

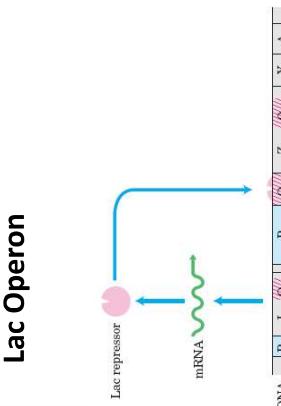
# Computational Aspects of Gene Regulation

Lecture 13

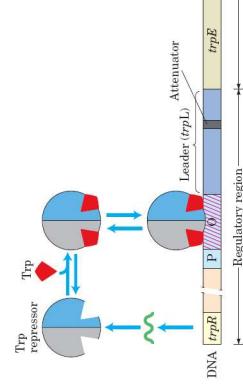
## Objectives

- In this lecture you will learn about:
1. What are gene regulatory networks
  2. Steps in building a gene regulatory network
  3. Model of a transcription module
  4. Network motifs and logic gates
  5. Boolean representations

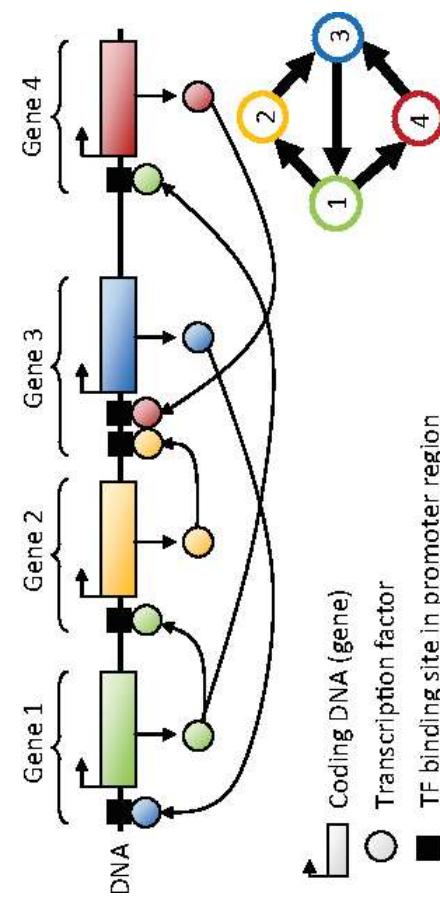
## The Lac and Trp Operons



### Trp Operon



## A typical gene regulatory network

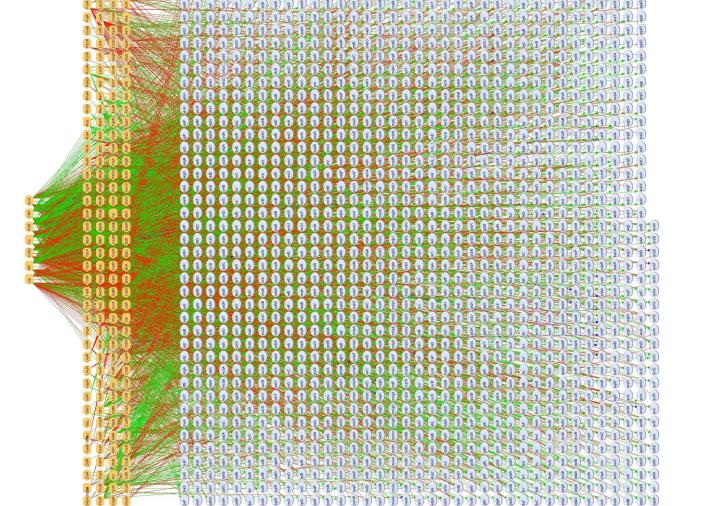


Legend:  
■ Coding DNA (gene)  
○ Transcription factor  
■ TF binding site in promoter region

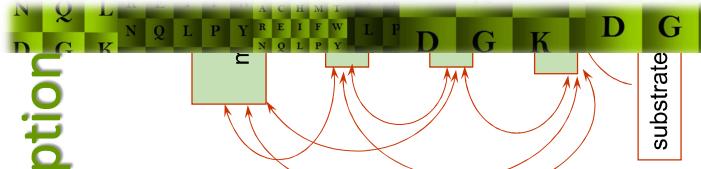
# What are Gene Regulatory (Transcription) Networks?

