

# ELEN E6010 Systems Biology: Design Principles of Biological Circuits

## Lecture1: Intro, Basics of Transcription Networks

Prof. Predrag R. Jelenković  
Time: Tuesday 4:10-6:40pm  
227 Seeley W. Mudd Building

Dept. of Electrical Engineering  
Columbia University , NY 10027, USA  
Office: 812 Schapiro Research Bldg.  
Phone: (212) 854-8174  
Email: [predrag@ee.columbia.edu](mailto:predrag@ee.columbia.edu)  
URL: <http://www.ee.columbia.edu/~predrag>

# E6010 Systems Biology: Brief Description

- Recent successes in describing genomes of humans and model organisms raise a new set of challenges aimed at describing the complex dynamical mechanisms of gene regulation and protein interactions.
- Some of the fundamental features of these complex and large-scale systems include: **nonlinearity, transport delay, intricate feedback mechanisms, deterministic and stochastic kinetics, random networks, multiple time scales phenomena, modularity, hierarchical organization, robustness, increased reliability, kinetic proof reading and optimal evolutionary design.**
- The course provides an introduction to transcription regulation networks. We will see that these networks are made of repeating occurrences of simple patterns network motifs, i.e., elementary building blocks/circuits. Network motifs in other biological networks, including, developmental, signal transduction and neural networks are also discussed.

# E6010 Systems Biology: Brief Description

- The course will also focus on robustness of biological circuits: basically they are designed so that their function is insensitive to the naturally occurring stochastic fluctuations in the components of the circuit; this robustness principle will be illustrated on well-studied systems, including bacterial chemotaxis and patterning in fruit fly development.
- In the later part of the course, we will study how constrained evolutionary optimization can be used to understand the optimal circuit design, and how kinetic proof reading can minimize errors made in biological information processing.
- Interestingly, these features of biological systems, reuse of a small set of basic building blocks (network motifs), robustness (insensitivity) to component variations, modularity, hierarchical and constrained optimal design are also found in men made systems, suggesting a deeper connection that can unify our understanding of evolved and designed systems.

# E6010 Systems Biology: Course Logistics

**Prerequisites:** All the biology and modeling concepts will be covered from basic principles. However, some knowledge of molecular biology (E3060/E4060) and elementary concepts from calculus is desirable.

**Required text:** Lecture notes, research papers and the following textbook will be used:

1. An Introduction to Systems Biology: Design Principles of Biological Circuits by Uri Alon, Chapman & Hall, ISBN 1-58488-642-0.

Also, we recommend (not required) the following books:

2. Systems Biology, E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwing, ISBN 978-3-527-31874-2.
3. Physical Biology of the Cell, R. Philips, J. Kondev, J. Theriot, ISBN 978-0-8153-4163-5.
4. Essential Cell Biology, by B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, ISBN 0-8153-2045-0.

**Homework:** Assignments will be given weekly. Quantitative homework assignments may require the use of mathematical software packages MATHEMATICA or MATLAB.

**Grading:** Hwk (15%) + Midterm (35%) + Final Project (50%).

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

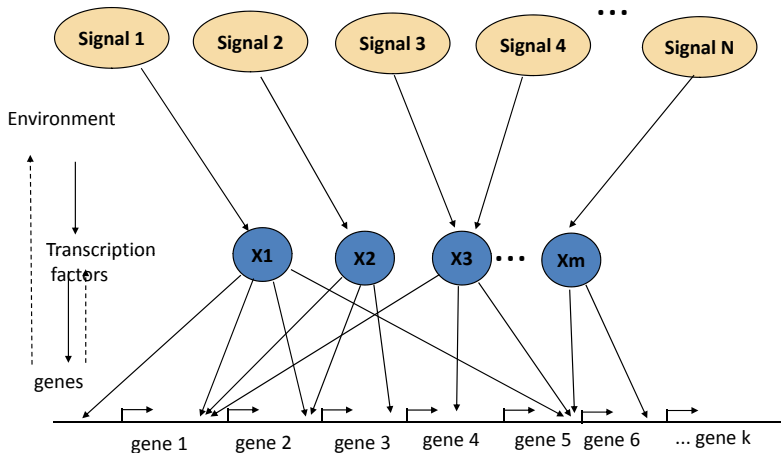
# Transcription Networks: Basic Concepts

