Modeling Gene Regulatory Networks (GRN)

Yinghao Wu
Department of Systems and Computational Biology
Albert Einstein College of Medicine
Fall 2014

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Outline

- Introduction to gene regulation

- Construction of GRN

- Modeling the dynamics of GRN

Outline

- Introduction to gene regulation

- Construction of GRN

  - Unsupervised

  - Supervised

- Modeling the dynamics of GRN

  - Discrete Models (Boolean Network)

  - Differential and Stochastic Equations

Outline

- Introduction to gene regulation

- Construction of GRN

- Modeling the dynamics of GRN

1: Proteins interacting with DNA turn genes on or off in response to environmental changes

- **Gene expression** is the overall process of information flow from genes to proteins

  - Mainly controlled at the level of transcription

  - A gene that is "turned on" is being transcribed to produce mRNA that is translated to make its corresponding protein

  - Organisms respond to environmental changes by controlling gene expression

1: Proteins interacting with DNA turn genes on or off in response to environmental changes

- An **operon** is a group of genes under coordinated control in bacteria

- The lactose (*lac*) operon includes

  - Three adjacent genes for lactose-utilization enzymes

  - **Promoter** sequence where RNA polymerase binds

  - Operator sequence is where a **repressor** can bind and block RNA polymerase action
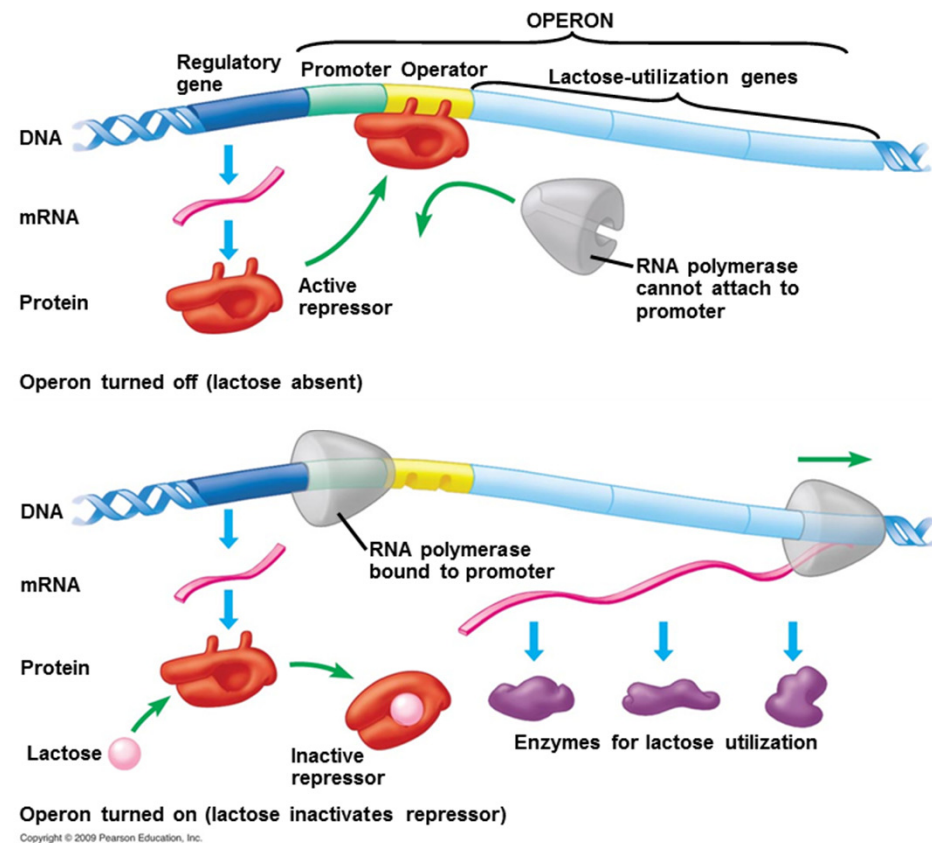
1: Proteins interacting with DNA turn genes on or off in response to environmental changes

- Regulation of the *lac* operon

  - **Regulatory gene** codes for a repressor protein

  - In the absence of lactose, the repressor binds to the operator and prevents RNA polymerase action

  - Lactose inactivates the repressor, so the operator is unblocked
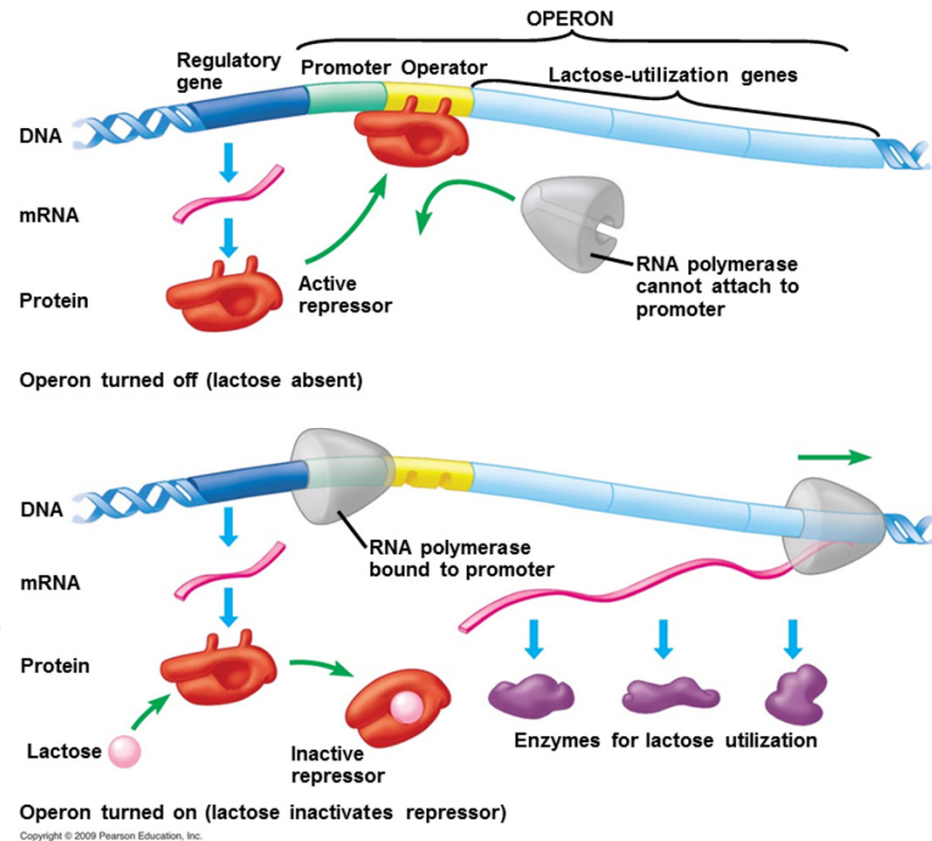
1: Proteins interacting with DNA turn genes on or off in response to environmental changes

- Types of operon control
  - Inducible operon (*lac* operon)
    - Active repressor binds to the operator
    - Inducer (lactose) binds to and inactivates the repressor
  - Repressible operon (*trp* operon)
    - Repressor is initially inactive
    - Corepressor (tryptophan) binds to the repressor and makes it active
  - For many operons, **activators** enhance RNA polymerase binding to the promoter

2: Differentiation results from the expression of different combinations of genes

- **Differentiation** involves cell specialization, in both structure and function

- Differentiation is controlled by turning specific sets of genes on or off

Haematopoiesis
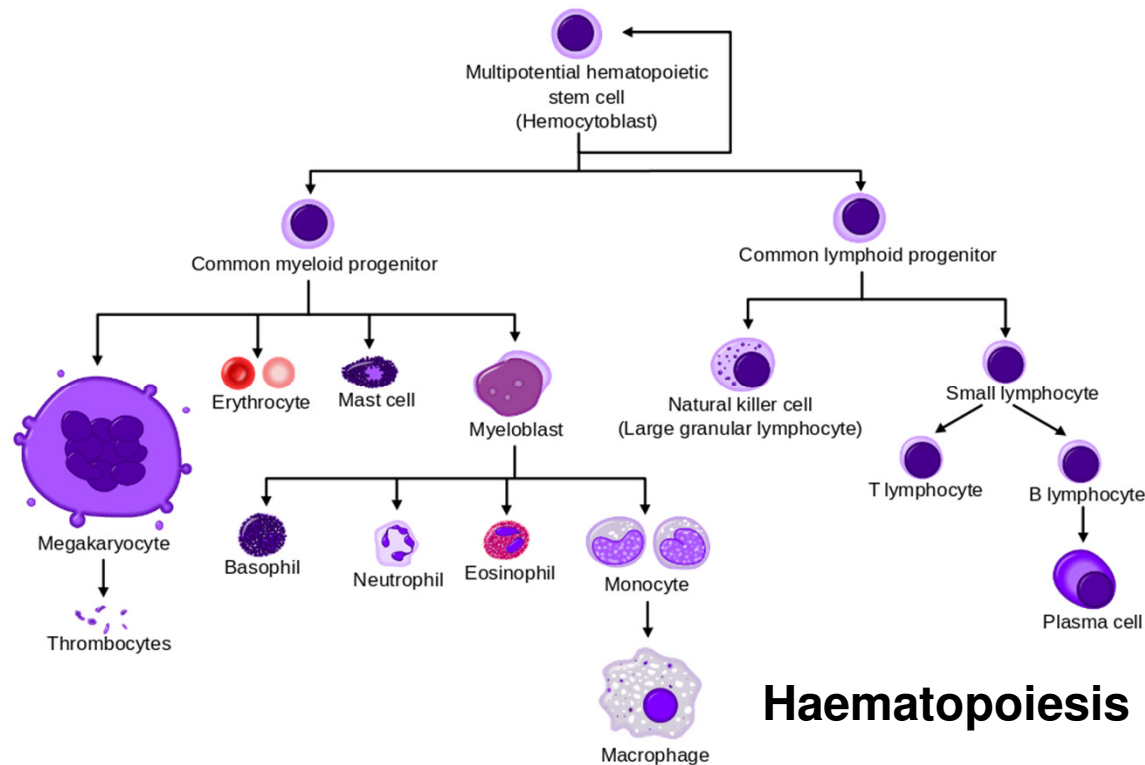
3: DNA packing in eukaryotic chromosomes helps regulate gene expression

- Eukaryotic chromosomes undergo multiple levels of folding and coiling, called DNA packing

