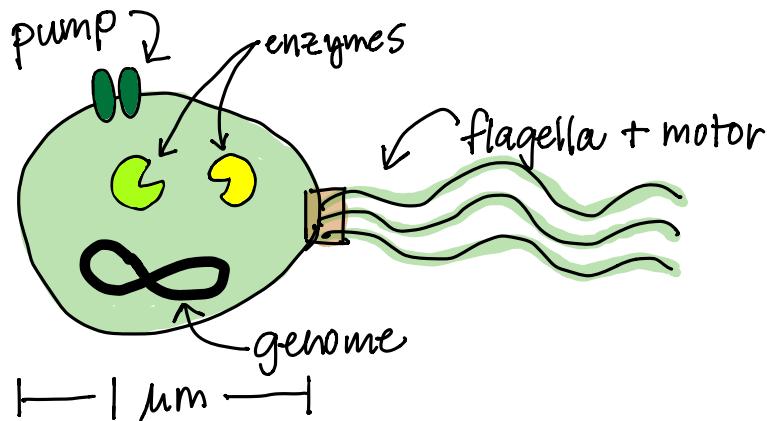


GENE REGULATION & TRANSCRIPTION FACTORS

How do cells process information to allow for survival?

Here is E. coli:



E. coli is constantly monitoring its internal and external environment to tune its composition.

Pumps, enzymes, motors, flagella — these are all proteins.

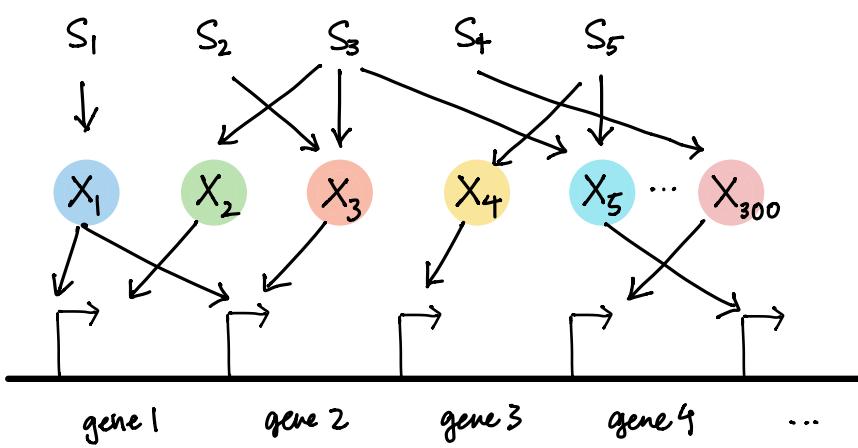
1. How does *E. coli* know which proteins to make?
2. How quickly does it do that?

E. coli senses signals in its environment:

FOR CONTINUED SURVIVAL,
in response to these signals,
E. coli has to make the right "decisions".

- pH
- temperature
- nutrients
- toxins
-

Cells represent their internal and external environments using ~300 special information-carrying proteins called TRANSCRIPTION FACTORS.



SIGNALS e.g., "It is too hot here" or "I am starving"

TRANSCRIPTION FACTORS respond to signals & change the transcription rates of genes.

GENES ... are only expressed when the cell needs them.

How do transcription factors activate gene expression?

TRANSCRIPTION FACTOR X binds to SIGNAL S_x
 $\downarrow (\mu s)$

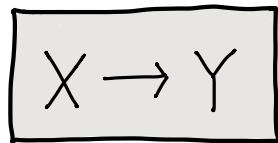
the ACTIVATED FORM of X, X^* , binds the promoter region of GENE Y

