Systems Thinking

@Soumil Gupta

Introduction to Systems Biology

Basic Biology

- Difference Between Introns and Exons:

Introns	Exons
Introns are non-coding sequence of a gene	Exons are coding sequence of a gene
Introns do not appear in mature mRNA molecules	Exons collectively make the final RNA molecule

- Types of Signaling:

Endocrine	Autocrine	Paracrine
Signals from distant cells that originate from endocrine cells, usually producing a slow response, but having a long-lasting effect	Produced by signaling cells that can also bind to the ligand that is released: the signaling cell and the target cell can be the same or a similar cell (auto = self)	A form of cell signaling in which the target cell is near (para = near) the signal-releasing cell

- Difference Between Translation and Transcription:

Transcription	Translation
Process by which information encoded in DNA is copied onto mRNA	Process by which information encoded in mRNA is used to assemble a protein at a ribosome
Occurs in Nucleus	Occurs in Ribosome

- Difference between Eukaryotic and Prokaryotic Cell:

Prokaryotic	Eukaryotic
Size of the cell is small	Size of the cell is large
Nucleolus is absent	Nucleolus is present
Mitochondria is absent	Mitochondria is present
Contains single chromosome	Contains multiple chromosomes.

> `Q:` How does the cell read the genome?
>
> `A:` The process of reading the genome involves transcription and translation. In the nucleus, a segment of DNA is transcribed into a complementary RNA molecule by the enzyme RNA polymerase. This newly synthesized messenger RNA (mRNA) undergoes processing, including the addition of a protective cap and poly-A tail, and removal of introns. The processed mRNA is then transported to the cytoplasm, where ribosomes read it in codons of three nucleotides.

Transcription Networks

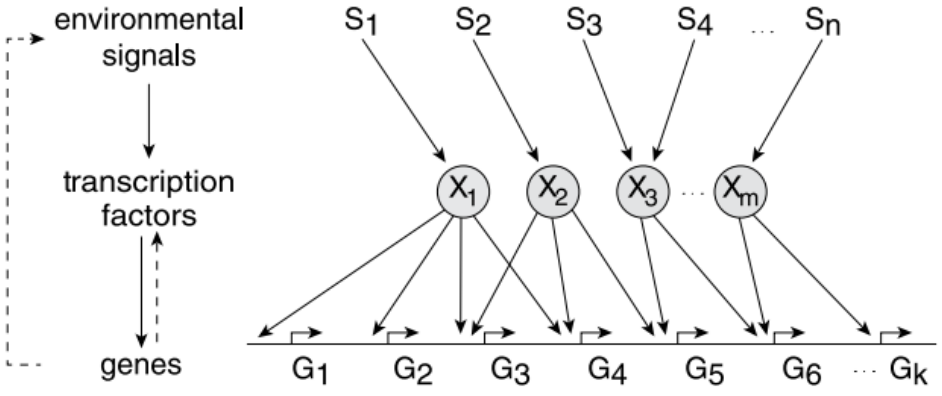
▼ Introduction

- The cell is made of several thousand types of proteins.
- Each protein is a nanometer-size molecule that carries out a specific task with precision.

- `Example:` Escherichia coli is a cell that contains a few million proteins, of about 4500 different types.
- Cells encounter different situations that require different proteins. The cell continuously monitors its environment and calculates the amount at which each type of protein is needed.
- This information-processing function is largely carried out by transcription networks.