EINSTEIN

Albert Einstein College of Medicine
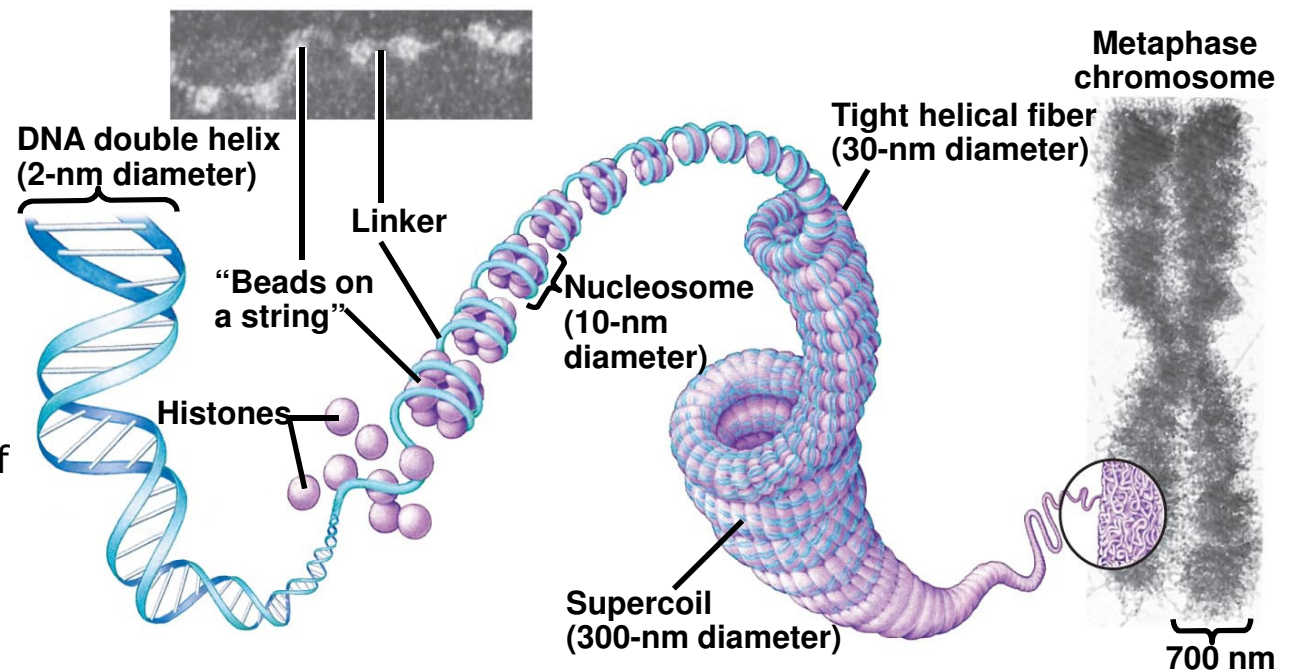OF YESHIVA UNIVERSITY
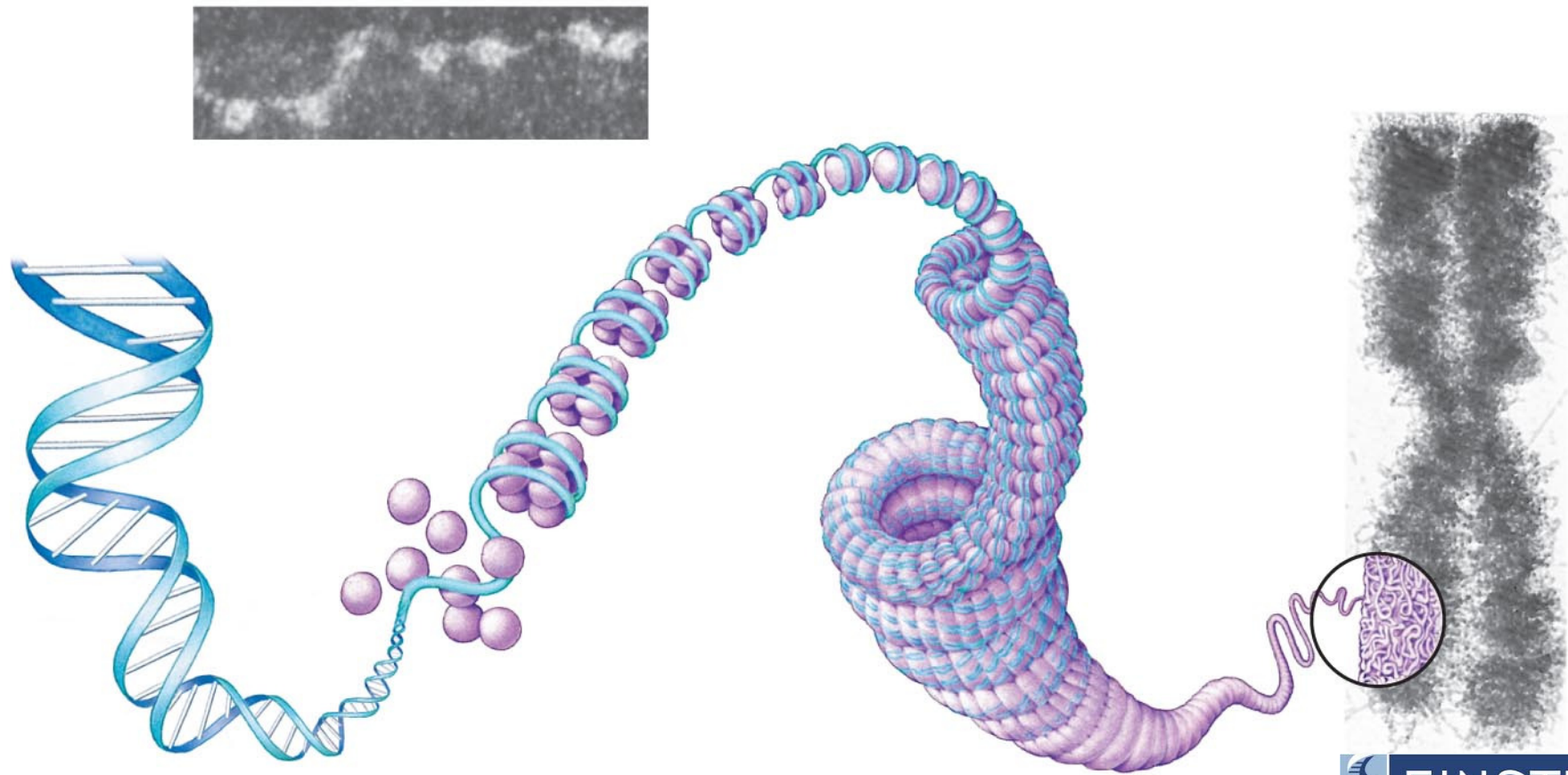
3: DNA packing in eukaryotic chromosomes helps regulate gene expression

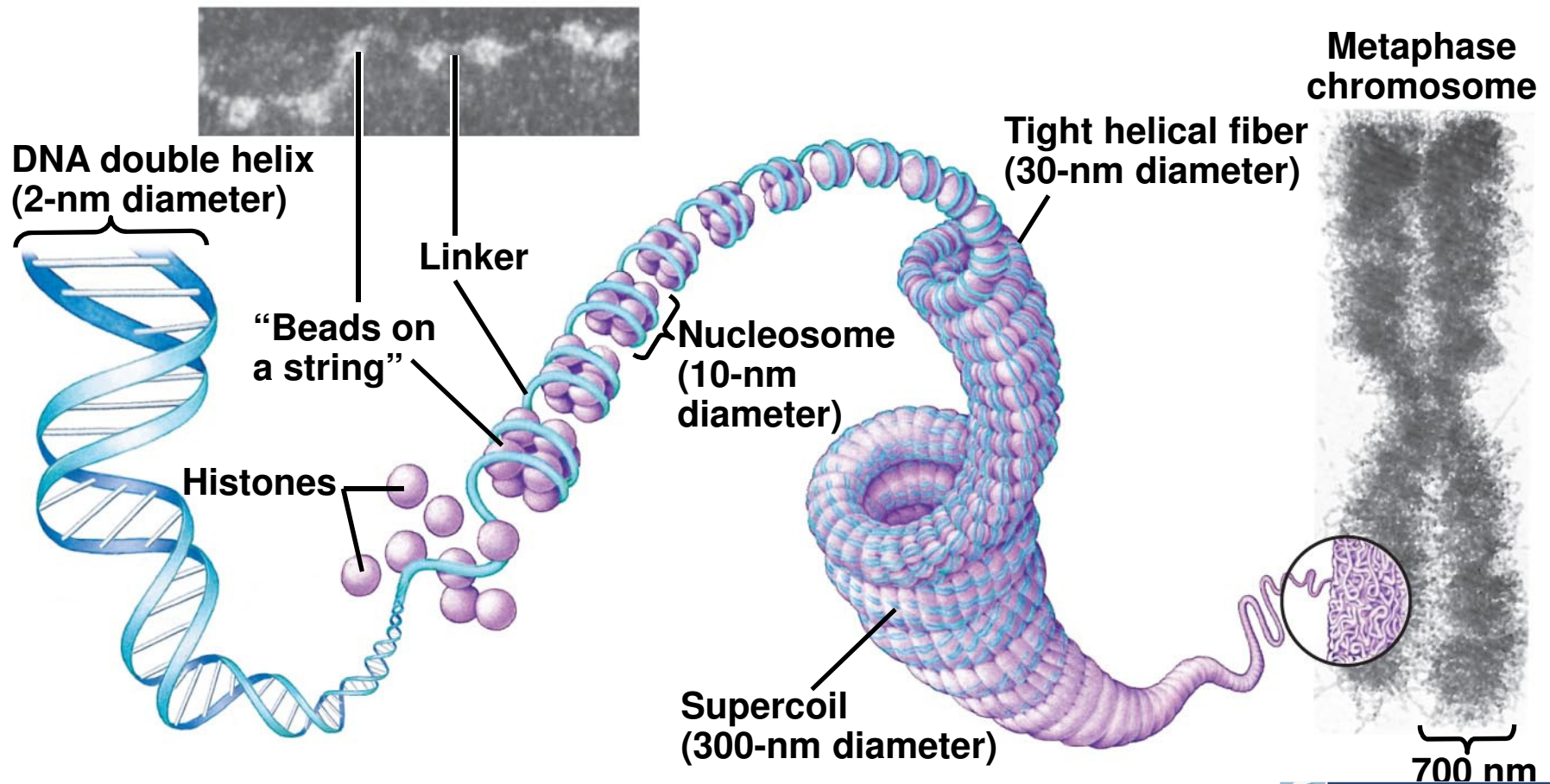- **Nucleosomes** are formed when DNA is wrapped around **histone** proteins
  - "Beads on a string" appearance
  - Each bead includes DNA plus 8 histone molecules
  - String is the linker DNA that connects nucleosomes
- Tight helical fiber is a coiling of the nucleosome string
- Supercoil is a coiling of the tight helical fiber
- Metaphase chromosome represents the highest level of packing

DNA double helix (2-nm diameter)

"Beads on a string"

Histones

Linker

Nucleosome (10-nm diameter)

Supercoil (300-nm diameter)

Tight helical fiber (30-nm diameter)

Metaphase chromosome

700 nm

Copyright © 2009 Pearson Education, Inc.

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

3: DNA packing in eukaryotic chromosomes helps regulate gene expression

- DNA packing can prevent transcription

DNA double helix (2-nm diameter)

"Beads on a string"

Linker

Histones

Nucleosome (10-nm diameter)

Tight helical fiber (30-nm diameter)

Supercoil (300-nm diameter)

Metaphase chromosome

700 nm

EINSTEIN

Albert Einstein College of Medicine
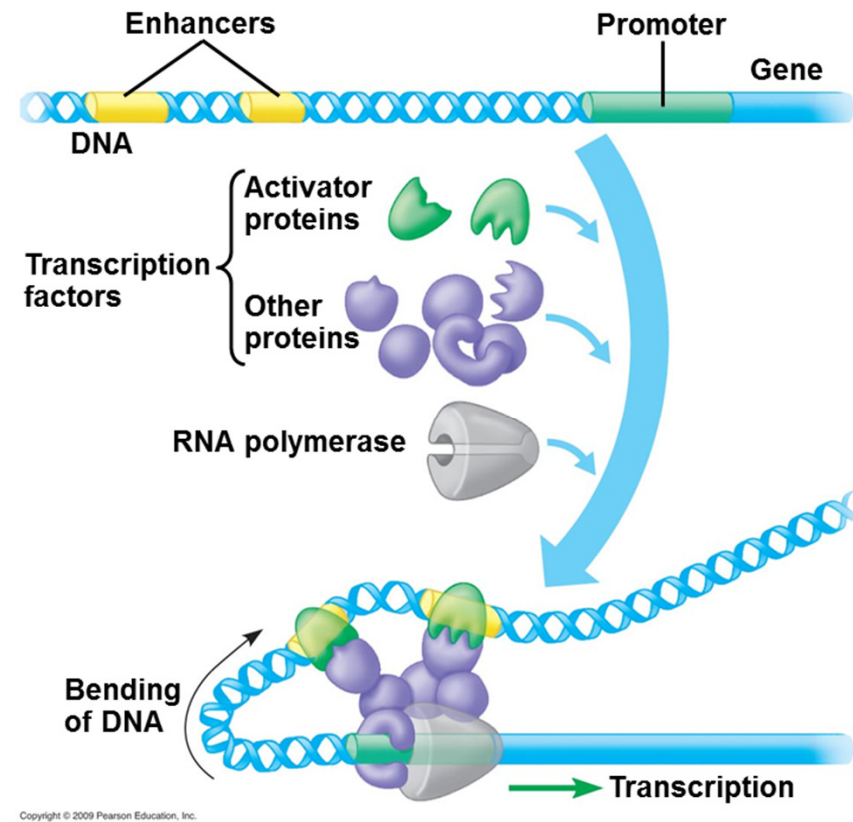OF YESHIVA UNIVERSITY

4: Complex assemblies of proteins control eukaryotic transcription

- Eukaryotic genes
  - Each gene has its own promoter and terminator
  - Are usually switched off and require activators to be turned on
  - Are controlled by interactions between numerous regulatory proteins and control sequences

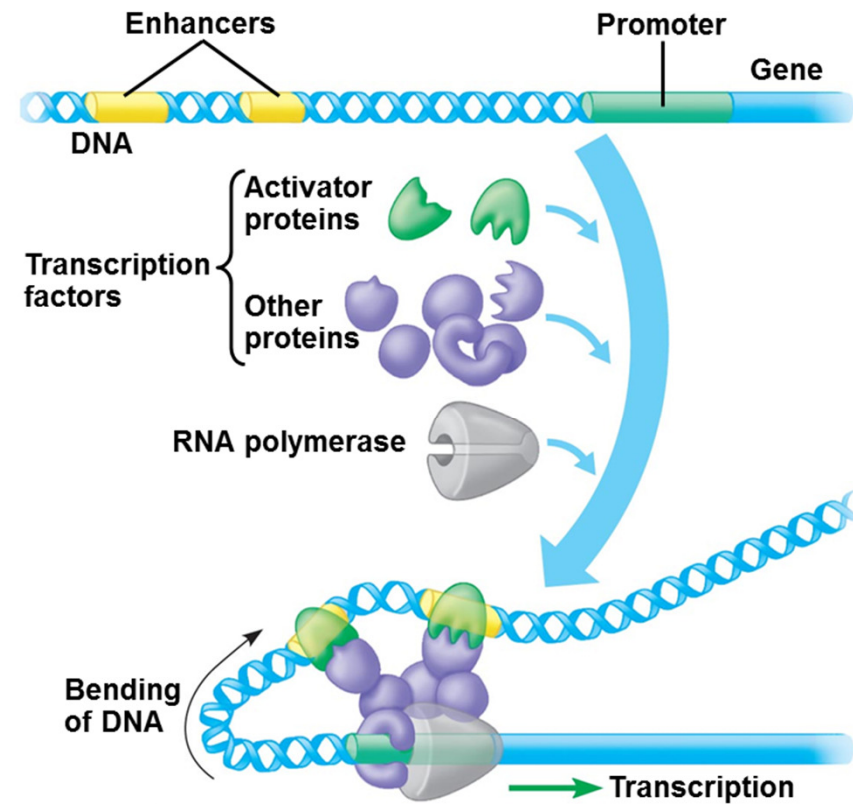# 4: Complex assemblies of proteins control eukaryotic transcription