- The cell is an integrated device made of several thousand types of interacting proteins (E. Coli has about 4000 types).
- Cell needs different proteins for different functions. It monitors the environment and determines the amount of each protein needed.
- E.g., when sugar is sensed, the cell produces proteins that can transport sugar into the cell and utilize it. When damaged, the cell produces repair proteins. Etc.
- Hence, the cell continuously monitors the environment and calculates the need for each protein. This information processing, which determines the rate of production of each protein, is primarily carried out by **transcription networks**.
- I.e., in engineering terms, transcription network is a controller or control network.

# The cognitive problem of the cell

- Cells live in a complex environment from which they sense many different signals, including physical parameters such as temperature and osmotic pressure, signaling molecules from other cells, nutrients, harmful chemicals, etc.
- Cells can also sense their internal state, such as the level of key metabolites and internal damage (e.g., to DNA, proteins).
- Cells respond to these signals by producing appropriate proteins that act upon these external or internal observations/signals.
- To represent these environmental signals/states, the cells use special proteins called **transcription factors** (TFs).
- Transcription factors are designed to change rapidly (sub-second time scale) between active and inactive states, depending on the inputs/signals from the environment. Each active TF can bind the DNA to regulate the rate of transcription of target genes.

# Elements of transcription networks

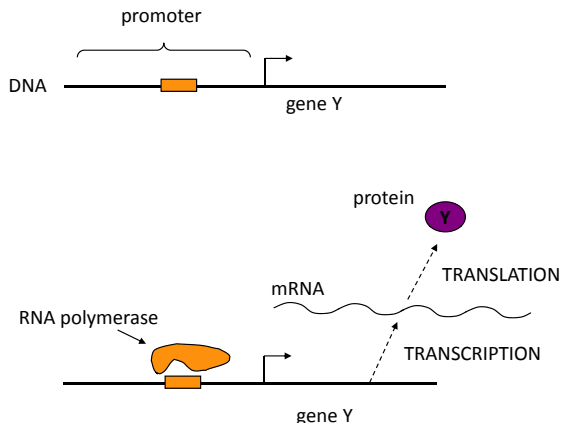


The mapping between environmental signals, transcription factors inside the cell and the genes that they regulate. The environmental signals activate specific transcription factor proteins. The transcription factors, when active, bind DNA to change the transcription rate of specific target genes, the rate at which mRNA is produced. The mRNA is then translated into protein. Hence, transcription factors regulate the rate at which the proteins encoded by the genes are produced. These proteins affect the environment (internal and external). Some proteins are themselves transcription factors, that can activate or repress other genes, etc.

# Elements of transcription networks

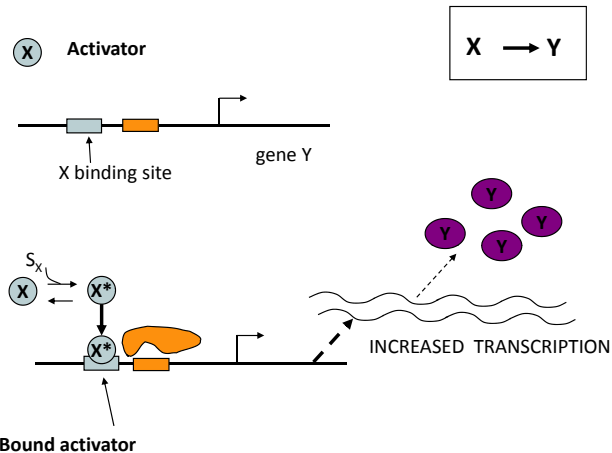
- The interaction between transcription factors (TF) and genes is described by transcription networks. Elements of these networks are genes and transcription factors.
- The rate at which the gene is transcribed is controlled by the **promoter**, a regulatory region of the gene that precedes the gene.
- When TFs are bound to the promoter region, they change the probability per unit time that RNAP binds the promoter and produces an mRNA molecule.
- TFs can act as **activators** that increase the transcription rate of the gene, or as **repressors** that reduce the transcription rate.
- TFs are themselves encoded by genes, which are regulated by yet another TFs, and so on. This set of interactions forms a **transcription network**.