## This is one of the layers of information generation and transmission within a cell

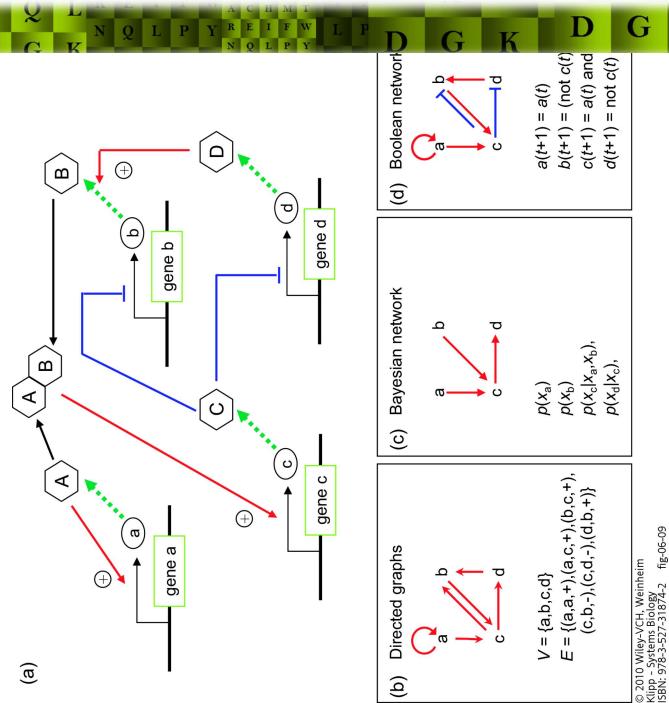
- ❖ It is the regulation of gene expression at the stage of transcription
- ❖ This is broadly the first step to gene expression and its control regulates the temporal programming in genes
- ❖ It is of interest because it mediates changes in cells and helps in understanding the onset and progression of disease



## Regulatory Network of *E. coli* K12

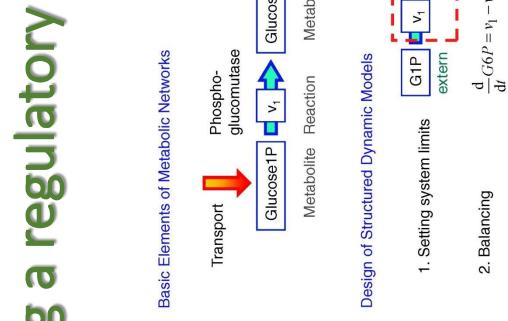


## There are different methods of analysis



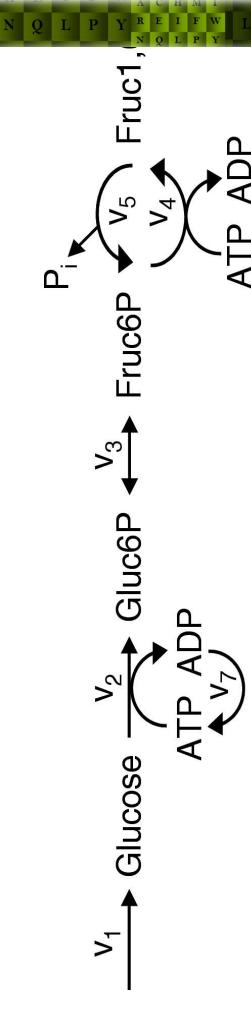
## Steps to follow for building a regulatory network model