The Cognitive Problem of the Cell

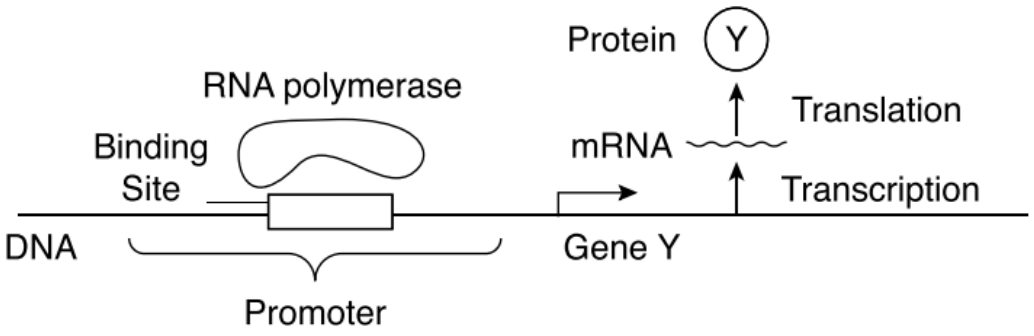
- Cells live in a complex environment and can sense many different signals.
- Cells respond to these signals by producing appropriate proteins that act upon the internal or external environment.
- `Transcription Factor:` To represent these environmental states, the cell uses special proteins called transcription factors as symbols. Transcription factors are designed to transit rapidly between active and inactive molecular states, at a rate that is modulated by a specific environmental signal.
- Each active transcription factor can bind the DNA to regulate the rate at which specific target genes are read. The genes are transcribed into mRNA, which is then translated into protein, which can act on the environment.
- `Note:` We don't need a transcription factor for every situation. Many situations demand the same response, so these can be addressed with a single transcription factor.



- These transcription factors regulate their target genes to mobilize the appropriate protein responses in each case.

▼ Elements of Transcription Networks

- `Gene:` Each gene is a stretch of DNA whose sequence encodes the information needed for production of a protein.
- The protein is produced in two steps, **transcription and translation**.
- `Transcription:` The gene is copied into a disposable mRNA molecule by a protein machine called RNA polymerase ($\mathrm{RNA_p}$).
- The mRNA is then translated into a protein.



- `Promoter:` The rate at which the gene is transcribed i.e., the number of mRNA produced per unit time, is controlled by a regulatory region of DNA that precedes the gene (it is present a few molecules before the Gene itself), called the promoter.
- Whereas RNAp acts on all of the genes, changes in the expression of specific genes are due to transcription factors. Each transcription factor modulates the transcription rate of a set of target genes. Transcription factors affect the transcription rate by binding **specific sites in the promoters** of the regulated gene.
- When bound, they change the probability per unit time that RNAp binds the promoter and produces an mRNA molecule. The transcription factors thus affect the rate at which RNAp initiates transcription of the gene.
- `Note:` Transcription factors can act as activators that increase the transcription rate of a gene, or as repressors that reduce the transcription rate.

- [Skipped Page 6. Read it yourself]
- Each environmental signal is a small molecule, protein modification or molecular partner that directly affects the activity of one of the transcription factors.
- The signals usually cause a physical change in the shape of the transcription factor protein, causing it to assume an active molecular state.

  `Note:` Thus, signal S_{x} can cause X to rapidly shift to its active state X^*, bind the promoter of gene Y and change the rate of transcription, leading to increased or decreased production of protein Y

▼ Activators and Repressors

- Each arrow in a transcription network corresponds to an interaction in which a transcription factor directly controls the transcription rate of a gene. These interactions can be of two types:
- `Activation:` Occurs when the transcription factor increases the rate of transcription when it binds the promoter.
- `Repression:` Occurs when the transcription factor reduces the rate of transcription when it binds the promoter.
- Each arrow in the network has a sign: $+$ for activation, $-$ for repression.
- Plus arrows are denoted by a regular arrow $X \rightarrow Y$, whereas minus arrows are denoted by a blunt-headed arrow $X \dashv Y$
- Each transcription factor acts primarily in one mode for its target genes, as either an activator or a repressor.
- In contrast, the input modes (input to a gene) of regulation are often mixed: a typical gene is activated by some transcription factors and repressed by others.
- Thus, the signs on outgoing arrows (arrows that point out from a given node) are highly correlated, but the signs on incoming arrows (arrows that point into a given node) are not, because we do not know the net effect.
- The strength of the effect of a transcription factor on a target gene is described by an input function.

▼ Logic Input Functions

- Hill input functions are useful for detailed models. For simplicity, it is useful to use even simpler functions that capture the essential behavior of these input functions.

▼ Input Functions

- Typically, the input function $f(X^*)$ is a monotonic function. It is an increasing function when X is an activator and a decreasing function when X is a repressor.
- **Hill function**: A useful function that realistically describes many gene input functions.