- Control sequences

  - Promoter

  - Enhancer

    - Related genes located on different chromosomes can be controlled by similar enhancer sequences

4: Complex assemblies of proteins control eukaryotic transcription

- Regulatory proteins that bind to control sequences

  - **Transcription factors** promote RNA polymerase binding to the promoter

  - Activator proteins bind to DNA **enhancers** and interact with other transcription factors

 - **Silencers** are repressors that inhibit transcription

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

5: Eukaryotic RNA may be spliced in more than one way

- **Alternative RNA splicing**

    - Production of different mRNAs from the same transcript

    - Results in production of more than one polypeptide from the same gene

   - Can involve removal of an exon with the introns on either side
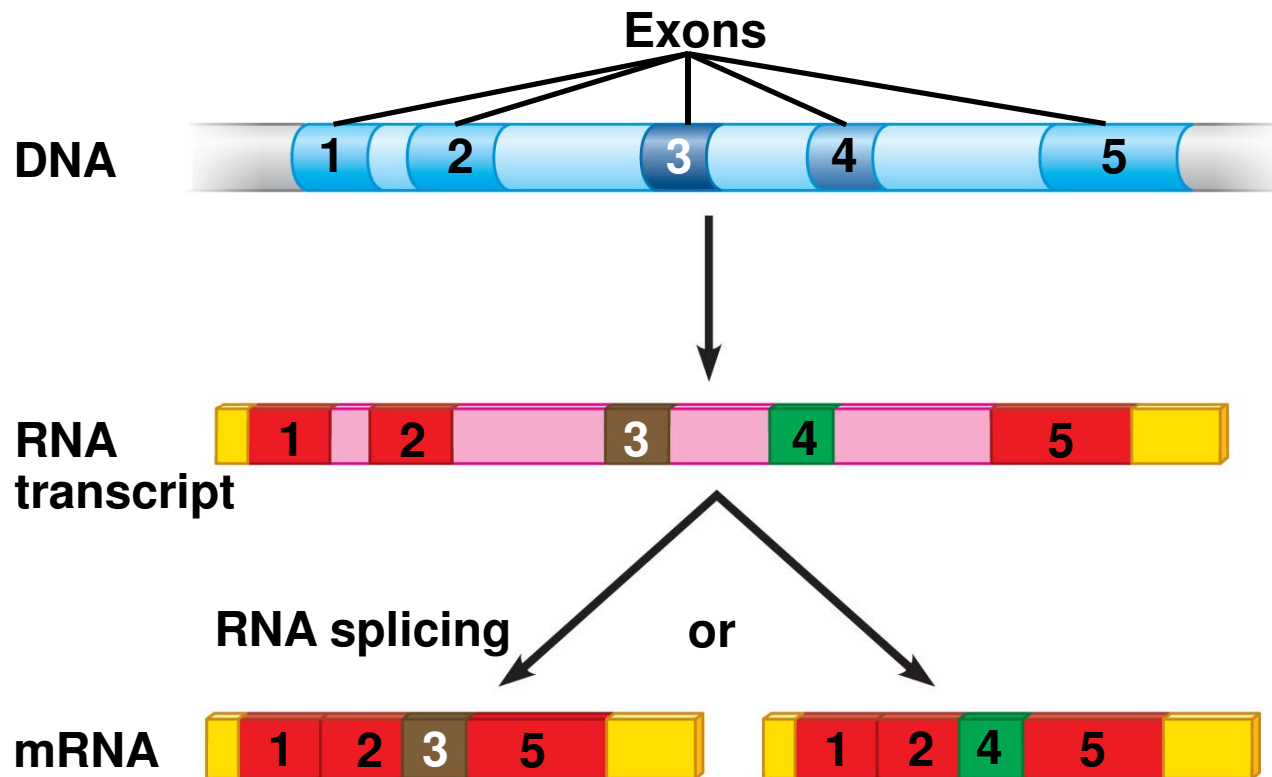
6: Small RNAs play multiple roles in controlling gene expression

- **RNA interference** (**RNAi**)
  - Prevents expression of a gene by interfering with translation of its RNA product
  - Involves binding of small, complementary RNAs to mRNA molecules
  - Leads to degradation of mRNA or inhibition of translation

6: Small RNAs play multiple roles in controlling gene expression
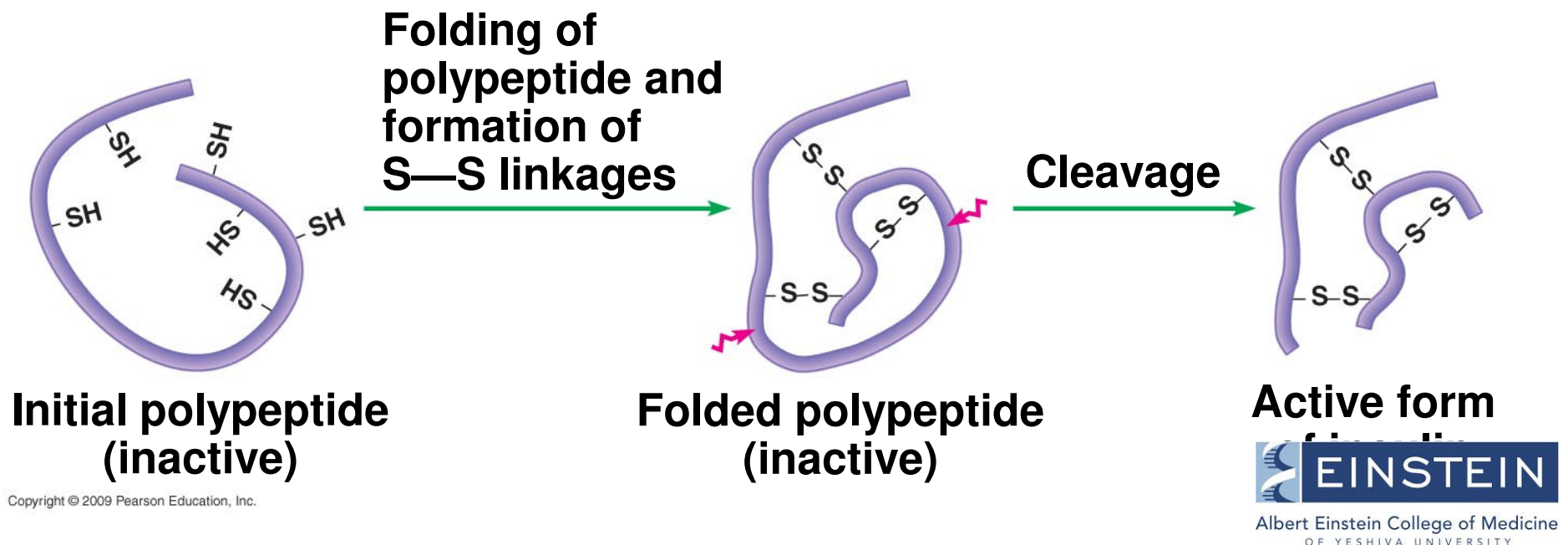
- ## MicroRNA
    - Single-stranded chain about 20 nucleotides long
    - Binds to protein complex
    - MicroRNA + protein complex binds to complementary mRNA to interfere with protein production

miRNA

Protein

❶

miRNA-protein complex

❷

Target mRNA

❸

❹

mRNA degraded   OR   Translation blocked
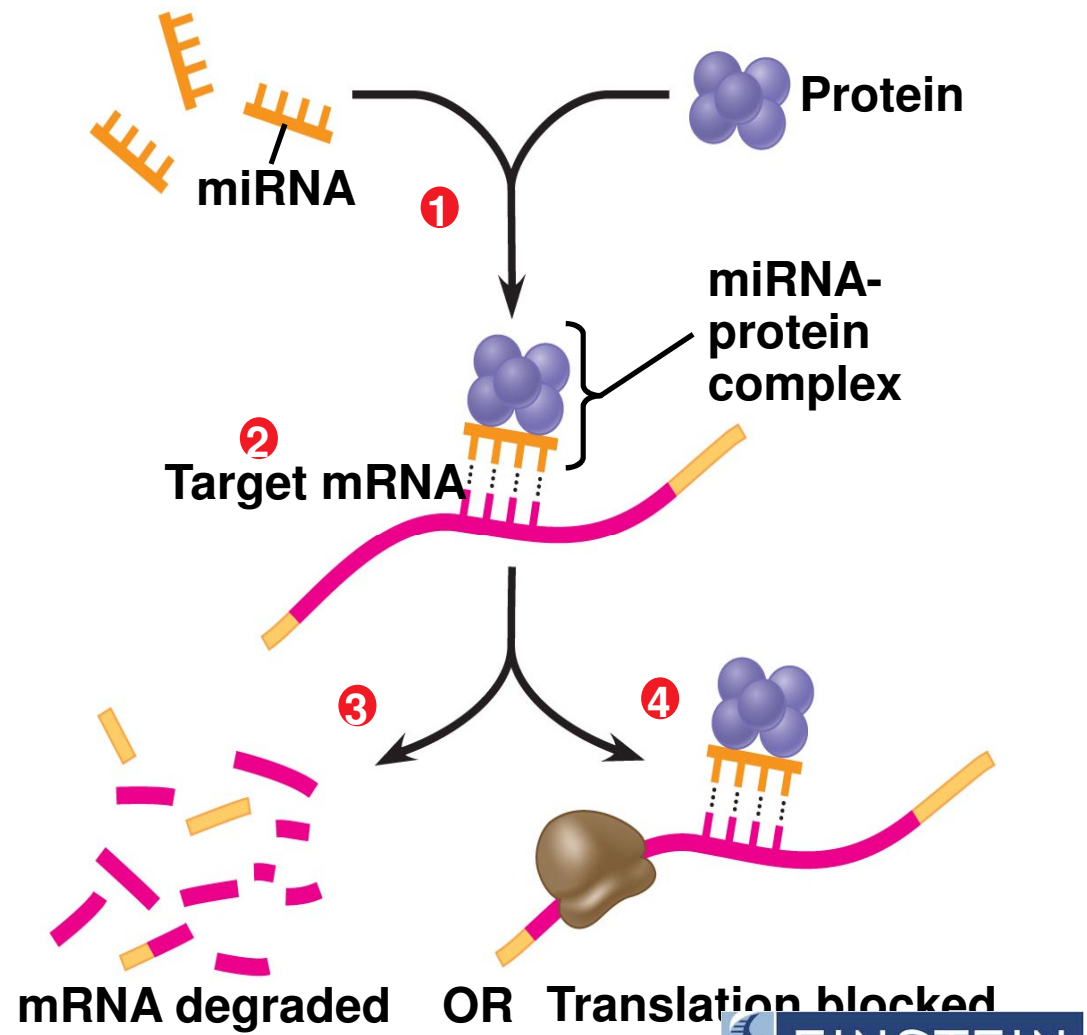
7: Translation and later stages of gene expression are also subject to regulation

- Control of gene expression also occurs with
  - Breakdown of mRNA
  - Initiation of translation
  - Protein activation
  - Protein breakdown

Folding of polypeptide and formation of S—S linkages

Cleavage

Initial polypeptide (inactive)

Folded polypeptide (inactive)

Active form of insulin

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

8: Signal transduction pathways convert messages received at the cell surface to responses within the cell

- **Signal transduction pathway** is a series of molecular changes that converts a signal at the cell's surface to a response within the cell

  - Signal molecule is released by a signaling cell

  - Signal molecule binds to a receptor on the surface of a target cell
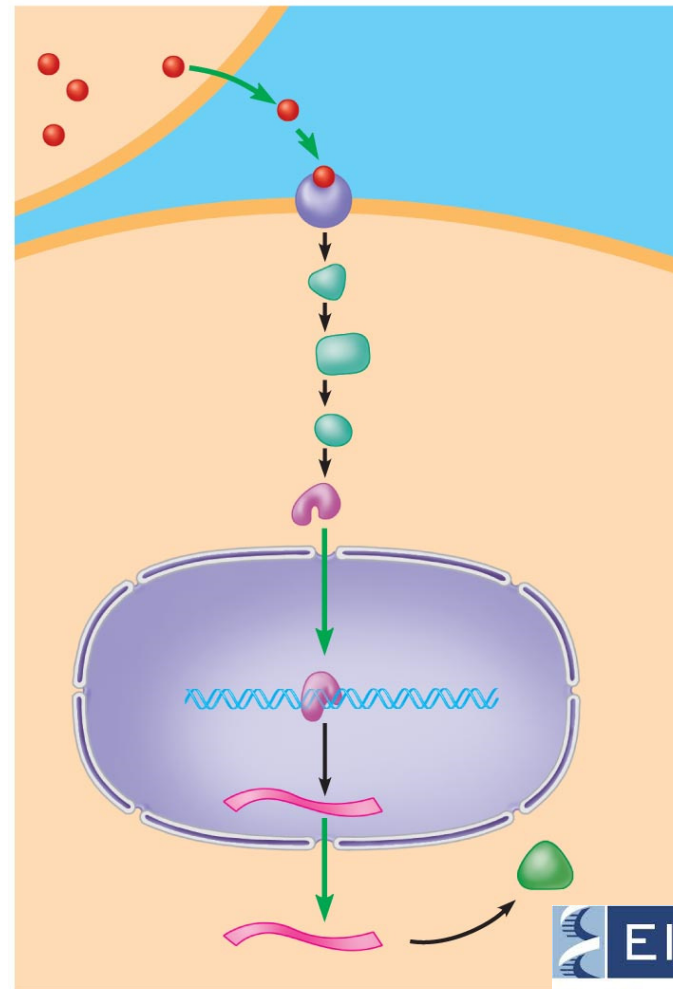
EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