# Gene transcription regulation: basic picture



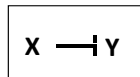
Each gene is usually preceded by a regulatory DNA region called the promoter. The promoter contains a specific site (DNA sequence) that can bind RNA polymerase (RNAP), a complex of several proteins that forms an enzyme that can synthesize mRNA that is complementary to the gene's coding sequence. The process of forming the mRNA is called transcription. The mRNA is then translated into protein.

# Gene transcription regulation: activators

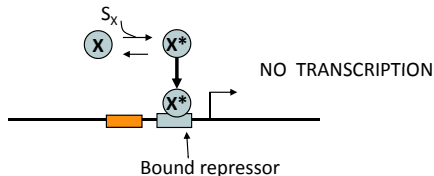


Activator X, is a transcription- factor protein that increases the rate of mRNA transcription when it binds the promoter. The activator transits rapidly between active and inactive forms. In its active form, it has a high affinity to a specific site (or sites) on the promoter. The signal  $S_x$  increases the probability that X is in its active form  $X^*$ . Thus,  $X^*$  binds the promoter of gene Y to increase transcription and production of protein Y.

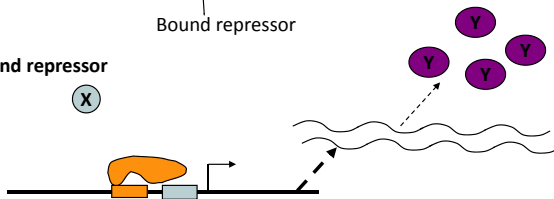
# Gene transcription regulation: repressors



Bound repressor



Unbound repressor

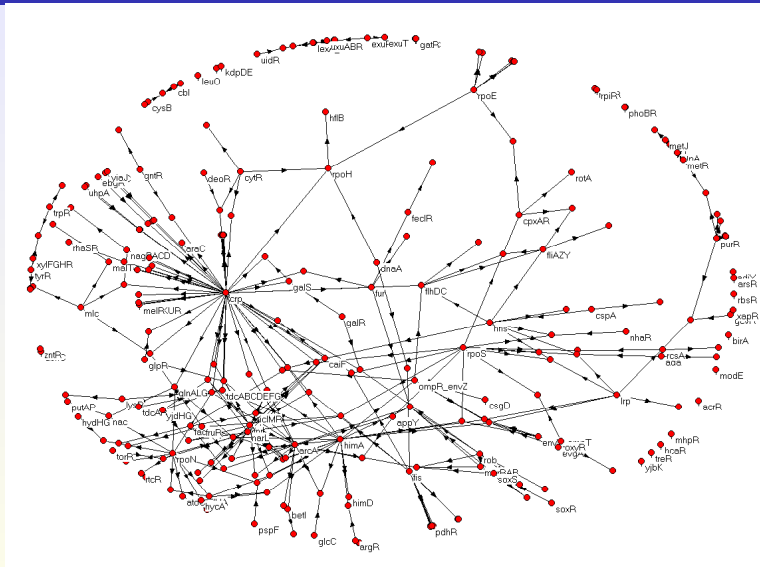


A repressor  $X$ , is a transcription-factor protein that decreases mRNA transcription when it binds the promoter. The signal  $S_X$  increases the probability that  $X$  is in its active form  $X^*$ .  $X^*$  binds a specific site in the promoter of gene  $Y$  to decrease transcription and production of protein  $Y$ . Many genes show a weak (basal) transcription when repressor is bound.

# Elements of transcription networks

- In this network, **nodes** are genes and **edges** are transcriptional regulations of genes by the protein products of others.
- $X \rightarrow Y$  - means that the product of gene  $X$  is a transcription factor protein that binds to the promoter of gene  $Y$ .
- The input to the network are **signals** that carry information from the environment. Each signal is a small molecule, protein modification, or molecular partner that directly affects the activity of one of the transcription factors. Signal  $S_X$  can cause  $X$  to rapidly change to its active state  $X^*$ .
- The network thus represent a dynamical system that, after an input signal arrives, TFs activities change, leading to the changes in the transcription rates of proteins.
- At last, let us note that not all the proteins are TFs, but rather carry out the diverse functions of the living cells, such as building structures (membranes) and catalyzing reactions.