$\downarrow (s)$

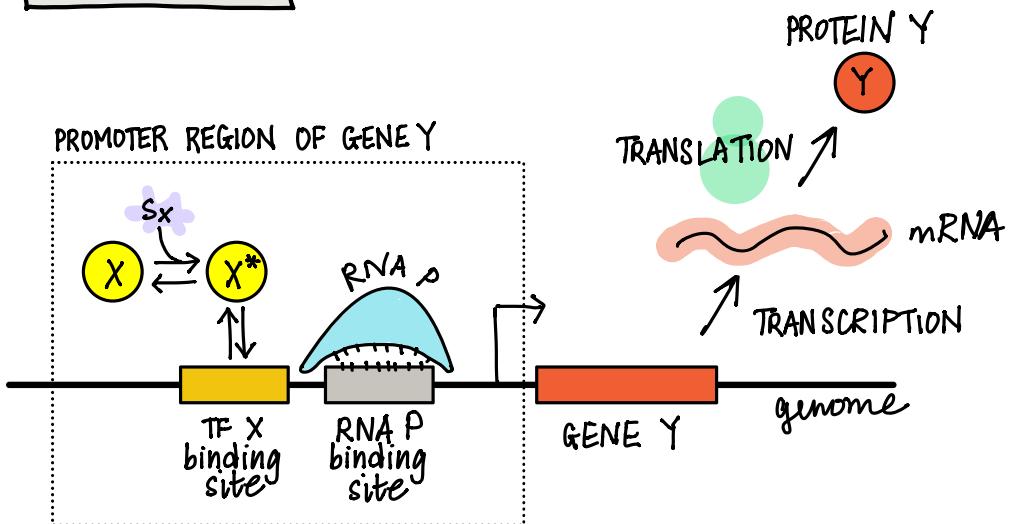
TRANSCRIPTION & TRANSLATION

$\downarrow (min)$

accumulation of PROTEIN Y in the cell.
 $(hours)$



Transcription factor X regulates gene Y.

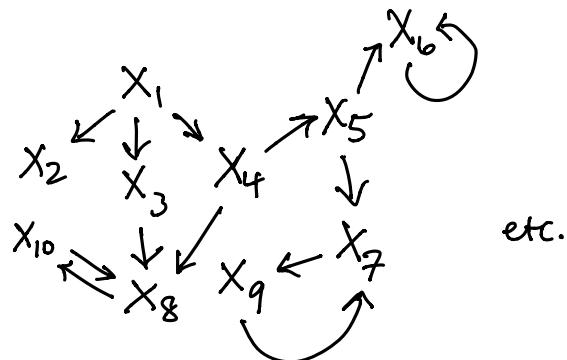


SEPARATION OF TIMESCALES

allows us to assume slow reactions are not changing while fast reactions are changing.

When we collect all the nodes and edges, we get the GENE REGULATION NETWORK.

In E. coli, $N \approx 4500$ genes
 $E \approx 10,000$ arrows.

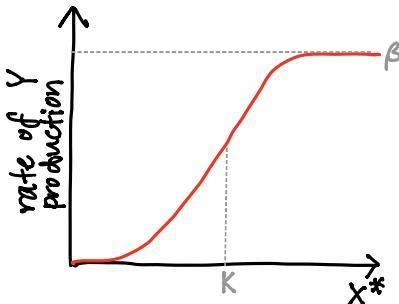


- TFs are the only nodes w/ outgoing arrows. (b/w 1-10³)
- each gene has 0-6 incoming arrows.

To predict gene expression in the GRN, we need to understand the arrows.



Input function : what is the rate of Y production given a certain concentration of X^* , the active form of X that binds DNA?



- This function comes from a chemical binding event with an equilibrium constant, K.
- $K = [X^*]$ at which 50% of X^* is bound.
- When there is infinite X^* and all sites are bound, we have the maximum rate of Y production, β .

The HILL FUNCTION describes binding event behavior.

For an activating TF, the rate of Y production is

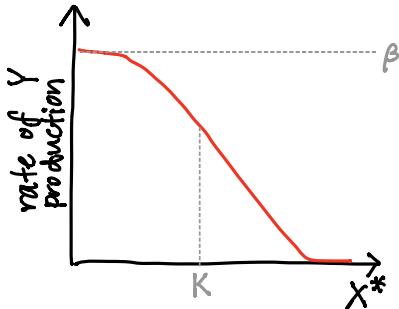
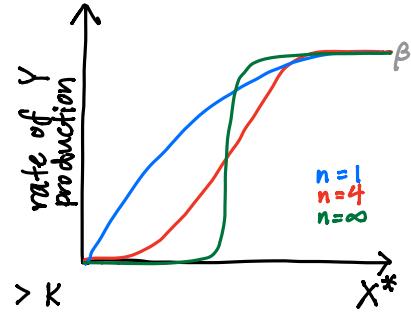
$$\hookrightarrow \text{when } K = X^*, \text{ rate} = \frac{1}{2}\beta.$$

$\hookrightarrow n$ says how steep the curve is.