8: Signal transduction pathways convert messages received at the cell surface to responses within the cell

– Related proteins are activated in a series of reactions

– A transcription factor is activated and enters the nucleus

– Specific genes are transcribed to initiate a cellular response
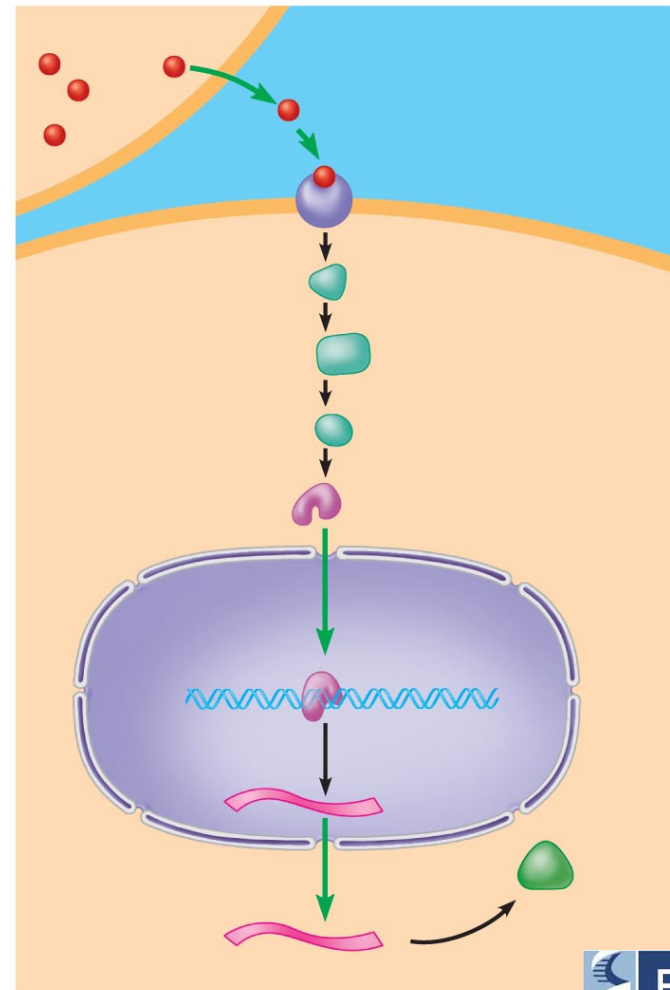
EINSTEIN

Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Review: Multiple mechanisms regulate gene expression in eukaryotes

- Many possible control points exist; a given gene may be subject to only a few of these

  - Chromosome changes

    - DNA unpacking

  - Control of transcription

    - Regulatory proteins and control sequences

  - Control of RNA processing

    - Addition of 5' cap and 3' poly-A tail

    - Splicing

  - Flow through nuclear envelope

Review: Multiple mechanisms regulate gene expression in eukaryotes

- Many possible control points exist; a given gene may be subject to only a few of these

  - Breakdown of mRNA

  - Control of translation

  - Control after translation

    - Cleavage/modification/activation of proteins

    - Breakdown of protein

Outline

- Introduction to gene regulation

- Construction of GRN

- Modeling the dynamics of GRN

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Gene-regulatory networks

What are gene-regulatory networks (GRNs)?

> \- networks between genes coding for transcription factors and
>
> genes they regulate

Gene-regulatory networks

How does one generate GRNs?

    - from co-expression + regulatory information (e.g. presence of

    TF binding sites)

What can these GRNs be used for?

    - functional interpretation of exp. data, guide inhibitor design etc.

Limitations of current GRN models:

    - incomplete in terms of TF-interactions,

    - usually do not account for epigenetic effects and miRNAs

How does one generate GRNs?

...

(1) "by hand" based on individual experimental
    observations

(2) Infer GRNs by computational methods from
    gene expression data

Unsupervised methods

Unsupervised methods are either based on **correlation** or on **mutual information**.

Correlation-based network inference methods assume that correlated expression levels between two genes are indicative of a regulatory interaction.

Correlation coefficients range from -1 to 1.

A positive correlation coefficient indicates an activating interaction, whereas a negative coefficient indicates an inhibitory interaction.

The common correlation measure by **Pearson** is defined as

$$corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sigma(X_i) \cdot \sigma(X_j)}$$

where X_i and X_j are the expression levels of genes i and j, cov(.,.) denotes the covariance, and σ is the standard deviation.

Rank-based unsupervised methods

Pearson's correlation measure assumes normally distributed values.
This assumption does not necessarily hold for gene expression data.

Therefore rank-based measures are frequently used.
The measures by Spearman and Kendall are the most common.

Spearman's method is simply Pearson's correlation coefficient for the ranked expression values

Kendall's τ coefficient :
$$\tau(X_i, X_j) = \frac{con(X_i^r, X_j^r) - dis(X_i^r, X_j^r)}{\frac{1}{2}n(n-1)}$$

where X_i^r and X_j^r are the ranked expression profiles of genes i and j.

Con(.) denotes the number of concordant value pairs (i.e. where the ranks for both elements agree). *dis(.)* is the number of disconcordant value pairs in X_i^r and X_j^r.
Both profiles are of length n.

Unsupervised methods based on mutual information

Relevance networks (RN) introduced by Butte and Kohane measure the **mutual information (MI)** between gene expression profiles to infer interactions.

The MI I between discrete variables X_i and X_j is defined as

$$I(X_i, X_j) = \sum_{x_i \in X_i} \sum_{x_j \in X_j} p(x_i, x_j) \log\left(\frac{p(x_i, x_j)}{p(x_i)p(x_j)}\right)$$

where $p(x_i, x_j)$ is the **joint probability distribution** of X_i and X_j

$p(x_i)$ and $p(x_i)$ are the **marginal probabilities** of the two variables
(ignoring the value of the other one).

X_i and X_j are required to be discrete variables.

Unsupervised methods: Z-score

Z-SCORE is a network inference strategy by Prill et al. that takes advantage of knockout data.

It assumes that a knockout affects directly interacting genes more strongly than others.

The z-score z_{ij} describes the effect of a knockout of gene i in the k-th experiment on gene j as the normalized deviation of the expression level X_{jk} of gene j for experiment k from the average expression $\mu(X_j)$ of gene j:

$$z_{ij} = \left| \frac{X_{jk} - \mu(X_j)}{\sigma(X_j)} \right|$$

supervised inference method: SVM

In contrast to unsupervised methods, e.g. correlation methods, the supervised approach does not directly operate on pairs of expression profiles but on feature vectors that can be constructed in various ways.

E.g. one may use the outer product of two gene expression profiles X_i and X_j to construct feature vectors:

$$\mathbf{x} = X_i X_j^T$$

A sample set for the training of the SVM is then composed of feature vectors $\mathbf{x}_i$ that are labeled $\gamma_i = +1$ for gene pairs that interact and $\gamma_i = -1$ for those that do not interact.

Measure accuracy of GRNs

Inference methods aim to recreate the topology of a genetic regulatory network e.g. based on expression data only.

The **accuracy** of a method is assessed by the extent to which the network it infers is similar to the true regulatory network.

We quantify similarity e.g. by the area under the Receiver Operator Characteristic curve (AUC)

$$AUC = \frac{1}{2} \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k-1})$$

where X_k is the false-positive rate and Y_k is the true positive rate for the k-th output in the ranked list of predicted edge weights.

An AUC of 1.0 indicates a perfect prediction, while an AUC of 0.5 indicates a performance no better than random predictions.

Summary

Network inference is a very important active research field.

Inference methods allow to construct the topologies of gene-regulatory networks solely from expression data (unsupervised methods).

Supervised methods show far better performance.

Performance on real data is lower than on synthetic data because regulation in cells is not only due to interaction of TFs with genes, but also depends on epigenetic effects (DNA methylation, chromatin structure/histone modifications, and miRNAs).

Outline

- Introduction to gene regulation

- Construction of GRN

- Modeling the dynamics of GRN

Reconstructed GRN is not static, but dynamic

The dynamics of GRN is reflected by motifs

Motifs are functional related

Network Motif 1: Feed-Forward-Loop
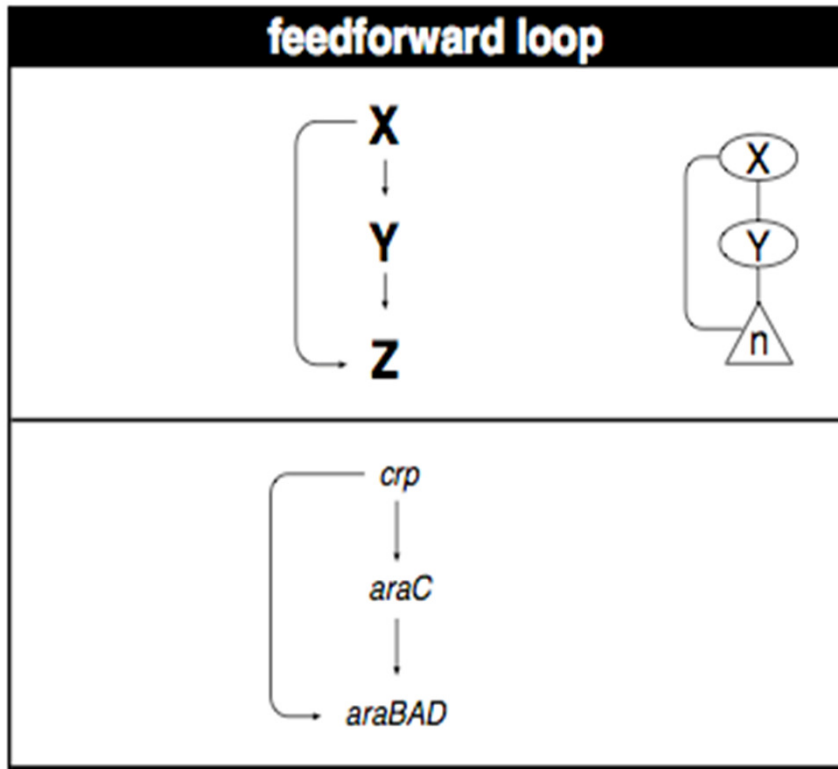
X = general transcription factor
Y = specific transcription factor
Z = effector operon(s)

X and Y **together** regulate Z:

"**coherent**", if X and Y have the **same** effect on Z (activation vs. repression), otherwise "incoherent"

85% of the FFL in E coli are coherent

FFL dynamics
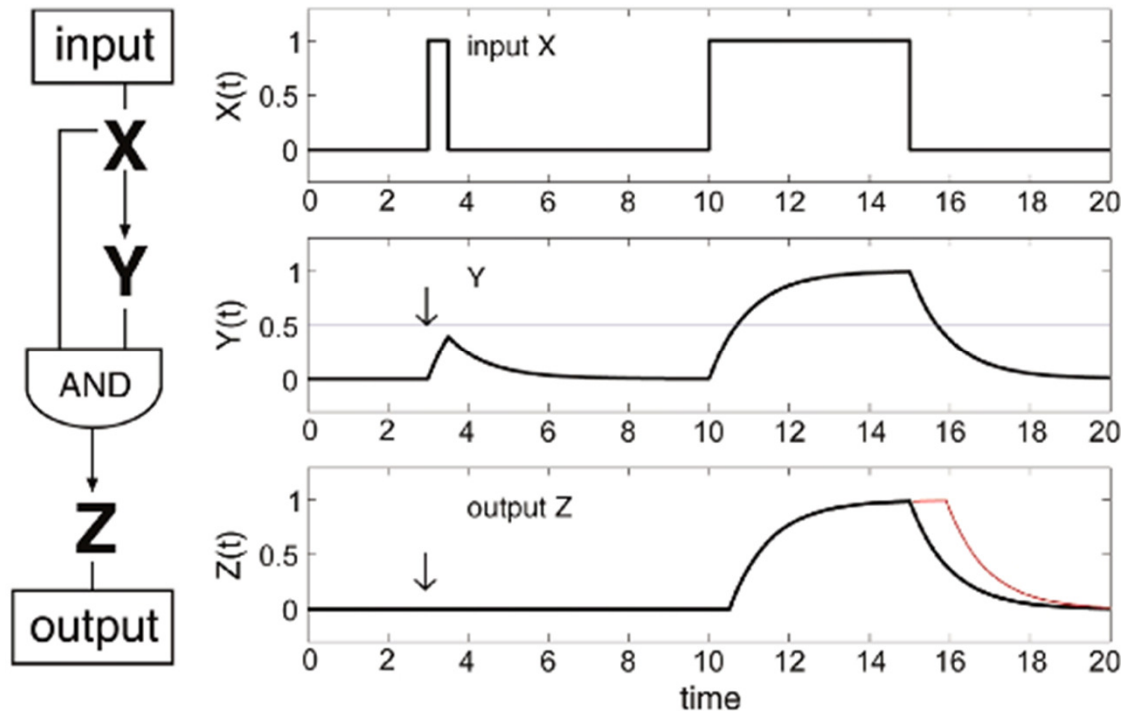
In a coherent FFL:
X **and** Y activate Z