# Transcription network of *E. Coli*



About 400 nodes  $\Rightarrow$  Complex network  $\Rightarrow$  What now?

# Understanding and Simplifying the Complexity in Space and Time

- Strong **separation of time scales**.
- **Modularity** of transcription networks.
- **Logic and step-function** approximations (of Hill input functions).
- **Network motifs**: basic building blocks of biological networks/circuits.
- **Robustness**: biological circuits have robust designs such that their essential function is nearly independent of biochemical parameters.
- **Kinetic proofreading**: How can a biochemical recognition system pick out a specific molecule in a sea of similar molecules?
- **Optimal gene circuit design**: Are bio-circuits designed in some optimal way for a given environment?
- Similarity between **men made/engineered and evolved biological systems**. Are there some deeper explanations for these similarities?

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

# Separation of time scales

- Transcription networks are designed with a strong separation of time scales: the input signals usually activate TFs on a sub-second time scale.
- Binding of an active TF to its DNA reaches equilibrium in seconds.
- Transcription and translation takes minutes.
- Accumulation of the protein takes many minutes to hours.
- Typical approximate time scales for *E. coli*:
  - 1 Binding signaling molecule to a TF  $\sim 1$  msec.
  - 2 Binding active TF to its DNA site  $\sim 1$  sec.
  - 3 Transcription + translation of the gene  $\sim 5$  min.
  - 4 50% change of protein concentration  $\sim 1$  h.
- Hence, **1 and 2 can be considered instantaneous** when studying transcription networks.

- 1 Transcription Networks: Basic Concepts
- 2 **Understanding and Simplifying the Complexity in Space and Time**
  - Separation of time scales
  - **Modularity of transcription networks**
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

# Modularity of transcription networks

- Remarkable property of transcription networks is the modularity of their components. One can take a DNA form one organism and express it in a different organism.
- For example, one can take the gene for green fluorescent protein (GFP) from jellyfish and introduce it to bacteria. As a result, the bacteria produce GFP, causing its color to turn green.
- Regulation can also be added by adding a promoter region.
- E.g., control of GFP in the bacteria can be achieved by passing the gene next to the promoter for another gene, say, the one that is controlled by a sugar-inducible transcription factor. This causes *E. coli* to express GFP and turn green only in the presence of sugar.

# Modularity of transcription networks

- Hence, promoters and genes are generally interchangeable. This modularity makes transcription nets very plastic during evolution by being able to readily incorporate new genes and regulations.
- In fact, transcription nets can evolve very rapidly: the edges appear to evolve on a faster time scale than the nodes (coding regions).
  - For example, related animals, such as humans and mice, have very similar genes, but the transcription regulation of these genes, is evidently quite different.
  - I.e., many of the differences between animal species appear to lie in the differences in the edges of the transcription networks, rather than in the differences in their genes.

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - **Transcription networks: functions of the edges**
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

# Transcription networks: signs on the edges

- Each edge in transcription network corresponds to an interaction in which TFs directly control the transcription rate of a gene.
- These interactions can be of two types:
  - **Activation, or positive control:** TF increases the rate of transcription when binds to a promoter.
  - **Repression, or negative control:** TF reduces the rate of transcription when it binds to the promoter.
- Interestingly, transcription networks often have comparable number of positive and negative controls, e.g., *E. coli* or yeast have 60-80% of activation interactions.

# Transcription networks: signs on the edges

- Typically, TFs act either as activators or repressors. However, activators that regulate many genes act as repressors for some of their target genes.
- Hence, in general, TFs tend to employ one mode of regulation for most of their target genes.
- In contrast, the signs on the edges that go into a node are less correlated.
- In short, the signs of the **outgoing edges are rather correlated**, but the signs of the **incoming edges are rather not**.