▼ For Activator →

- The Hill input function **for an activator** is a curve that rises from zero and approaches a maximal saturated level

$$f(X^*) = \beta \frac{X^{*n}}{K^n + X^{*n}}, \quad \text{Hill function for an activator}$$

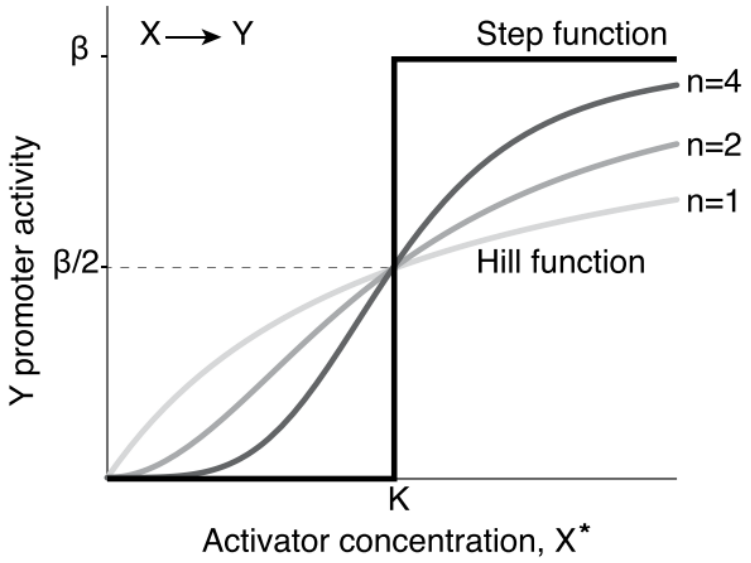
- **Parameters:**

  κ : **The activation coefficient**, and has units of concentration. It defines the concentration of active X needed to significantly activate expression.

  β **Maximal Promoter Activity**, It is reached at high activator concentrations, $X^* \gg K$.

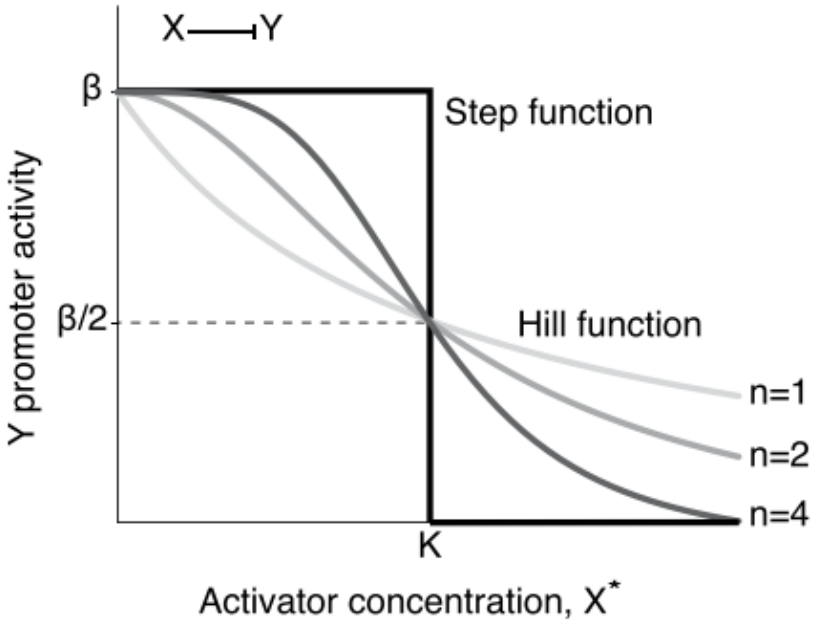
  n : It determines the **steepness** of the input function. The larger n is, the more step-like the input function.

- The Hill function approaches a limiting value at high levels of X*, rather than increasing indefinitely.



▼ For Repressor →

- For a repressor, X ⊣ Y, the Hill input function is a decreasing curve.
- Since a repressor allows strong transcription of a gene only when it is not bound to the promoter, this function can be derived by considering the probability that the promoter is unbound by X^*
- The maximal promoter activity β is obtained when the repressor does not bind the promoter at all that is, when $X^* = 0$



$$f(X^*) = \beta \frac{K^n}{K^n + X^{*n}} \quad \text{Hill function for a repressor}$$

  - The input functions above go from a transcription rate of zero to a maximal transcription rate β. Many genes have a non-zero minimal expression level, called the gene's basal expression level. A basal level can be described by adding to the input function a term β_0.