Dynamics:
• input activates X
• X activates Y (delay)
• (X && Y) activates Z

Delay between X and Y → signal must persist longer than delay
→ reject transient signal, react only to **persistent** signals
→ fast shutdown

Helps with **decisions** based on **fluctuating signals**

Network Motif 2: Single-Input-Module
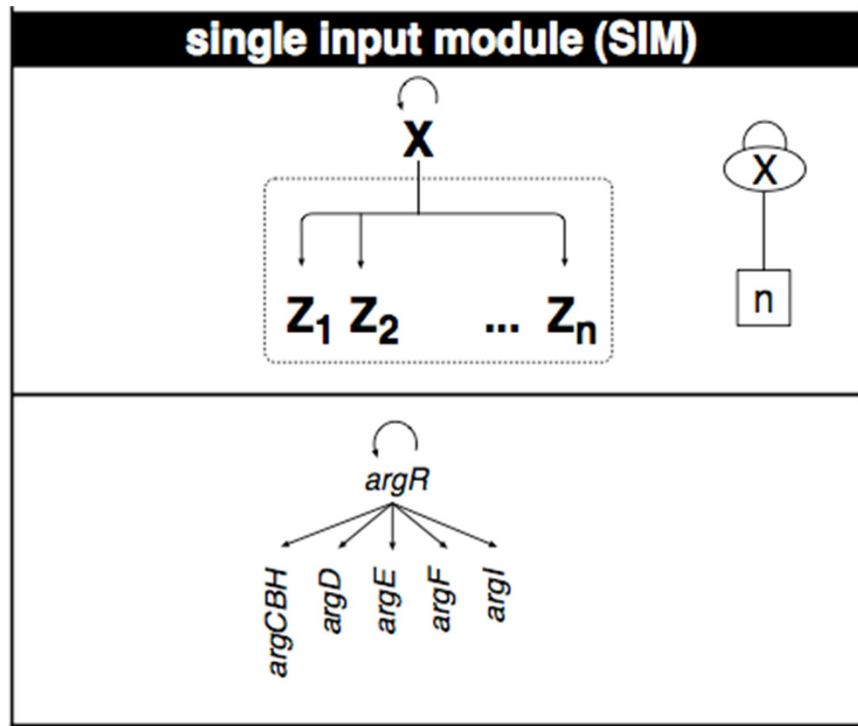
Set of operons controlled by a single transcription factor
- same sign
- no additional regulation
- control usually autoregulatory

Mainly found in genes that code for **parts** of a protein **complex** or metabolic **pathway**
→ relative stoichiometries

SIM-Dynamics
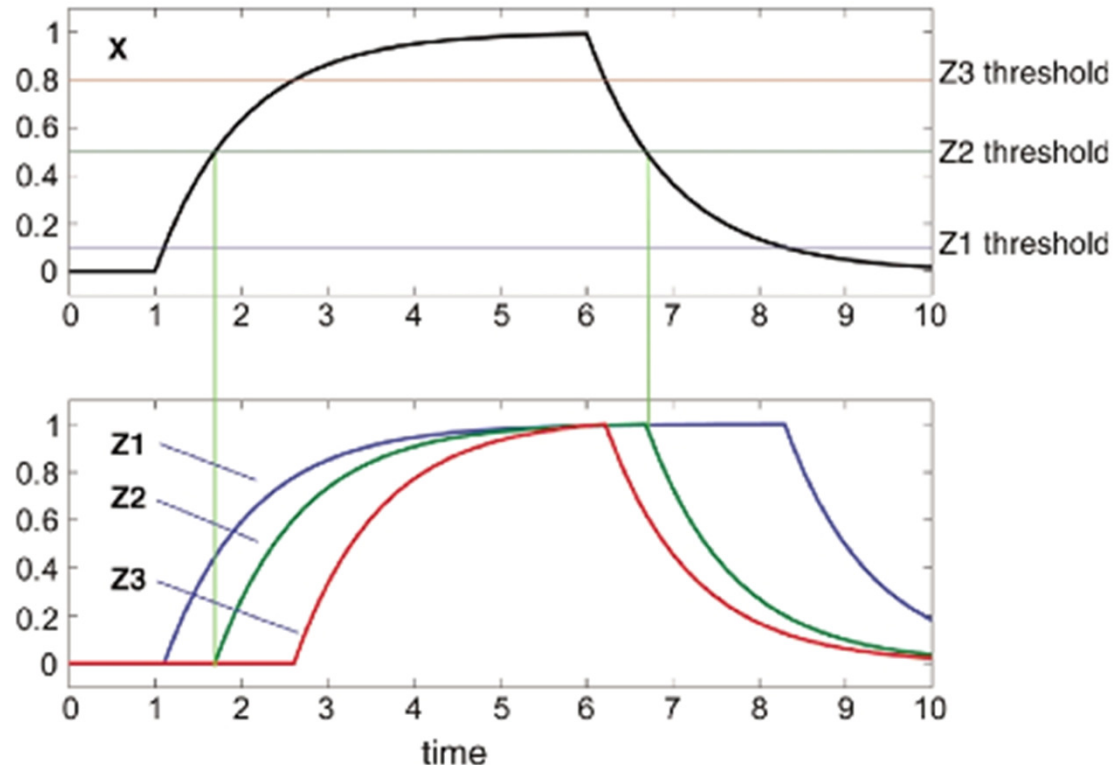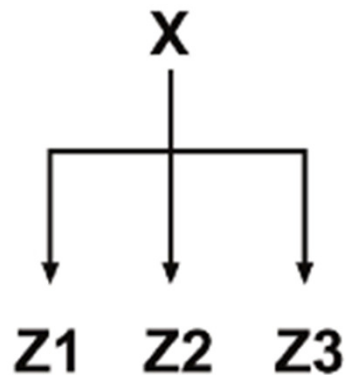
With different thresholds for each regulated operon:
→ first gene that is activated is the last that is deactivated
  → well defined temporal ordering (e.g. flagella synthesis) + stoichiometries

Densely Overlapping Motifs in Gene Regulation

dense overlapping regulons (DOR)
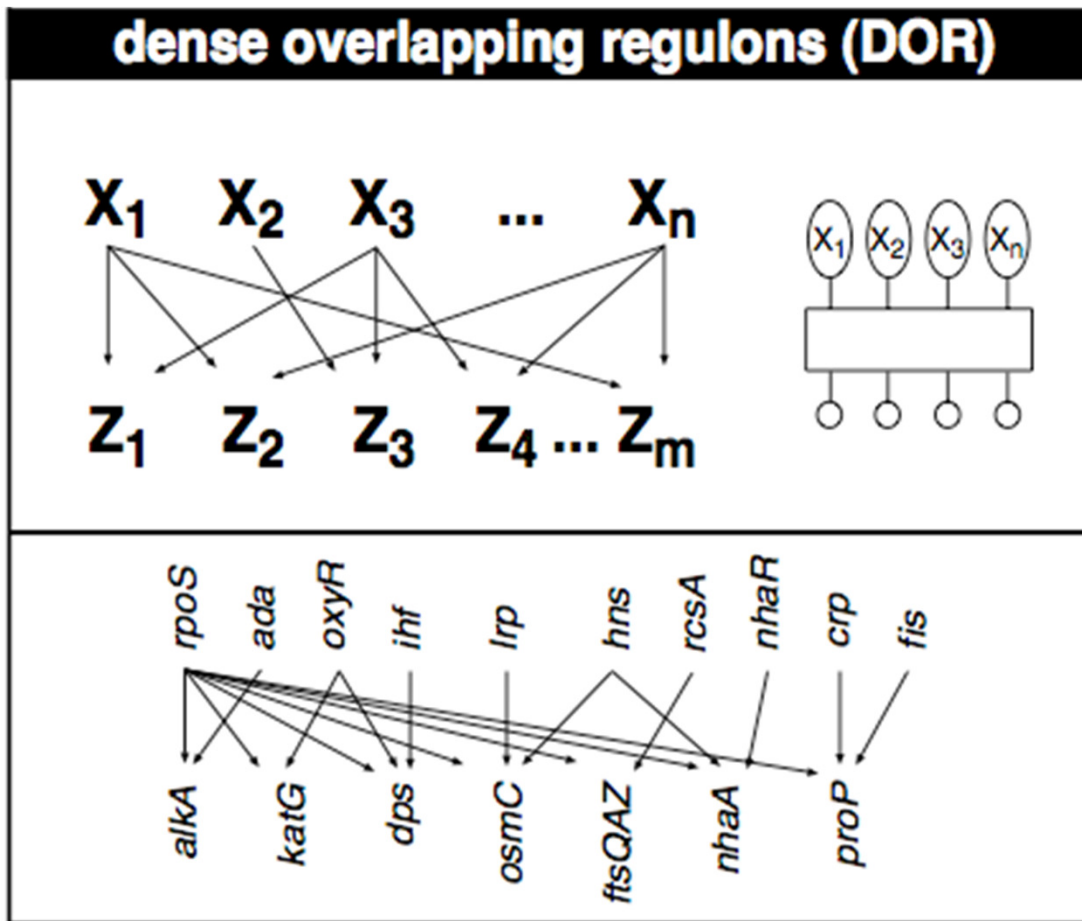
Dense layer between groups of transcription factors and operons
→ much denser than network average

Usually each operon is regulated by a different combination of TFs.

Main "**computational**" **units** of the regulation system

How can we study the dynamics of GRN

One simplest model is using Boolean Network

What is Boolean Networks

Conditional transitions

- "If LuxI is present, then AI will be produced…"

- "If there is AI and there's no LuxR:AI bound to the genome, then LuxR will be expressed and complexes can form…"
- "If LuxR:AI is bound to the genome, then LuxI is expressed…"

Simplified mathematical **description** of the dependencies:

Densities of the species   <=>   discrete states: on/off, 1/0

Network of dependencies <=>   condition tables

Progress in time   <=>   discrete propagation steps

Boolean Networks II

State of the system: described by **vector** of **discrete** values

$$S_i = \{0, 1, 1, 0, 0, 1, \ldots\}$$

$$S_i = \{x_1(i),\ x_2(i),\ x_3(i),\ \ldots\}$$

fixed number of species with **finite number** of states each

$\rightarrow$ finite number of system states

Propagation:

$$S_{i+1} = \{x_1(i+1),\ x_2(i+1),\ x_3(i+1),\ \ldots\}$$

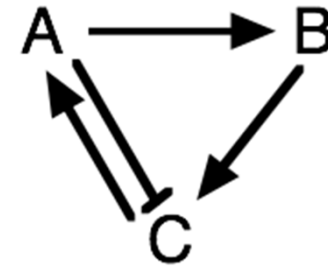$x_1(i+1) = f_1(x_1(i), x_2(i), x_3(i), \ldots)$ with f_i given by condition tables

Propagation Trajectories:

$\rightarrow$ periodic trajectories

$\rightarrow$ **periodic** sequence of states = **attractor**

$\rightarrow$ all states leading to an attractor = **basin of attraction**

A Small Example

State vector $S = \{A, B, C\} \rightarrow 8$ possible states

Conditional evolution:

A is on if C is on

A_{i+1}	C_i
0	0
1	1

A activates B

B_{i+1}	A_i
0	0
1	1

C is on if (B is on && A is off)

C_{i+1}	A_i	B_i
0	0	0
1	0	1
0	1	0
0	1	1

Start from $\{A, B, C\} = \{1, 0, 0\}$

#	S_i	A	B	C
0	S_0	1	0	0
1	S_1	0	1	0
2	S_2	0	0	1
3	$S_3 = S_0$	1	0	0

assume that inhibition through A is stronger than activation via B

periodic orbit of length 3

Test the Other States

Test the other states

#	A	B	C
0	1	1	1
1	1	1	0
2	0	1	0
3	0	0	1
4	1	0	0
5	0	1	0

A_{i+1}	C_i
0	0
1	1

B_{i+1}	A_i
0	0
1	1

C_{i+1}	A_i	B_i
0	0	0
1	0	1
0	1	0
0	1	1

#	A	B	C
0	1	0	1
1	1	1	0
2	0	1	0

#	A	B	C
0	0	1	1
1	1	0	1

#	A	B	C
0	0	0	0
1	0	0	0

Same attractor as before:
100 → 010 → 001 → 100

also reached from:
110, 111, 101, 011
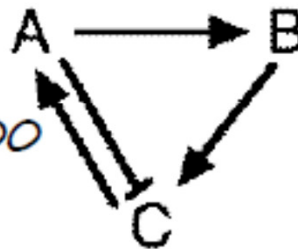
→ **Either all off or stable oscillations**