# Transcription networks: functions of the edges

- The edges not only have signs, but also have a certain strength that is described by an **input function**.
- When,  $X$  regulates  $Y$ , represented in the network as  $X \rightarrow Y$ , the number of molecules of protein  $Y$  produced per unit of time is a function of the concentration of  $X$  in its active form  $X^*$ :

$$\text{rate of production of } Y = f(X^*).$$

- Typically,  $f(X^*)$  is monotonic, S-shaped function. It is an increasing function when  $X$  is an activator and decreasing when it is a repressor.

# The numbers on the edges: the input function

- A function that describes many real gene input functions is called **Hill function**; here is a Hill function of an activator:

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

where

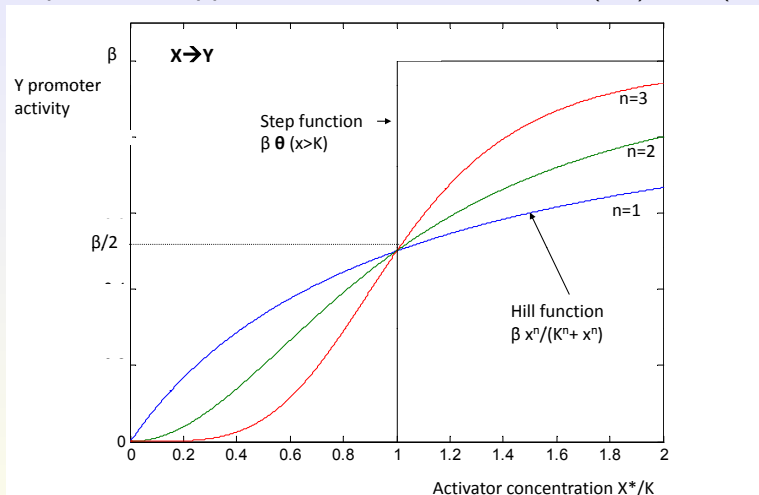
- $K$  - is termed the **activation coefficient**, and has units of concentration;
- $\beta$  - is the **expression level** of the promoter;
- $n$  - is the **Hill coefficient** that governs the steepness of the input function; the larger  $n$  - the steeper the input function. Typical values for  $n$  are 1 – 4.
- Hill input function for repressor:

$$f(X^*) = \frac{\beta}{1 + \frac{X^{*n}}{K^n}}.$$

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - **A simple approximation: logical/step function  $n \approx \infty$**
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

# A simple approximation: logic/step function $n \approx \infty$

Logic/step function approximation for an activator:  $f(X^*) = \beta \theta(X^* > K)$



Input functions for activator X described by Hill functions with Hill coefficient  $n=1,2$  and  $4$ . Black line: step/ logical input function. The maximal promoter activity is  $\beta$ , and  $K$  is the threshold for activation of a target gene (the concentration of  $X^*$  needed for 50% maximal activation).

# A simple approximation: logic/step function $n \approx \infty$

Logic/step function approximation for a repressor:  $f(X^*) = \beta \theta(X^* < K)$

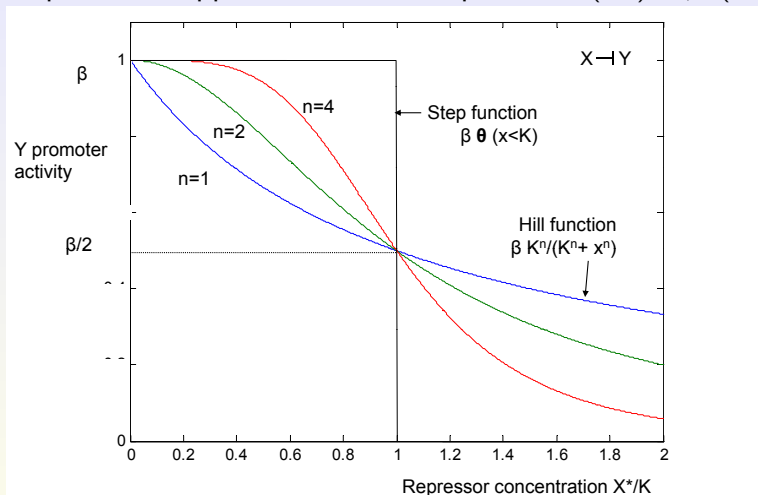


Figure 2.4b: Input functions for repressor X described by Hill functions with Hill coefficient  $n=1, 2$  and  $4$ . Black line: logical input function (step function). The maximal unrepressed promoter activity is  $\beta$ , and  $K$  is the threshold for repression of a target gene (the concentration of  $X^*$  needed for 50% maximal repression).