- $n=1-4$ typically (e.g. multiple binding sites)

- $n=\infty$ looks like a step function : $F(x) = \begin{cases} \beta, & x^* > K \\ 0, & x^* < K \end{cases}$

$$\frac{\beta (X^*)^n}{K^n + (X^*)^n}$$



For a repressing TF, the rate of Y production is

$$\frac{\beta K^n}{(X^*)^n + K^n}$$

What happens when 2+ TFs regulate the same gene?

→ These input functions often look like the product of Hill functions.

→ Complex logic also happens.

→ No systematic way to discover this behavior!!

GRNs are plastic over the slow timescale of generations. Evolution changes and rewrites the GRN to increase an organism's fitness.



dynamics: the equations that guide these networks.

FOR A SINGLE ARROW $X \rightarrow Y$:

$$\frac{dY}{dt} = \beta - \alpha Y \leftarrow \text{concentration of } Y.$$

↑ ↑ ↑
 rate of change of Y production rate removal rate

$$\alpha = \alpha_{deg} + \alpha_{dil}$$

↑ ↑
 proteasomal degradation dilution from cell division

⇒ WHAT IS THE STEADY-STATE LEVEL OF Y? (Y_{st})

Y_{st} is important because it's the concentration of Y that is functional for gene circuits to work.

At steady-state, $\frac{dY}{dt} = 0$.

$$0 = \beta - \alpha Y$$

$$Y_{st} = \frac{\beta}{\alpha} \quad \left\{ \begin{array}{l} \text{ratio between the production} \\ \text{and removal of } Y. \end{array} \right.$$

⇒ HOW LONG DOES IT TAKE TO REACH Y_{st} ?

Let's say we start at Y_{st} , then suddenly, $\beta \rightarrow 0$ because S_x vanishes. The new $Y_{st} = 0$. But how long will it take to get there?

$$\frac{dY}{dt} = -\alpha Y$$

$$\frac{dY}{Y} = -\alpha dt$$

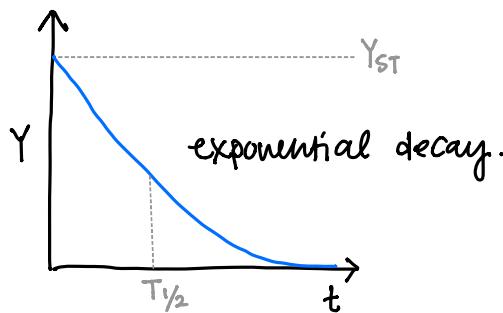
$$\int \frac{dY}{Y} = -\alpha \int dt$$

$$\ln Y + C_1 = -\alpha t + C_2$$

$$Y = Ce^{-\alpha t}$$

with the initial condition of $Y = Y_{ST}$,

$$Y = Y_{ST}e^{-\alpha t}$$



Halfway to the new Y_{ST} value, $Y=0$, is the RESPONSE TIME, $T_{1/2}$.

$$Y(T_{1/2}) = \frac{1}{2} Y_{ST}$$

$$\frac{1}{2} Y_{ST} = Y_{ST} e^{-\alpha(T_{1/2})}$$

$$\frac{1}{2} = e^{-\alpha(T_{1/2})}$$

$$-\ln 2 = -\alpha(T_{1/2})$$

$$T_{1/2} = \frac{\ln 2}{\alpha}$$

$T_{1/2}$ depends only on the removal rate, α , not the production rate, β .

- * If we try the opposite scenario where we begin with $\beta=0$ and introduce a new signal Sx , so that $\beta \rightarrow \beta$, we get the SAME RESPONSE RATE.
- * So: it doesn't matter if you're making or removing the protein in response to the signal. $T_{1/2}$ is always determined by α .

→ HOW LONG IS $\ln 2 / \alpha$?

Interestingly, for most proteins in E.coli (and also human cells), it is really TOO SLOW!! Most proteins are stable with $\alpha_{deg} = 0$ on the timescale of a cell generation. So α is mostly α_{dil} .

Imagine $\beta=0$. The cell doubles, so after one generation, the concentration in each cell is $Y_{ST}/2$. So after one generation, the concentration has gone down by a factor of 2, which is exactly the response time.

Since the direction of the signal doesn't matter, it always takes a full generation to reach Y_{ST} .

Cells function faster than a cell generation — so, how do they solve this problem?

GRNs contain NETWORK MOTIFS — recurring circuits of interactions — that allow cells to respond to signals and undergo development in a robust way and on the correct timescale.