A Knock-out Mutant
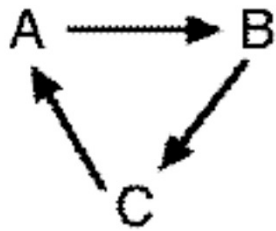
A_{i+1}	C_i
0	0
1	1

B_{i+1}	A_i
0	0
1	1

C_{i+1}	B_i
0	0
1	1

Attractors:

#	A	B	C
0	1	0	0
1	0	1	0
2	0	0	1
3	1	0	0

#	A	B	C
0	1	1	0
1	0	1	1
2	1	0	1
3	1	1	0

#	A	B	C
0	1	1	1
1	1	1	1

#	A	B	C
0	0	0	0
1	0	0	0

no feedback
→ no stabilization, network just "rotates"

One specific example: Quorum Sensing in bacterial
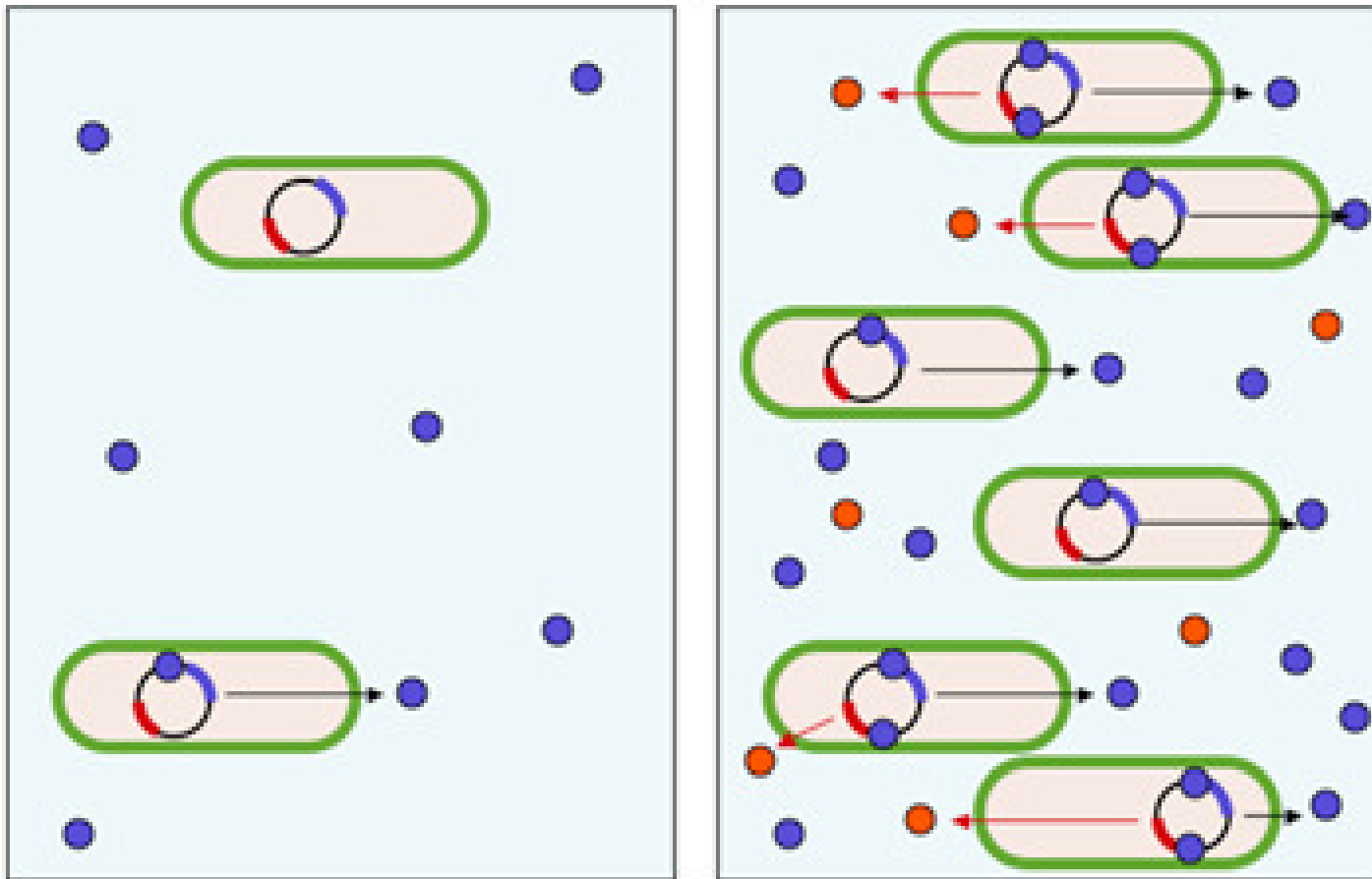
Gene regulation in Quorum Sensing
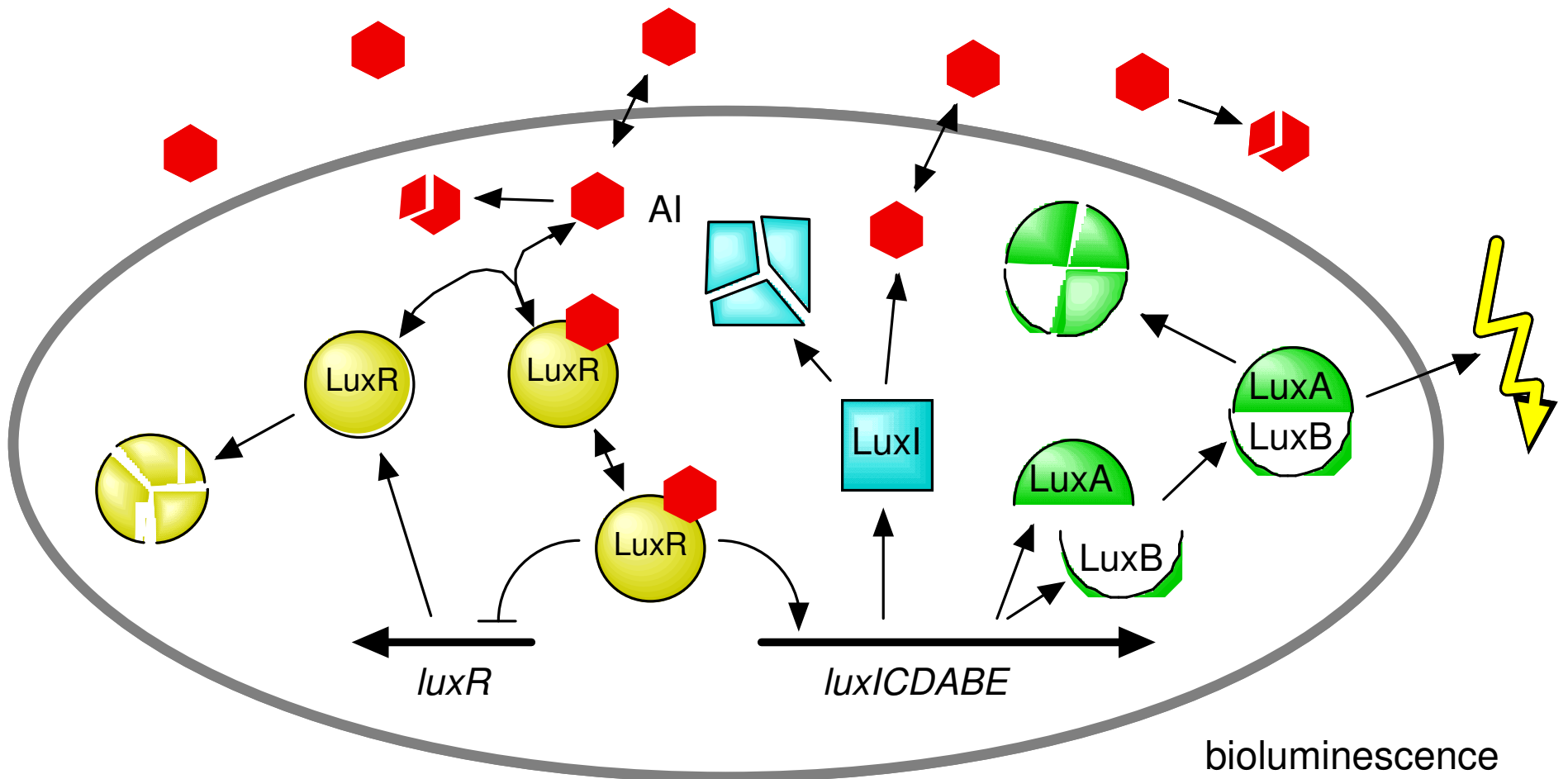
bioluminescence

Boolean Network of Quorum Sensing
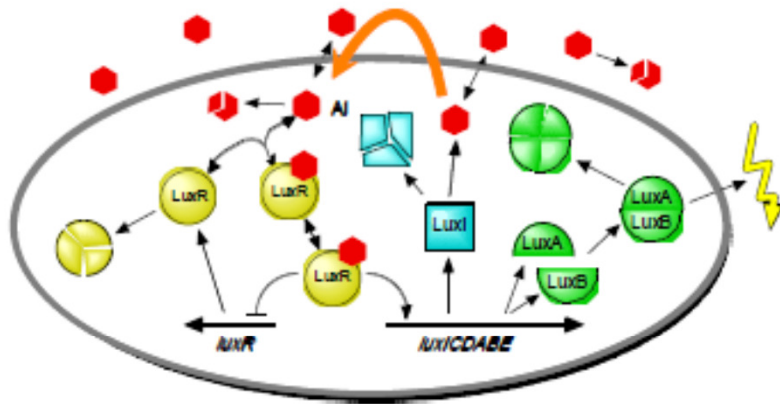
Minimum set of species:

LuxR, AI, LuxR:AI,
LuxR:AI:genome, LuxI
Here: Light signal (LuxAB) α LuxI

Condition tables: describe the state of a species in the next step given the current states of all relevant species.

LuxI	LuxR:AI:Genome
0	0
1	1

How does LuxI depend on LuxR:AI:Genome?

LuxR:AI:Genome	LuxR:AI
0	0
1	1

How does LuxR:AI:Genome depend on LuxR:AI?

Condition Tables for QS II
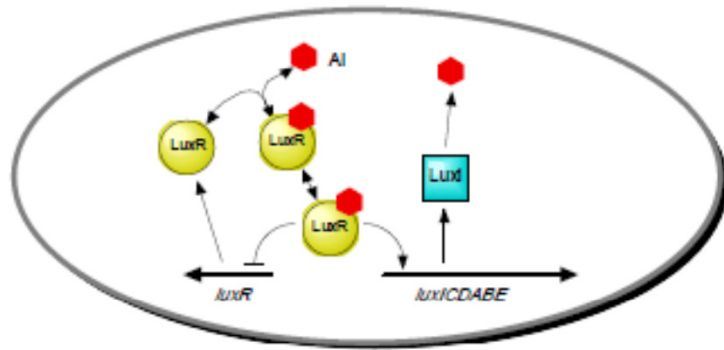
Lux R	LuxR	AI	LuxR:AI:Genome
1	0	0	0
1	1	0	0
1	0	1	0
1	1	1	0
0	0	0	1
1	1	0	1
0	0	1	1
0	1	1	1

LuxR:AI	Lux R	AI	LuxR:AI:Genome
0	0	0	0
0	1	0	0
1	0	1	0
1	1	1	0
0	0	0	1
0	1	0	1
0	0	1	1
1	1	1	1

$\rightarrow$

LuxR:AI	LuxR	AI	LuxR:AI:Genome
0	X	0	X
1	1	1	X
1	0	1	0
0	0	1	1

Note: no dissociation
(LuxR:AI:Genome → LuxR:AI + Genome)
only degradation of AI
LuxR:AI:Genome → LuxR

Condition Tables for QS III
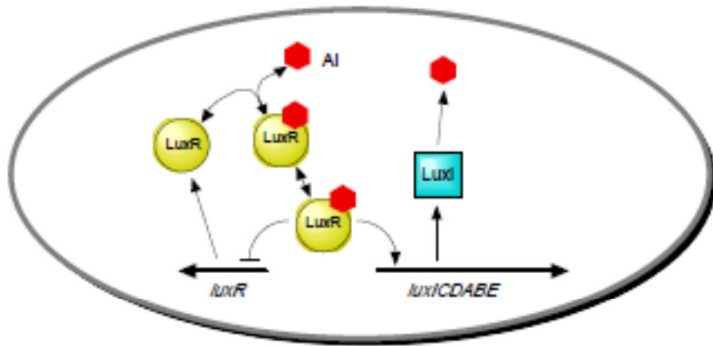
AI	LuxR	AI	LuxI
0	0	0	0
0	1	0	0
1	0	1	0
0	1	1	0
1	0	0	1
1	1	0	1
1	0	1	1
1	1	1	1