- ❖ Identify the elements of the model
- ❖ Characterize the kind of interaction/reaction
- ❖ Define the boundary of your observation (system)
- ❖ Identify the information/flow “into” and “out of” the boundary
- ❖ What are the intrinsic generation/degradation rates?
- ❖ Are all the parameters known?
- ❖ Assign the kinetics
- ❖ Code and Simulate

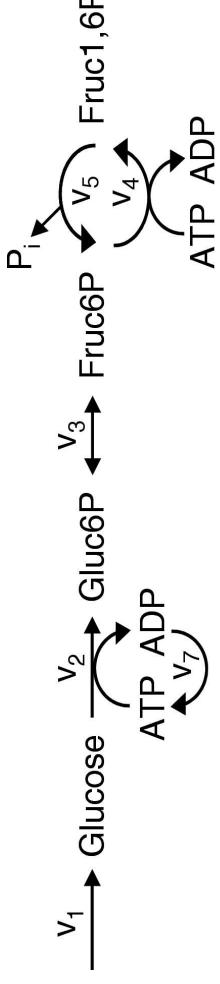


## Add the details of the important constituents

## Define the rate equations



- ❖ Layout the reactions involved
- ❖ Are the rates balanced?



❖ This is the most important step

- ❖ Go to the literature and ensure correctness of the reactions
- ❖ Use various resources to determine the parameters that have been reported for the same or similar reactions (or cellular events)
- ❖ Do your experiments results appear reasonable? Have the parameters been evaluated correctly?

$$\begin{aligned}\frac{d}{dt} Glu &= v_1 - v_2 \\ \frac{d}{dt} G6P &= v_2 - v_3 \\ \frac{d}{dt} F6P &= v_3 - v_4 + v_5 \\ \frac{d}{dt} F1,6P2 &= v_4 - v_5 + v_6 \\ \frac{d}{dt} ATP &= -\frac{d}{dt} ADP = -v_2 - v_4\end{aligned}$$

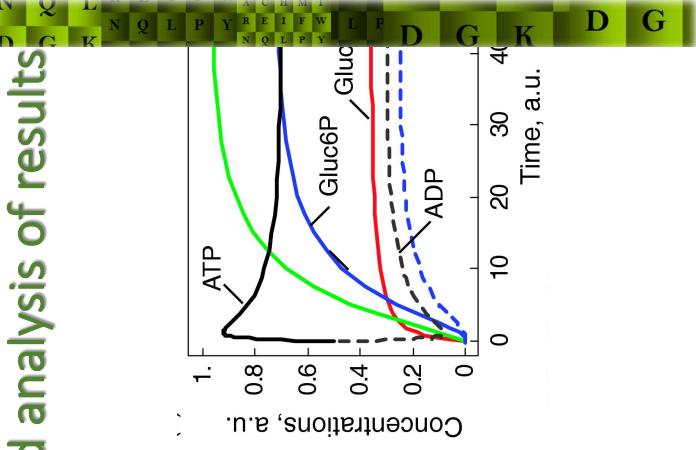
## Performing simulation and analysis of results

## Transcription Modules

- ❖ In the general form, the ODE that models the output of the gene ( $Z$ ) in response to a regulatory input  $S$  is given by the equation

$$\begin{aligned}\frac{dZ}{dt} &= k' + \frac{k \cdot (S^n / K^n)^\mu}{1 + (S^n / K^n)^\mu} - k_d \cdot Z \\ Z_{ss} &= \frac{k}{k_d} \left( a + \frac{(S^n / K^n)^\mu}{1 + (S^n / K^n)^\mu} \right), \quad k' = a \cdot k\end{aligned}$$

- ❖ Repression and Activation are taken care of by the parameter  $\mu$
- ❖  $\mu = 0 \rightarrow$  repression;  $\mu = 1 \rightarrow$  activation
- ❖ The parameters  $k'$  and  $k$ , represent the signal-independent and the signal-dependent gene expression.