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - **Multi-dimensional input functions**
  - Dynamics and response time of simple gene regulation
  - Network motifs: in search of basic building blocks

# Multi-dimensional input functions

- Many genes are regulated by multiple transcription factors.
- Often, these multi-dimensional input functions can be approximated by logic functions.
- Many genes require simultaneous binding of two (or more) TFs in order to show significant expression; this is similar to AND gate:

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

- For other genes binding of either activator is sufficient; this resembles the OR gate:

$$f(X^*, Y^*) = \beta\theta(X^* > K \text{ OR } Y^* > K).$$

- However, not all genes have Boolean-like input functions, e.g., some genes display SUM input functions:

$$f(X^*, Y^*) = \beta_X X^* + \beta_Y Y^*.$$

- 1 Transcription Networks: Basic Concepts
- 2 **Understanding and Simplifying the Complexity in Space and Time**
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - **Dynamics and response time of simple gene regulation**
  - Network motifs: in search of basic building blocks

# Dynamics and response time of simple gene regulation

- Change of concentration of  $Y$  is due to the difference between its production and degradation/dilution, as described by the following equation:

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

- In steady state  $dY/dt = 0$  and, thus

$$Y_s \equiv Y(\infty) = \frac{\beta}{\alpha}.$$

- When the production stops, i.e.,  $\beta = 0$ , then  $Y(t)$  degrades as

$$Y(t) = Y_s e^{-\alpha t}.$$

- its half response time is defined at the time  $t$  when  $Y(t)$  reaches  $Y_s/2$

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

# Dynamics and response time of simple gene regulation

- If an unstimulated gene becomes stimulated by a strong signal  $S_x$ , then the solution to

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

- is given by

$$Y(t) = Y_s(1 - e^{-\alpha t}).$$

- Initially, when  $\alpha t \ll 1$ ,  $Y(t)$  grows basically linearly

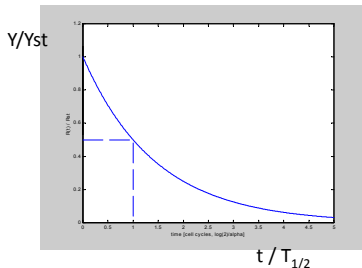
$$Y(t) \approx \beta t.$$

- The response time, defined at the time  $t$  when  $Y(t)$  reaches  $Y_s/2$ , is again

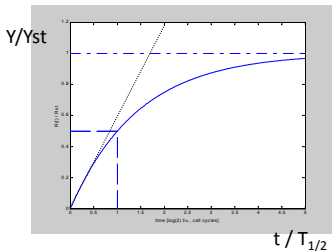
$$T_{1/2} = \log(2)/\alpha.$$

# Dynamics and response time of simple regulation

Change of concentration of  $Y$  is due to the difference between its production and degradation/dilution:  $dY/dt = \beta - \alpha Y$ .



Decay in protein concentration following a sudden drop in production rate. Note the response-time, the time it takes the concentration to reach half of its variation, is  $T_{1/2} = \log(2)/\alpha$ .