→

AI	LuxR	AI	LuxI
1	x	x	1
0	x	0	0
1	0	1	0
0	1	1	0
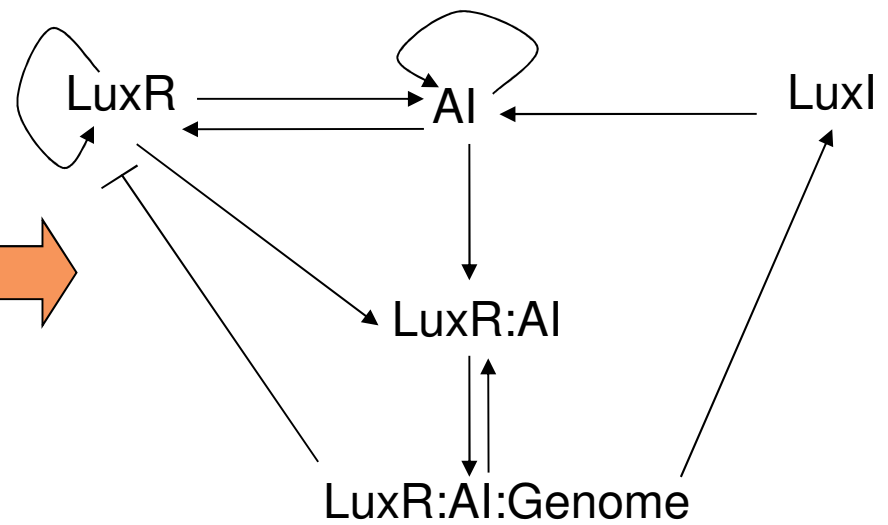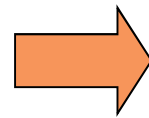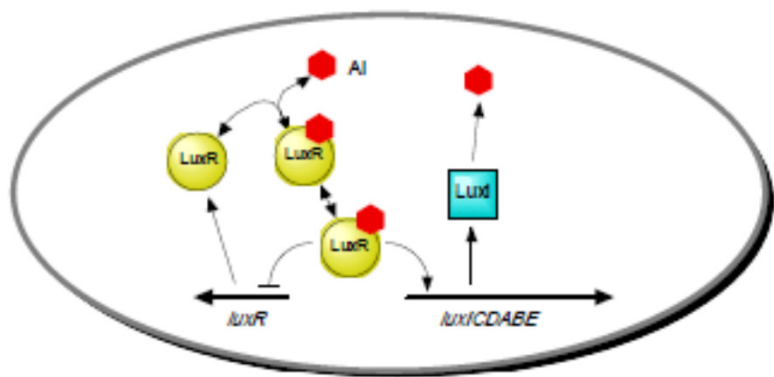
Scanning for Attractors

States of *V. fischeri* QS system mapped onto integers

{LuxR (LR), LuxR:AI (RA), AI, LuxR:AI:Genome (RAG), LuxI (LI)}
$$= \{1, \ 2, \ 4, \ 8, \ 16\}$$

For each **attractor**:
- periodic orbit and its length (period)
- basin of attraction and its relative size (32 states in total)
  → how likely will the system end in each of the attractors?

Attractor 1: orbit: 1   → period 1

states: 0, 1 → size 2,   2/32 = 6.25 %

start from state 0:
```
 #   LR  RA  AI  RAG LI - state
 0   .   .   .   .   . -  0
 1   X   .   .   .   . -  1        <=
 2   X   .   .   .   . -  1        attractor
```

Attractor 2:   orbit: 3, 9, 17, 5   → period 4

states: 2, 3, 5, 8, 9, 16, ~~17~~ size 7,   21.9 %

start from state 8:

#	LR	RA	AI	RAG	LI	-	state
0	.	.	.	X	.	-	8
1	.	.	.	.	X	-	16
2	X	.	X	.	.	-	5
3	X	X	.	.	.	-	3
4	X	.	.	X	.	-	9
5	X	.	.	.	X	-	17
6	X	.	X	.	.	-	5

attractor

averaged occupancies in this periodic orbit:

LR	RA	AI	RAG	LI
4/4 = 1	1/4 = 0.25	1/4 = 0.25	1/4 = 0.25	1/4 = 0.25

Attractors III

Attractor 3:   period 4, basin of 16 states → 50 %

```
#   LR  RA  AI  RAG LI – state0
.   X   X   .   .   –  6:
.   X   X   X   .   – 14.
.   .   X   X   X   – 28:
.   .   X   .   X   – 20
```

Attractor 4:   period 4, basin of 4 states → 12.5 %

```
#   LR  RA  AI  RAG LI – state0
X   X   X   .   .   –  7:
X   X   .   X   .   – 11:
X   .   .   X   X   – 25:
X   .   X   .   X   – 21:
```

Attractor 5:   period 2, basin of 3 states → 9.4 %

```
#   LR  RA  AI  RAG LI – state0
X   .   X   X   .   – 13:
.   X   .   .   X   – 18
```

Classifying the Attractors

→ Interpret the system's behavior from the properties of the attractors

Attractor	period	basin size	<LuxR>	<LuxR:AI>	<AI>	<LuxR:AI:Gen>	<LuxI>
1	1	6.25 % (2)	1	0	0	0	0
2	4	21.9% (7)	1	0.25	0.25	0.25	0.25
3	4	50 % (16)	0	0.5	1	0.5	0.5
4	4	12.5 % (4)	1	0.5	0.5	0.5	0.5
5	2	9.4% (3)	0.5	0.5	0.5	0.5	0.5

Three **regimes**:

dark: LuxI = 0   **intermediate**: LuxI = 0.25   **bright**: LuxI = 0.5

free LuxR, no AI   free LuxR + little AI   little free LuxR (0.24) + much AI (0.85)

Summary and limitations of Boolean Netowk Model

Generally: → quality of the **results** depends on the quality of the **model**

→ quality of the model depends on the quality of the **assumptions**

Assumptions for the Boolean network description:

- subset of the species considered → reduced system state space)

- only discrete density levels → dynamic balances lost,
  reduced to oscillations

- conditional yes–no causality → no continuous processes

- discretized propagation steps → timing of concurrent paths?

"You get what you pay for"

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Petri-Nets: More Accurate
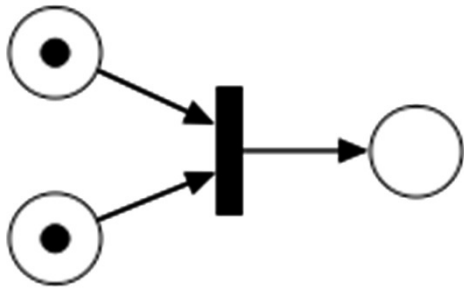
Bipartite **graph** of

- places
- transitions

two types of nodes

- directed weighted arcs
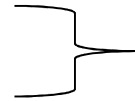
Places: have a capacity (1 … ∞)
→ max. number of tokens   (default: ∞)

Arcs: have costs (1 … ∞)
→ number of tokens that are consumed/produced (default: 1)

Transitions: can fire, when the conditions are fulfilled
→ enough tokens on the in-places: ≥ costs for in-arcs
→ enough remaining capacity on the out-places: ≥ costs for out-arcs
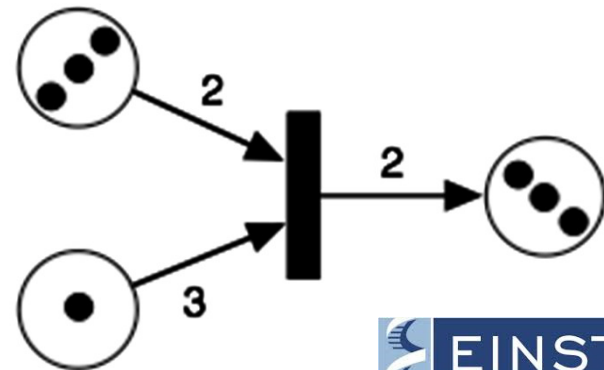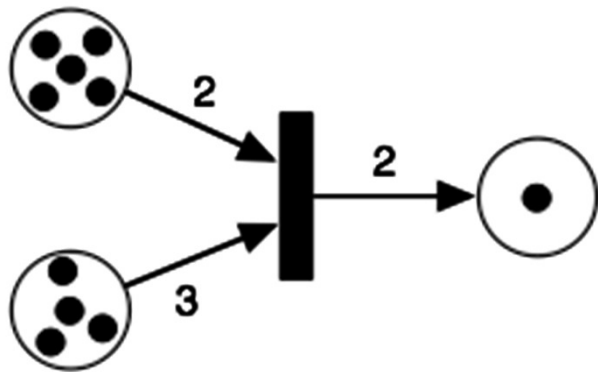
Multiple Possibilities
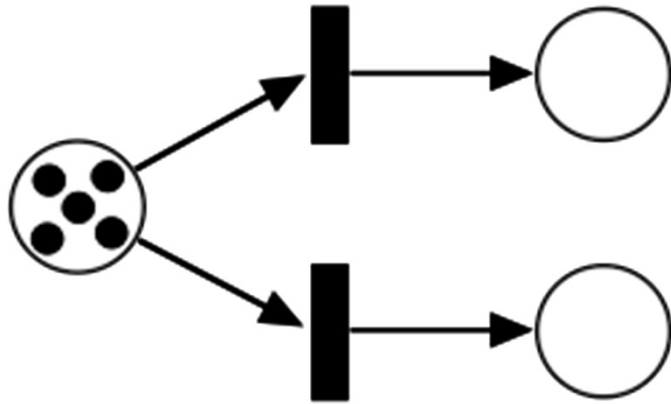
When **multiple transitions** may fire:

- all are equal
  → choose one randomly

- if priorities are defined
  → transition with highest priority fires

Study GRN dynamics by differential equation models

Simple example 1:

Study GRN dynamics by differential equation models

Simple example 1:

Model of negative feedback system
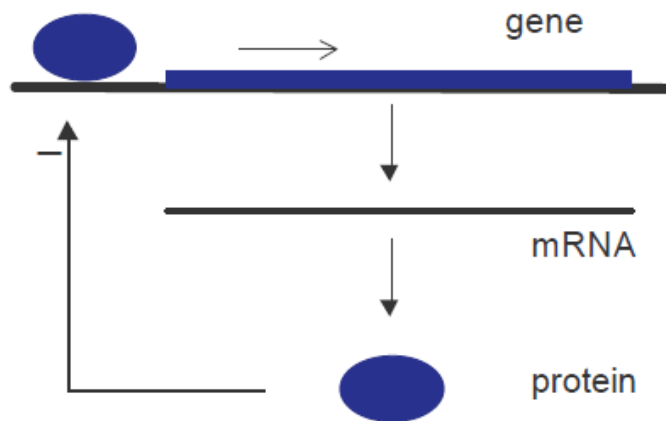
x_1 = mRNA concentration
x_2 = protein concentration

$$\dot{x}_1 = \kappa_1 f(x_2) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 x_1 - \gamma_2 x_2$$

κ_1, $\kappa_2 > 0$, production rate constants
γ_1, $\gamma_2 > 0$, degradation rate constants
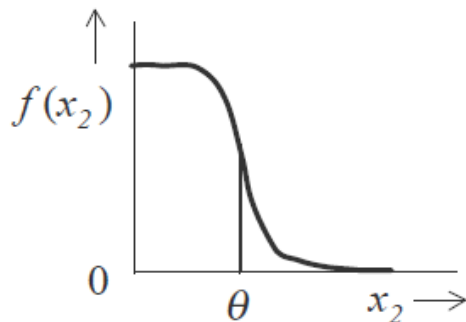
$$f(x_2) = \frac{\theta^n}{\theta^n + x_2^n} \quad , \quad \theta > 0 \text{ threshold}$$

Study GRN dynamics by differential equation models

Simple example 1:

Steady state analysis

❖ No analytical solution of nonlinear differential equations describing feedback system

❖ System has single **steady state** at $\dot{x} = 0$
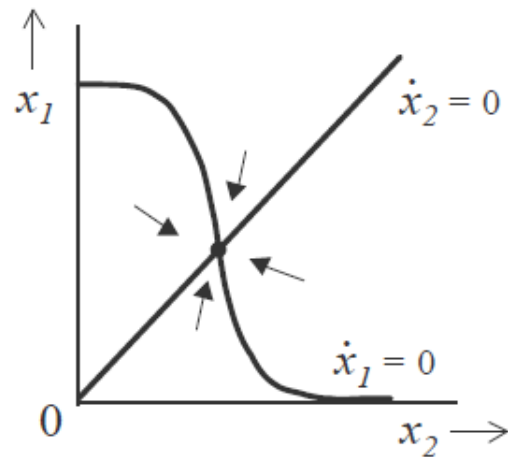
$$\dot{x}_1 = 0 : \quad x_1 = \frac{\kappa_1}{\gamma_1} f(x_2)$$

$$\dot{x}_2 = 0 : \quad x_1 = \frac{\gamma_2}{\kappa_2} x_2$$

❖ Steady state is **stable**, that is, after perturbation system will return to steady state (homeostasis)

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Study GRN dynamics by differential equation models

Simple example 1:

Transient behavior after pertubation

❖ Numerical simulation of differential equations shows **transient behavior** towards steady state after perturbation