- The essence of input functions is a transition between low and high values, with a characteristic threshold K. We will sometimes approximate input functions using a logic approximation
- In this approximation, the gene is either OFF, $f(X^*) = 0$, or ON, $f(X^*) = \beta$. The threshold for activation is K.
  Thus, logic input functions are step-like approximations for the smoother Hill functions.
- **For Activator:**

  $$f(X^*) = \beta\theta(X^* > K)$$

  For Repressor:

  $$f(X^*) = \beta\theta(X^* < K)$$

where
θ is equal to 0 or 1 according to the logic statement in the parentheses.

▼ Multi-Dimensional Input Functions Govern Genes with Several Inputs

- We saw how Hill functions and logic functions can describe input from a single transcription factor.
- Many genes, however, are regulated by multiple transcription factors. They are nodes in the network with two or more incoming arrows. Their promoter activity is thus a multi-dimensional input function of the different input transcription factors.
- Often, multi-dimensional input functions can be usefully approximated by logic functions, just as in the case of single-input functions.
- For example, consider genes regulated by two activators. Many genes require that both activator proteins bind to the promoter in order to show high expression. This is similar to an AND gate:

$$f(X^*, Y^*) = \beta\theta\,(X^* > K_x)\,\theta(Y^* > K_y) \sim X^* \text{ AND } Y^*$$

- For other genes, binding of either activator is sufficient. This resembles an OR gate:

$$f(X^*, Y^*) = \beta\theta(X^* > K_x \text{ OR } Y^* > K_y) \sim X^* \text{ OR } Y^*$$

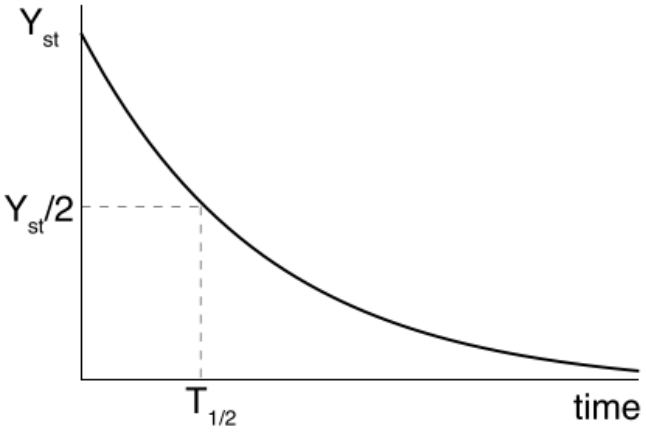
▼ Simple Regulation: Dynamics and Response Time

- We begin with the dynamics of a single arrow in the network. Consider a gene that is regulated by a transcription factor with no additional inputs.
- This transcription interaction is described in the network by $X \rightarrow Y$ which reads "**transcription factor X regulates gene Y**" Once X becomes activated by a signal (S_X), Y concentration begins to change.
- In the absence of its input signal, transcription factor X is inactive and Y is not produced. When the signal S_X appears, X rapidly transits to its active form X^* and binds the promoter of gene Y.

- Gene Y begins to be transcribed, and the mRNA is translated, resulting in accumulation of protein Y. The cell produces protein Y at a rate β.
- The production of Y is balanced by two processes, protein degradation and dilution. The degradation rate is α_{deg}, and the dilution rate is α_{dil}, giving a total removal rate of $\alpha = \alpha_{\text{deg}} + \alpha_{\text{dil}}$.
- Thus, β is responsible for production of Y while α is responsible for degradation of Y. The steady-state concentration can be found by solving for $dY/dt = 0$

$$dY/dt = \beta - \alpha Y$$

- If we now take away the input signal S_X, so the production of Y stops (β = 0). The result is an exponential decay of Y.



- **Response Time:** The response time, $T_{1/2}$, is defined as the time to reach halfway between the initial and final levels (half-life). The response time is therefore given by solving for the time when $Y(t) = Y_{st}/2$.

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

- The removal rate α directly determines the response time: fast removal allows rapid changes in concentration. The production rate β affects the steady-state level but not the response time.