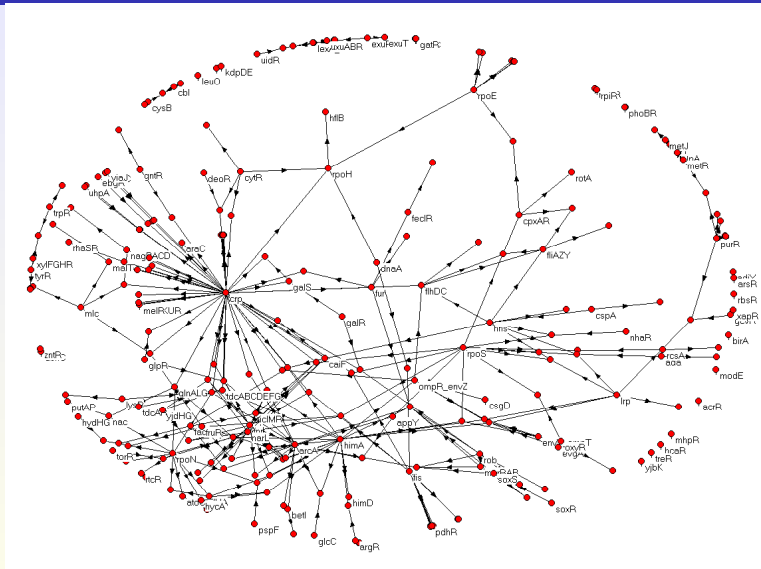


Rise in protein concentration following a sudden increase in production rate. Note the response time, the time it takes the dynamics to reach half of its variation, is  $T_{1/2} = \log(2)/\alpha$ . At early times, the protein accumulation is approximately linear with time,  $Y = \beta t$ .

**Response time is very important for designing efficient gene circuits.**

This is basically the end of Lecture 1, the rest gives some hints on what is more to come.

# Transcription network of *E. Coli*



Still left with a complex network  $\Rightarrow$  What now?

- 1 Transcription Networks: Basic Concepts
- 2 Understanding and Simplifying the Complexity in Space and Time
  - Separation of time scales
  - Modularity of transcription networks
  - Transcription networks: functions of the edges
  - A simple approximation: logical/step function  $n \approx \infty$
  - Multi-dimensional input functions
  - Dynamics and response time of simple gene regulation
  - **Network motifs: in search of basic building blocks**

# Network motifs: in search of basic building blocks

- Our goal will be to define understandable patterns of connections that serve as building blocks of the network.
- Ideally, we would like to understand the dynamics of the entire network based on the dynamics of the individual building blocks.
- Now, we will:
  - 1 Define a way to detect building-block patterns in complex networks, called **network motifs**.
  - 2 Examine the simplest network motif in transcription networks, **negative autoregulation**.
  - 3 Show that this motif has useful functions: speeding up the **response time of transcription interactions** and stabilizing them.

# Patterns, randomized networks, and network motifs

- Our approach will be to look for meaningful patterns on the basis of statistical significance. To define statistical significance, we compare the network to an ensemble of **randomized networks**.
- The randomized networks are networks with the same characteristics as the real network, but where the connections between nodes and edges are made at random. Patterns that occur in the real network significantly more often than in randomized ones are called **network motifs**.
- The basic idea is that patterns that occur in the real network much more often than in randomized networks must have been preserved over evolutionary timescales against mutations that randomly change edges.

# Autoregulation: a network motif

- Now we can begin to compare features of the real *E. coli* transcription network with the randomized networks. Let us start with **self-edges**, edges that originate and end at the same node.
- The *E. coli* network that we use as an example has 40 self-edges. These self-edges correspond to transcription factors that regulate the transcription of their own genes.
- Regulation of a gene by its own gene product is known as autogenous control, or **autoregulation**. Thirty-four of the autoregulatory proteins in the network are repressors that repress their own transcription: **negative autoregulation**.
- Is autoregulation significantly more frequent in the real network than at random?

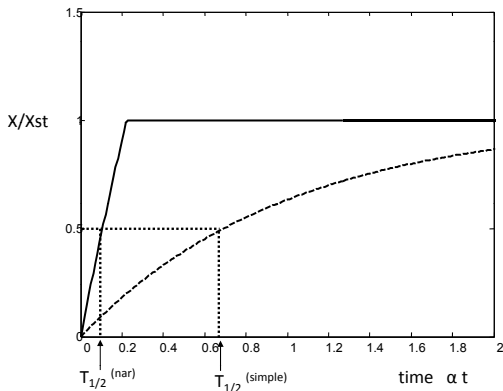
# Autoregulation: a network motif

- In the *E. coli* transcription network, the number of nodes and edges are  $N = 424$  and  $E = 519$ .
- Thus, a corresponding random network with the same  $N$  and  $E$  would be expected to have only about one self-edge, plus minus one:

$$\langle N_{self} \rangle_{rand} \sim E/N \sim 1.2 \quad \sigma_{rand} \sim \sqrt{1.2} \sim 1.1$$