- ❖ Cross check the results to make sure that it makes sense
- ❖ Check the boundary results
- ❖ Do the values at the boundaries satisfy the physical constraints?
- ❖ Plot the data/results and analyze
- ❖ What is the insight you have gained?
- ❖ What is the hypothesis you can propose?



## Example of a cascade of genes

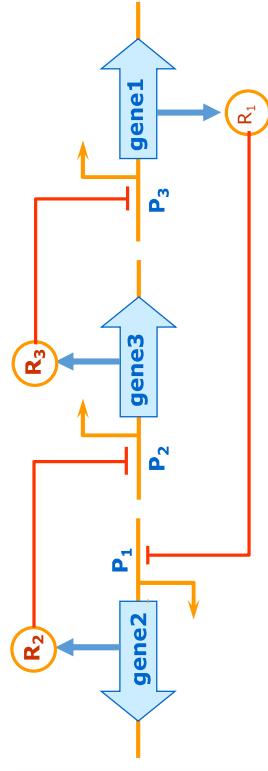
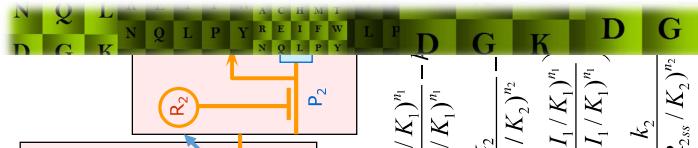
❖ Promoter  $P$  has no regulatory inputs

❖  $P$  is constitutive and drives the expression of  $R_1$  which in turn inhibits  $R_2$

❖ The inducer  $I_1$  regulates the signal  $R_2/P_2$  by modulating the cellular abundance of  $R_2$

$$P1: n = 2.4, K = 5.5 \text{ nM}, k = 220 \text{ min}^{-1},$$

$$P2: n = 1.7, K = 120 \text{ nM}, k = 255 \text{ min}^{-1},$$

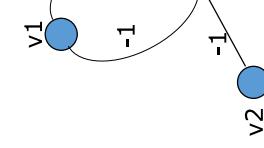
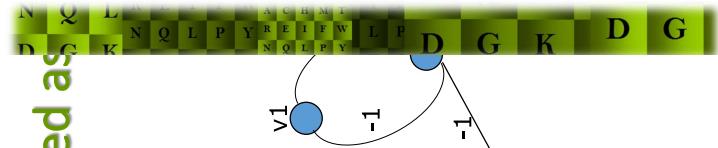


$$\begin{aligned} \frac{dR_2}{dt} &= a_1 k_1 + \frac{k_1 \cdot (I_1 / K_1)^{n_1}}{1 + (I_1 / K_1)^{n_1}} - k_2 R_2 \\ \frac{dZ}{dt} &= a_2 k_2 + \frac{k_2}{1 + (R_2 / K_2)^{n_2}} - k_3 Z \\ R_{2ss} &= \frac{k_1}{k_{a2}} \left( a_1 + \frac{k_1 \cdot (I_1 / K_1)^{n_1}}{1 + (I_1 / K_1)^{n_1}} \right) \\ Z_{ss} &= \frac{k_2}{k_d} \left( a_2 + \frac{k_2}{1 + (R_{2ss} / K_2)^{n_2}} \right) \end{aligned}$$



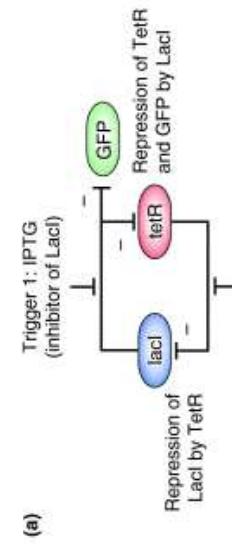
$$\begin{aligned} \frac{dm_i}{d\tau} &= \alpha \kappa + \frac{\kappa}{1 + r_j'^n} \\ \frac{dr_i}{d\tau} &= \varepsilon(m_i - r_i) \end{aligned}$$

