes
Models of gene expression

Models of gene expression focus on: promoter, mRNA, protein.

Promoter:

\[ D \xleftrightarrow{k_{ON}} B \]
\[ D \xrightarrow{k_{OFF}} B \]

\( D = \text{empty DNA} \)
\( B = \text{Polymerase bound to DNA} \)

mRNA:

\[ B \xrightarrow{k_M} M + B \]
\[ M \xrightarrow{g_M} \emptyset \]

\( M = \text{mRNA} \)

Protein:

\[ M \xrightarrow{k_P} P + M \]
\[ P \xrightarrow{g_P} \emptyset \]

\( P = \text{protein} \)
Models of gene expression focus on: promoter, mRNA, protein.

Promoter:

\[ D \xrightleftharpoons[k_{OFF}]^{k_{ON}} B \]

\[ \frac{dD}{dt} = -k_{ON}D + k_{OFF}B \]

\[ D + B = \text{const} = 1 \]

mRNA:

\[ B \xrightarrow[k_{M}]{} M + B \]

\[ \frac{dM}{dt} = k_{M}B - g_{M}M \]

Protein:

\[ M \xrightarrow[k_{P}]{} P + M \]

\[ \frac{dP}{dt} = k_{P}M - g_{P}P \]

\[ M \xrightarrow[g_{M}]{} \emptyset \]

\[ P \xrightarrow[g_{P}]{} \emptyset \]
Gene expression versus promoter state

Promoter:

\[ k_{ON} (1 - B) = k_{OFF} B \quad \Rightarrow \quad B = \frac{k_{ON}}{k_{OFF} + k_{ON}} \]

mRNA:

\[ M = \frac{k_M}{g_M} B \]

Protein:

\[ P = \frac{k_P}{g_P} M \]

\[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} B \]

\[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} \frac{k_{ON}}{k_{OFF} + k_{ON}} \]
Degradation = exponential decay

$A \xrightarrow{k_f} \emptyset$

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

$$-\frac{d[A]}{dt} = k_f [A]$$

$[A] = \text{amount of } A \text{ (in molecules or moles)}$

$k_f = \text{rate coefficient} =$

$= \text{probability of } A \text{ degrading per unit time}$

$$A(t) = A_0 e^{-k_f t}$$

Exponential decay

[k_f = 1]
Example #1: Coin elimination game

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

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

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

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

There can be fluctuations around the average. The chance of having 99 coins after 1 toss:

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

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

Ord. diff. equation: average of \([A]\)

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

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

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

\[ \frac{dP(N_A)}{dt} = \gamma(N_A + 1)P(N_A + 1, t) - \gamma N_A P(N_A, t) \]
Cells are microscopic reaction chambers

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

E. coli bacteria
Example: protein synthesis + degradation

\[ k=1 \quad \text{synthesis} \quad \rightarrow \quad \gamma=0.1 \quad \text{degradation} \]

\[ \emptyset \overset{k}{\rightarrow} P \quad \frac{dP}{dt} = k \]

\[ P \overset{\gamma}{\rightarrow} \emptyset \quad \frac{dP}{dt} = -\gamma P \]

ODE model:

\[ \frac{dP}{dt} = k - \gamma P \]

solution:

\[ P(t) = \frac{k}{\gamma} \left(1 - e^{-\gamma t}\right) \]
Stochastic gene expression in yeast cells
Quantifying cell-cell variability: the Coefficient of Variation (CV)

Coefficient of Variation:
CV = σ/μ

(1) Mean (μ) :
-Population average (arithmetic mean over many cells)

(2) Variance (σ²), standard deviation (σ), noise (CV = σ/μ) :
-Quantify deviations from the population mean
Stochasticity depends on synthesis & degradation rates

Lowering the mean causes CV to increase

<table>
<thead>
<tr>
<th>Synthesis rate</th>
<th>$k=0.1$</th>
<th>$k=10$</th>
<th>$k=1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma=\ln(2)/6$</td>
<td>$\mu=0.865617$</td>
<td>$\mu=86.5617$</td>
<td>$\mu=8656.17$</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2=0.865617$</td>
<td>$\sigma^2=86.5617$</td>
<td>$\sigma^2=8656.17$</td>
</tr>
<tr>
<td></td>
<td>CV=1.07482338</td>
<td>CV=0.1074823</td>
<td>CV=0.010748</td>
</tr>
<tr>
<td>$\gamma=\ln(2)/60$</td>
<td>$\mu=8.65617$</td>
<td>$\mu=865.617$</td>
<td>$\mu=86561.7$</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2=8.65617$</td>
<td>$\sigma^2=865.617$</td>
<td>$\sigma^2=86561.7$</td>
</tr>
<tr>
<td></td>
<td>CV=0.339889</td>
<td>CV=0.033989</td>
<td>CV=0.003399</td>
</tr>
<tr>
<td>$\gamma=\ln(2)/600$</td>
<td>$\mu=86.5617$</td>
<td>$\mu=8656.17$</td>
<td>$\mu=865617$</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2=86.5617$</td>
<td>$\sigma^2=8656.17$</td>
<td>$\sigma^2=865617$</td>
</tr>
<tr>
<td></td>
<td>CV=1.0748234</td>
<td>CV=0.1074823</td>
<td>CV=0.010748</td>
</tr>
</tbody>
</table>


$\mu=$mean, $\sigma^2=$variance, $CV=\sigma / \mu$
Mean and CV in real living yeast cells

The higher the mean, the lower the noise.

Summary

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