Autoregulation

▼ Autoregulation is a Network Motif

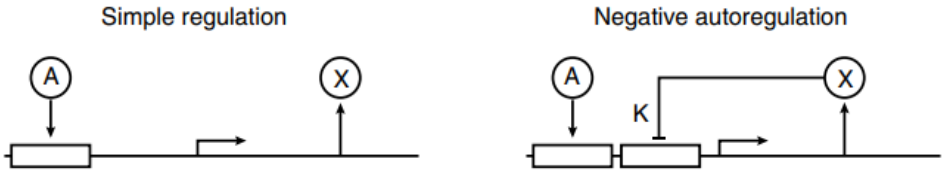
- Regulation of a gene by its own gene product is known as autogenous control, or autoregulation.
- Repressors that repress their own transcription is called negative autoregulation.
- For stable proteins that are not appreciably degraded in the cell, the response time is equal to the cell generation time.
  → The response time of a simply regulated gene is governed by its removal rate
  α.

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

- The negative autoregulation network motif can help speed up transcription responses.

▼ Effect of Negative Autoregulation on Response Time

- Negative autoregulation occurs when a transcription factor X represses its own transcription. This self-repression occurs when X binds its own promoter to inhibit production of mRNA. As a result, the higher the concentration of X, the lower its production rate.



- The dynamics of X are described by its production rate $f(X)$ and removal rate α:

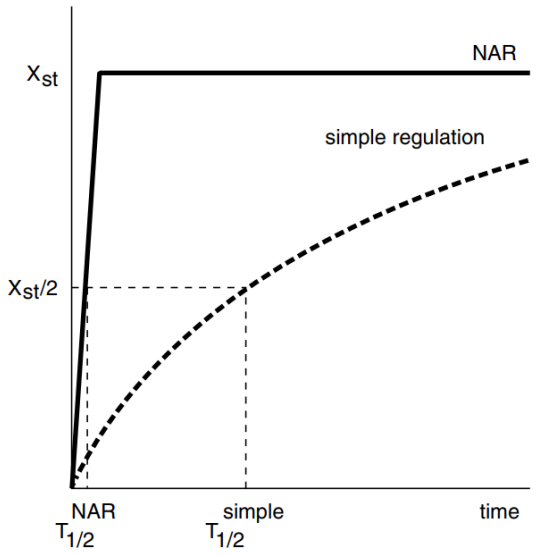
$$\frac{dX}{dt} = f(X) - \alpha X$$

- A good approximation for many promoters is a decreasing Hill function, when X is much smaller than the repression coefficient K, the promoter is free and the production rate reaches its maximal value, β.

On the other hand, when repressor X is at high concentration, no transcription occurs, $f(X) \sim 0$.

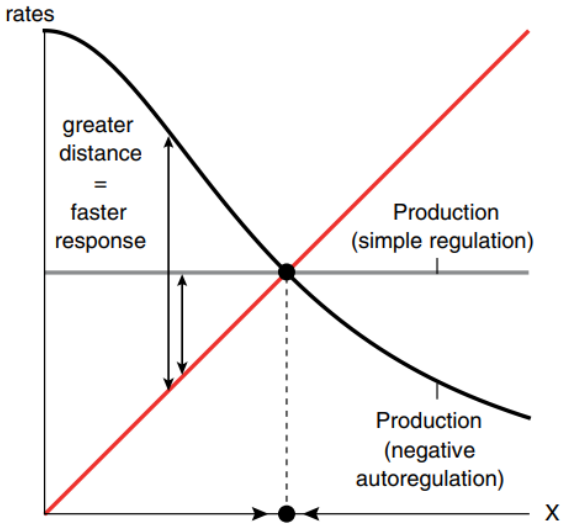
- Response time: Consider the case where X is initially absent, and its production starts at $t = 0$. At early times, while X conc is low, the promoter is unrepressed and production is at rate β. At early times, in fact, we can neglect removal ($\alpha X \ll \beta$) to find a linear accumulation of X with time →