- ❖ Repressor  $R_1$  inhibits the expression of repressor  $R_2$ , repressor  $R_2$  inhibits the expression of repressor  $R_3$ , and repressor  $R_3$  inhibits the expression of repressor  $R_1$
- ❖ The separation of transcription and translation contributes to a response delay that results in the emergence of oscillations



## Signed Graphs

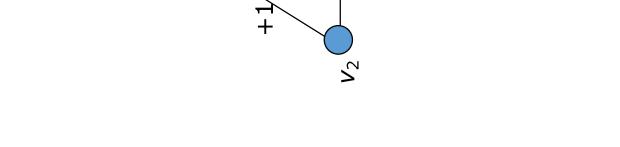
### Regulatory circuits can also be represented as signed graphs



- A **signed graph**  $S$  is an undirected network whose edges have functional values of  $+1$  or  $-1$ ; it is natural to refer to them as a positive edge or negative edge.

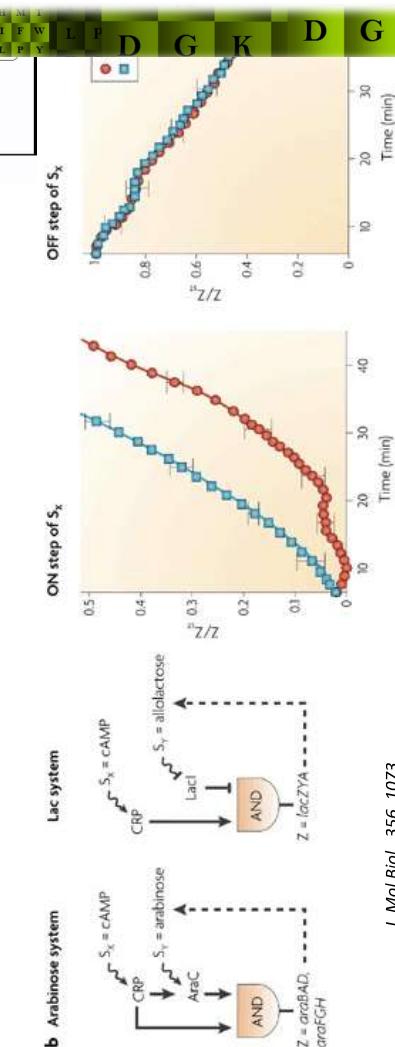
For example:

$$\begin{aligned} V &= (V_1, V_2, V_3), E = (V_1V_2, V_2V_3, V_3V_1) \text{ and} \\ f &= \{(V_1V_2, +1) (V_2V_3, -1) (V_3V_1, -1)\} \end{aligned}$$



# Gene Regulation can also be treated as logic

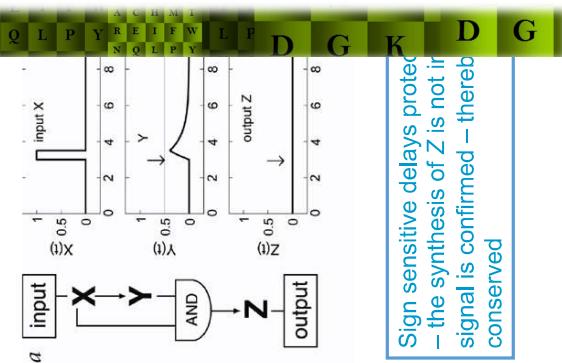
- ❖ Arabinose is only used if glucose is not present; proteins in this system are made only when condition arabinose “AND NOT” glucose is satisfied
- ❖ The delay  $T_{ON} \sim 20$  min
- ❖  $X = CRP, S_x = cAMP, S_y = \text{arabinose}, Y = araC,$   
In the lacrose operon,  $X$  does not regulate  $Y = lacI$