- **In contrast, the real network has 40 self-edges!**, which exceeds the random networks by many standard deviations.
- In particular negatively autoregulated gene is a network motif.  
**Why?**

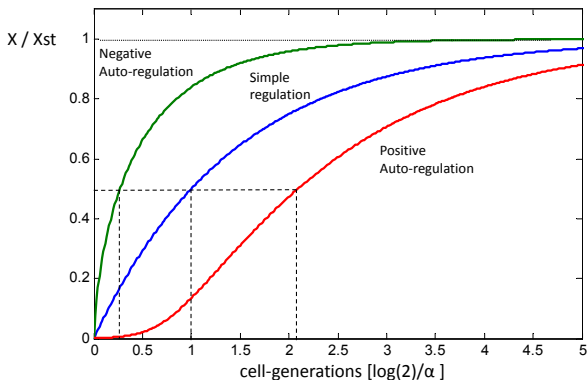
# Negative autoregulation speeds up the response time



Dynamics of negatively auto-regulated gene product (full line) and simply regulated gene product (dashed line) which reach the same steady-state level and have equal degradation/dilution rates  $\alpha$ . The response time is the time that the protein level reaches 50% of the steady state, denoted  $T_{1/2}^{(nar)}$  and  $T_{1/2}^{(simple)}$  for the negatively auto-regulated and simply regulated gene products. The parameters  $\beta=5$ ,  $\alpha=1$ ,  $\beta^{simple}=1$  were used.

Also, it increases robustness to the variation of production rate  $\beta$

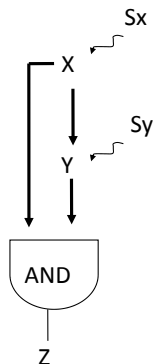
# What about positive autoregulation?



Dynamics of negatively auto-regulated gene, a simply regulated gene and a positively auto-regulated gene. The negatively and positively auto-regulated genes have a Hill-input function with Hill coefficient  $n=1$ . Shown is protein concentration normalized by its steady-state value  $X/X_{st}$ , following an increase in production rate. Time is in cell-generations, or for actively degraded proteins  $\log(2)/\alpha$ , where alpha is the protein degradation/dilution rate.

Slows response time, leads to bi-stability and increases input sensitivity.

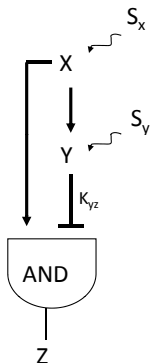
## Three node motifs: Feed-Forward Loops (FFLs)



The coherent type-1 FFL with an AND input function: Transcription factor X activates the gene encoding transcription factor Y, and both X and Y jointly activate gene Z. The two input signals are  $S_x$  and  $S_y$ . An input-function integrates the effects of X and Y at the Z promoter (an AND-gate in this figure).

C1-FFL is a sign sensitive delay element: filters brief input fluctuations.

# Incoherent type 1 FFL: pulse generator



The incoherent type-1 FFL with an AND gate at the Z promoter. The inputs are the inducers  $S_x$  and  $S_y$ . The repression threshold of gene Z by repressor Y is  $K_{yz}$ .

We will also attempt to answer why some other types of FFLs are rare.

# More to come

- Temporal programs and the global structure of transcription networks
- Network motifs in:
  - Developmental
  - Signal Transduction
  - Neuronal Networks
- Robustness of protein circuits: the example of bacterial chemotaxis
- Robust patterning in development
- Kinetic proof reading
- Optimal gene circuit design
- Demand rules for gene regulation: activators or repressors?
- Deeper underlying questions:
  - Simplicity in biology: bio nets evolved to function, not to be understandable?
  - Why are men made/engineered and evolved biological systems similar? Are they?

Please read the lecture notes and Chapter 2 of Uri Alon book; please try to understand the exercises 2.1-2.3 as well.

Also, if you have time please browse through the first 9 chapters of "Essential Cell Biology" book.