  $X(t) \sim \beta t$ while $X < K$and $X << \beta/\alpha$
- However, production stops when X levels reach the self-repression threshold, $X = K$, because production is zero when X exceeds K.
- Small oscillations might occur now if there are delays in the system. It might cause $X$to go above K slightly, but then the removal rate will back it to fall back to below K, when production starts again, and this continues.
- These oscillations are generally damped for realistic $f(X)$ unless delays are very long. Thus, $X_{\mathrm{st}} = K$.
- **For simplicity**, we calculate the response time using linear accumulation of X in which $X = \beta t$.
  →
  $T_{\frac{1}{2}}^{\mathrm{NAR}} = \frac{K}{2\beta}$
- The stronger the maximal unrepressed promoter activity β, the shorter the response time. Negative autoregulation can therefore use a strong promoter to give an initial fast production, and then use autorepression to stop production at the desired steady state.



▼ Rate Analysis Shows Speedup for Any Repressive Input Function $f(X)$

- Speedup in Response time is also found when using Hill input functions and not just a step function. In fact, any shape of the input function $f(X)$, as long as it is a decreasing function of X, causes speedup in NAR.
- Rate Plot: In the rate plot, we plot the rates of production and removal as a function of protein level X.
- The value of X at which these two lines cross → At this point, production equals removal, and hence X levels don't change ($dX/dt = 0$). This is called a fixed point of the equation, namely the steady-state value.
- The speed at which X approaches the fixed point is given by the distance between the two curves, because the speed is the temporal derivative dX/dt = production − removal.



▼ Robustness

- In addition to speeding the response time, negative autoregulation confers a second benefit which is the increased robustness of the steady-state expression level with respect to fluctuations in the production rate β.
- Simple gene regulation is affected quite strongly by fluctuations in production rate β. The steady-state level is linearly dependent on the production rate. $X_{\mathrm{st}} = \frac{\beta}{\alpha}$, and therefore, a change in β leads to a proportional change in X_{st}.
- This is a problem because the production rate of a given gene, β, fluctuates over time due to variations in the metabolic capacity of the cell. These cell–cell differences in β are typically on the **order of tens of percent**. Thus, a snapshot of genetically identical cells grown under identical conditions will show cell–cell differences in the expression of every protein.
- In contrast, negative autoregulation buffers fluctuations in the production rate. In the case of the sharp autorepression that we have discussed, the steady-state level does not depend on β at all, and depends only on the repression threshold of X for its own promoter: $X_{\mathrm{st}} = K$.
- This is desirable because the repression threshold K is determined by hardwired factors such as the chemical bonds between X and its DNA site. Such parameters vary much less from cell to cell than production rates.

- Moreover, NAR can make the steady-state robust to changes in removal rate α, such as those that occur when the growth rate of the cells changes.
- Removal rate affects the steady state in simple regulation quite strongly. In contrast, NAR with a steep regulation function has a steady state that depends only weakly on α, making protein levels less sensitive to changes in cell growth rate.

> **Q:** What about positive autoregulation?
>
> **A:** We will see that positive autoregulation acts in an opposite way: it slows
> down responses and can amplify noise in parameters. Such slowdown and stochasticity can be useful for processes that take many cell generations, as occurs when organisms develop from an egg to an embryo.

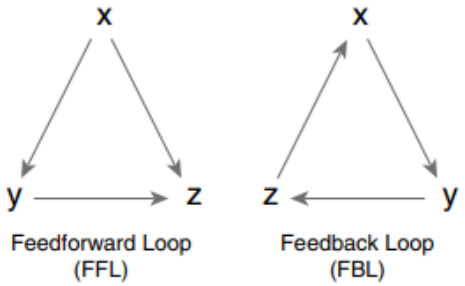
The Feedforward Loop

▼ Introduction

- Out of the many possible patterns that could appear in the transcription network, only a few are found significantly – the network motifs.
- Network motifs have defined information-processing functions.

▼ The Feedforward Loop is a Network Motif

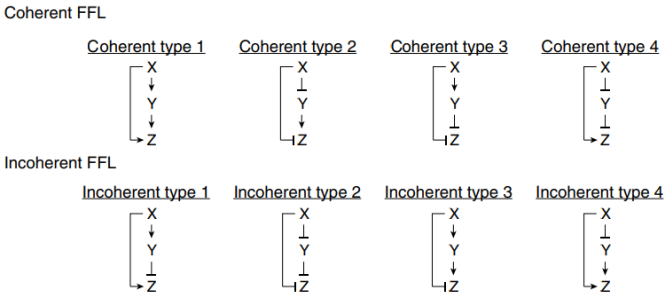
- Let us consider larger patterns of nodes and arrows, called subgraphs. In total, there are 13 possible ways to connect three nodes with directed arrows.