J. Mol Biol., 356, 1073.

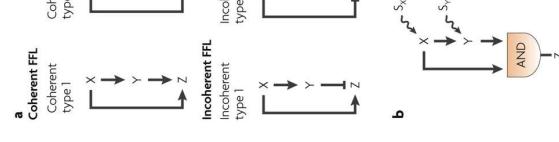
## The FFL is a persistence detector

- ❖ A sign sensitive delay element can be considered as an asymmetric filter
- ❖ A brief pulse of  $X$  is results in a signal shorter than  $T_{ON}$
- ❖ However, the motif responds immediately to a pulse OFF signal

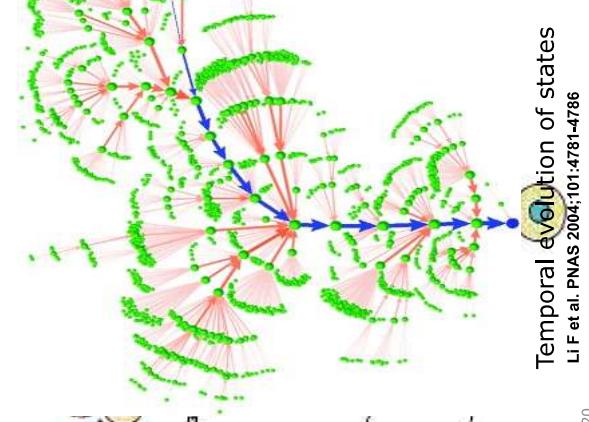


# Structure of the feed-forward loop

- ❖ The feedforward loop in consideration
- ❖ Has a direct path from  $X$  to  $Z$
- ❖ has an indirect path  $X \rightarrow Y \rightarrow Z$
- ❖ Each edge can be an activation or repression; so there are  $2^{3=8}$  FFLs
- ❖ These are classified into two groups
  - ❖ Coherent: the indirect path has the same overall sign as the direct path
  - ❖ Incoherent: the sign of the indirect path is opposite to that of the direct path
- ❖ There are two possible logic gates for the expression of  $Z$ : AND or OR



## Yeast cell cycle



## Boolean Attractors

- ❖ A boolean network is defined by  $G(V, F)$

$$V = \{v_1, v_2, \dots, v_n\}$$

$$F = \{f_1, f_2, \dots, f_n\}$$

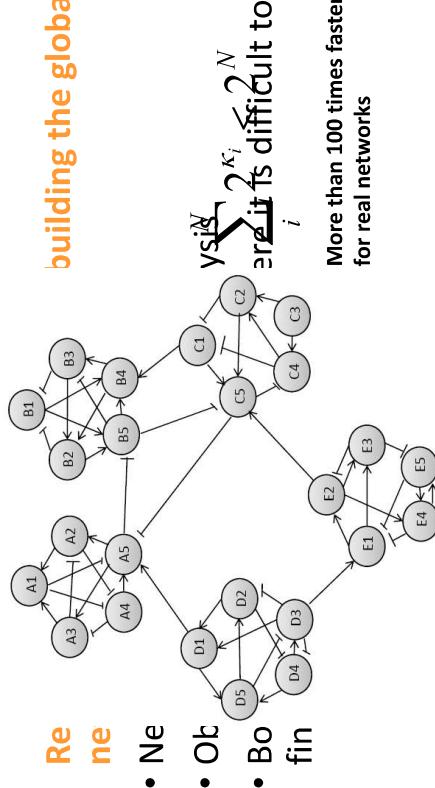
- ❖ Let  $v_i(t)$  represent the state of  $v_i$  at time  $t$ . The overall expression level of all the genes in the network at time step is given by the following vector:

$$v(t) = [v_1(t), v_2(t), \dots, v_n(t)]$$

- ❖ There are  $2^n$  possible states; the regulatory rules among the genes are given as follow

$$v_i(t+1) = f_i(v_{i_1}(t), v_{i_2}(t), \dots, v_{i_k}(t)), \quad i = 1, 2, \dots, n.$$

21



22

More than 100 times faster  
for real networks

## Toy Network Example

Determine the steady states of subgraphs



23