Difference between FFL and FBL

▼ The Structure of the Feedforward Loop Gene Circuit

- The feedforward loop (FFL), is composed of transcription factor X that regulates a second transcription factor, Y, and both X and Y regulate gene Z. Thus, the FFL has two parallel regulation paths, a direct path from X to Z and an indirect path that goes through Y.
- Each of the three arrows in the FFL can correspond to **activation (plus sign)** or **repression (minus sign)**. There are, therefore, $2^3 = 8$ possible types of FFLs. The eight FFL types can be classified into two groups: coherent and incoherent.
- In **coherent FFLs**, the indirect path has the same overall sign as the direct path. The overall sign of a path is given by multiplying the signs of the arrows on the path, so that two minus signs give an overall plus sign.
- In **incoherent FFLs**, the sign of the indirect path is opposite to that of the direct path.



- To understand the dynamics of the FFL we must also know how the inputs from the two regulators X and Y are integrated at the promoter of gene Z.
- Two biologically reasonable logic functions:

AND logic: in which both X and Y are needed to turn on Z expression.

OR logic: in which either X or Y is sufficient

▼ Coherent Type-1 FFL w/ AND Logic

→ Dynamics are as follows:

- At time $t = 0$, the signal S_{x} appears and triggers the activation of X. As a result, the transcription factor X rapidly transits to its active form X*. The active protein X* binds the promoter of gene Y, initiating production of protein Y, the second transcription factor in the FFL.
- In parallel, other copies of X* bind the promoter of gene Z. However, since the input function at the Z promoter is AND logic, X* alone cannot activate Z production.
- Production of Z requires binding of both X* and Y*. Z activation thus requires that the second input signal, Sy, is present, so that Y is in its active form, Y*. **Moreover (IMP)** the concentration of Y* must build up to sufficient levels to cross the activation threshold for gene Z, denoted K_{YZ}. This results in a delay in Z production.

→ Mathematics of the above Dynamics:

- Production of Y occurs at rate β_{Y} when $X*$ exceeds the activation threshold K_{XY}, as described by the step function θ.
  → Production rate of
  $Y = \beta_{\mathrm{Y}}\theta(X^* > K_{\mathrm{XY}})$
- The accumulation of Y is described by our familiar dynamic equation with a term for production and another term for removal:

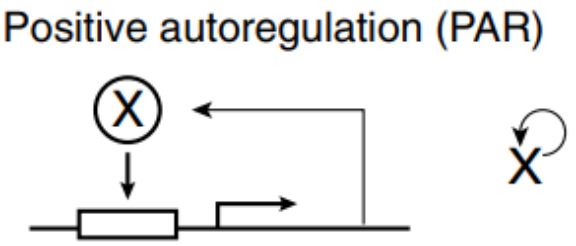
  $\frac{dY}{dt} = \beta_{\mathrm{Y}}\theta(X^* > K_{\mathrm{XY}}) - \alpha_{\mathrm{Y}}Y$
- The promoter of Z is governed by an AND-gate input function. The AND gate can be described by a product of two step functions, because both regulators need to cross their activation threshold (Remember, the activation threshold for X^* for production of Y is K_{XY} and for the production of Z is K_{XZ}):
  → Production of Z =
  $\beta_{\mathrm{Z}}\theta(X^* > K_{\mathrm{XZ}})\theta(Y^* > K_{\mathrm{YZ}}) - \alpha_{\mathrm{Z}}Z$
- Don't get confused, even Z will have a removal term $(\alpha_{\mathrm{Z}}Z)$.

Positive Feedback & Bistability

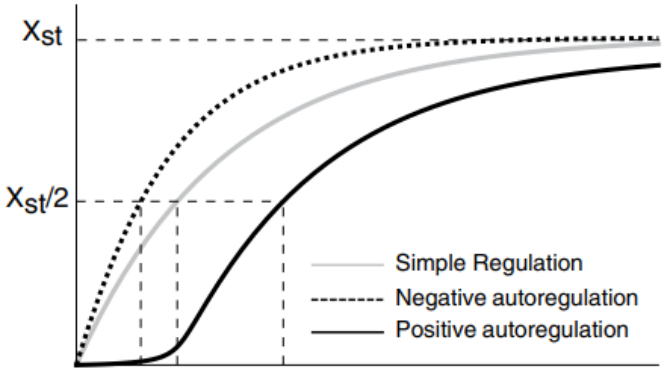
- The most common FFL types in developmental networks are the coherent type-1 and incoherent type-1 FFLs, just as in sensory networks. Developmental networks also display prominent autoregulation motifs.
- In addition to these motifs, developmental networks display a few additional network motifs that are not commonly found in sensory transcription networks. They are →

▼ Positive Autoregulation:

- Developmental networks have many positive autoregulation (PAR) loops. In PAR, a protein activates its own transcription.



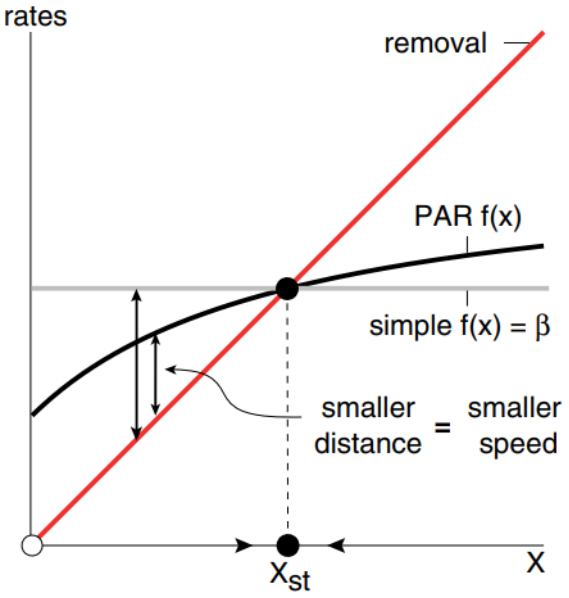
- Positive autoregulation has an opposite effect to that of negative autoregulation: it slows the response time relative to simple regulation.
- The dynamics are initially slow, but as the levels of X build up, it increases its own production and reaches halfway to steady state at a delay relative to simple regulation.



→ Rate Analysis for PAR:

- The equation is d$X/dt = f(x) - \alpha X$. Removal rate is a straight line, αX. Production rate is an increasing input function $f(X)$ appropriate

for the auto-activation of X



- The speed for approaching the fixed point is smaller in PAR than in simple regulation, which has a flat production curve $f(X) = \beta$. Thus, PAR shows slowdown for any increasing input function $f(X)$.
- The slow dynamics provided by PAR are useful in multi-stage processes that take a relatively long time, such as developmental processes. These processes can benefit from prolonged delays between the production of proteins responsible for different stages.
- Slow response times also help filter out rapidly varying noise in input signals, because slow responses integrate over this noise so that it cancels itself out.

Fold Change Detection

Here, we will study the ability of certain biological circuits to respond to relative changes in signal, instead of absolute changes.

▼ Universal Features of Sensory Systems

- Exact Adaptation: It is the ability to perfectly adjust to the background signal.
  Example: When we go from sunlight into a dark
  room lit by a candle, at first we don't see very well but after a while our pupils dilate to let in more light and our eyes adjust.

- Sensing of Relative changes rather than absolute changes: Suppose that we adapt to a room lit by a candle, and then we add a second candle. We sense a large change in light. But if we add the same candle to a room lit by a chandelier with 50 candles, we barely notice the change. The absolute number of photons added is the same, one candle's worth, but the relative change is very different.
- Weber's Law: Response to relative changes was described in human senses by Weber.

  $\Delta x_{\min} = kx_0$, where k is Weber's constant. This is called Weber's law: the just-noticeable difference is proportional to the background signal.
- In all cases, sensing of relative changes is found for an intermediate range of several decades of input signal (typically 2–5 decades). Relative sensing is lost at very weak signals on the brink of detection or very strong signals that saturate the receptors.