Initial values $x_1 (0)$, $x_2 (0)$ correspond to perturbation
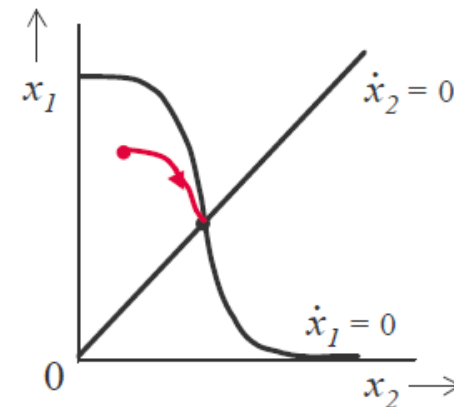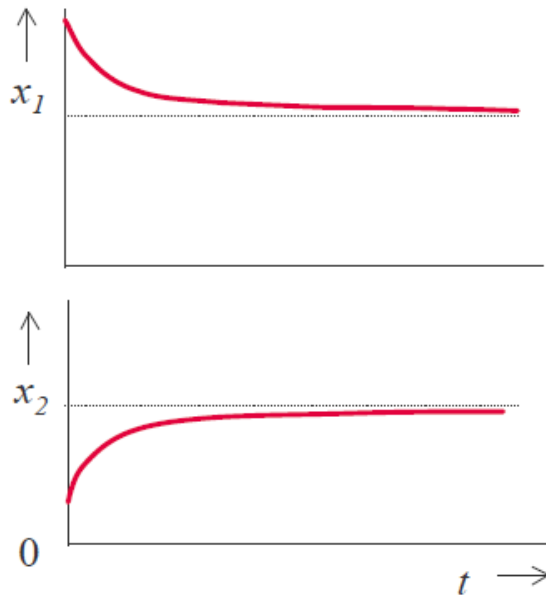
Study GRN dynamics by differential equation models

Simple example 2:

Positive feedback system

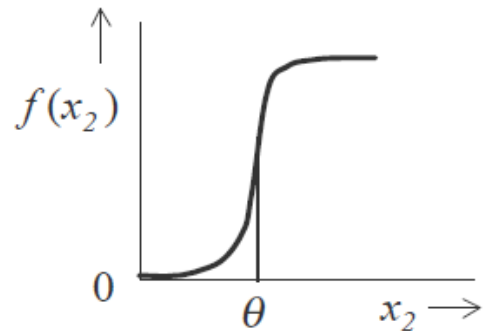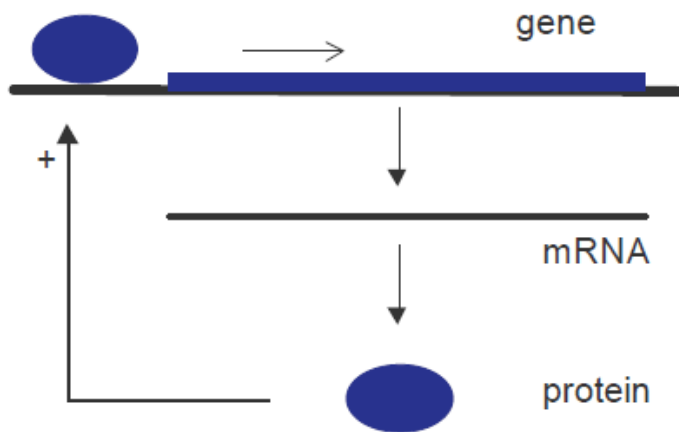- ❖ Gene encodes a protein activating its own expression: **positive feedback**
- ❖ Positive feedback important for **differentiation**, evolution towards one of two alternative states of system
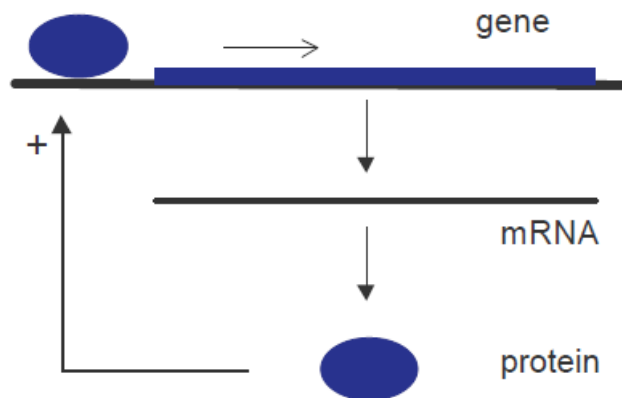
Study GRN dynamics by differential equation models

Simple example 2:

Model of positive feedback system

x_1 = mRNA concentration

x_2 = protein concentration

$$\dot{x}_1 = \kappa_1 f(x_2) - \gamma_1 x_1$$

$$\dot{x}_2 = \kappa_2 x_1 - \gamma_2 x_2$$

κ_1, $\kappa_2 > 0$, production rate constants
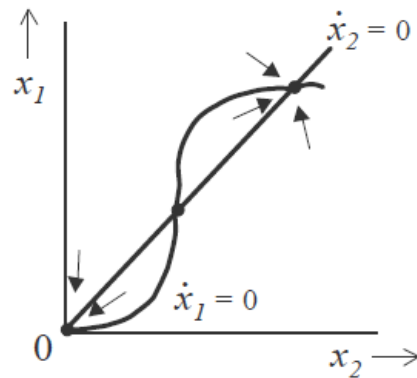
γ_1, $\gamma_2 > 0$, degradation rate constants

$$f(x_2) = \frac{x_2^n}{\theta^n + x_2^n}$$

Study GRN dynamics by differential equation models

Simple example 2:

Steady state analysis

❖ No analytical solution of nonlinear differential equations describing feedback system

❖ System has three **steady states**

$$\dot{x}_1 = 0 : x_1 = \frac{\kappa_1}{\gamma_1} f(x_2)$$
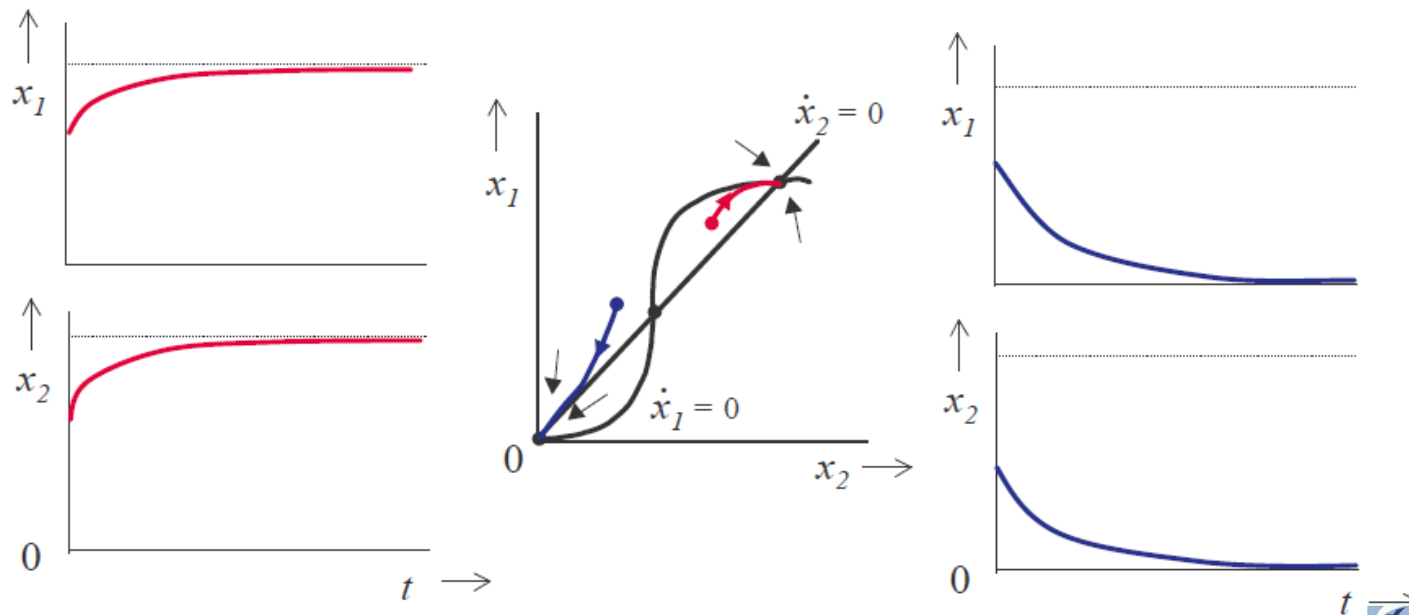
$$\dot{x}_2 = 0 : x_1 = \frac{\gamma_2}{\kappa_2} x_2$$

❖ Two **stable** and one **unstable** steady state. System will tend to one of two stable steady states (differentiation)

Study GRN dynamics by differential equation models
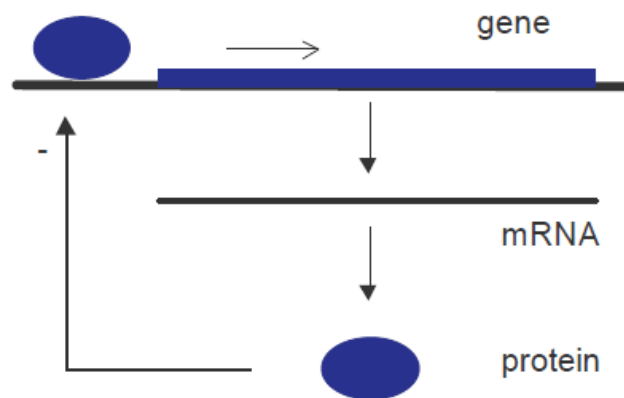
Simple example 2:

Study GRN dynamics by differential equation models

Simple example 3:

Model of time-delay feedback system

❖ Time to complete transcription and translation introduces **time-delay** in differential equations

x_1 = mRNA concentration

x_2 = protein concentration

$$\dot{x}_1 = \kappa_1 f(x_2^\tau) - \gamma_1 x_1$$

$$\dot{x}_2 = \kappa_2 x_1^\tau - \gamma_2 x_2$$

$$x_1^\tau(t) = x_1(t - \tau_1), \quad \tau_1 > 0 \text{ time-delay}$$

$$x_2^\tau(t) = x_2(t - \tau_2), \quad \tau_2 > 0 \text{ time-delay}$$

❖ Time-delay feedback systems may exhibit oscillatory behavior

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Study GRN dynamics by differential equation models

Simple example 4:

More complex feedback systems

❖ Gene encodes a protein activating synthesis of another protein inhibiting expression of gene: **positive and negative feedback**
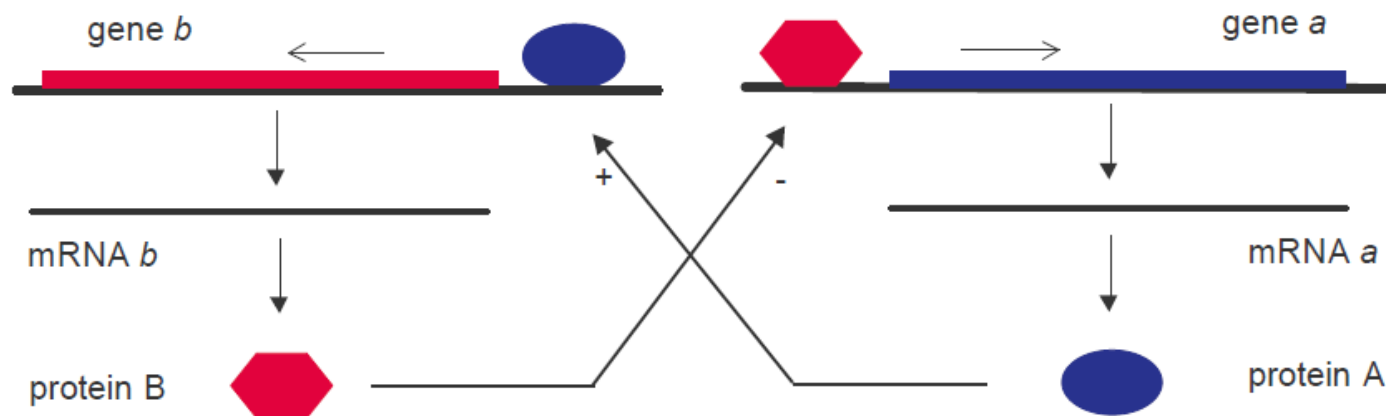
❖ Interlocking feedback loops give rise to models with complex dynamics: **numerical simulation** techniques necessary

Study GRN dynamics by differential equation models

❖ Differential equations have been used to model a variety of genetic regulatory networks:

- circadian rhythms in *Drosophila* (Leloup and Goldbeter, 1998)

- λ phage infection of *E. coli* (McAdams and Shapiro, 1998)

- segmentation of early embryo of *Drosophila* (Reinitz and Sharp, 1996)

- cell division in *Xenopus* (Novak and Tyson, 1993)

- Trp synthesis in *E. coli* (Santillán and Mackey, 2001)

- induction of *lac* operon in *E. coli* (Carrier and Keasling, 1999)

- developmental cycle of bacteriophage T7 (Endy *et al.*, 2000)

- ...

Study GRN dynamics by stochastic models

Gene expression is discrete process

❖ Gene expression is result of large number of **discrete** events: chemical reactions

$$DNA + RNAP \rightarrow DNA_0 \cdot RNAP$$

$$DNA_i \cdot RNAP \rightarrow DNA_{i+1} \cdot RNAP$$
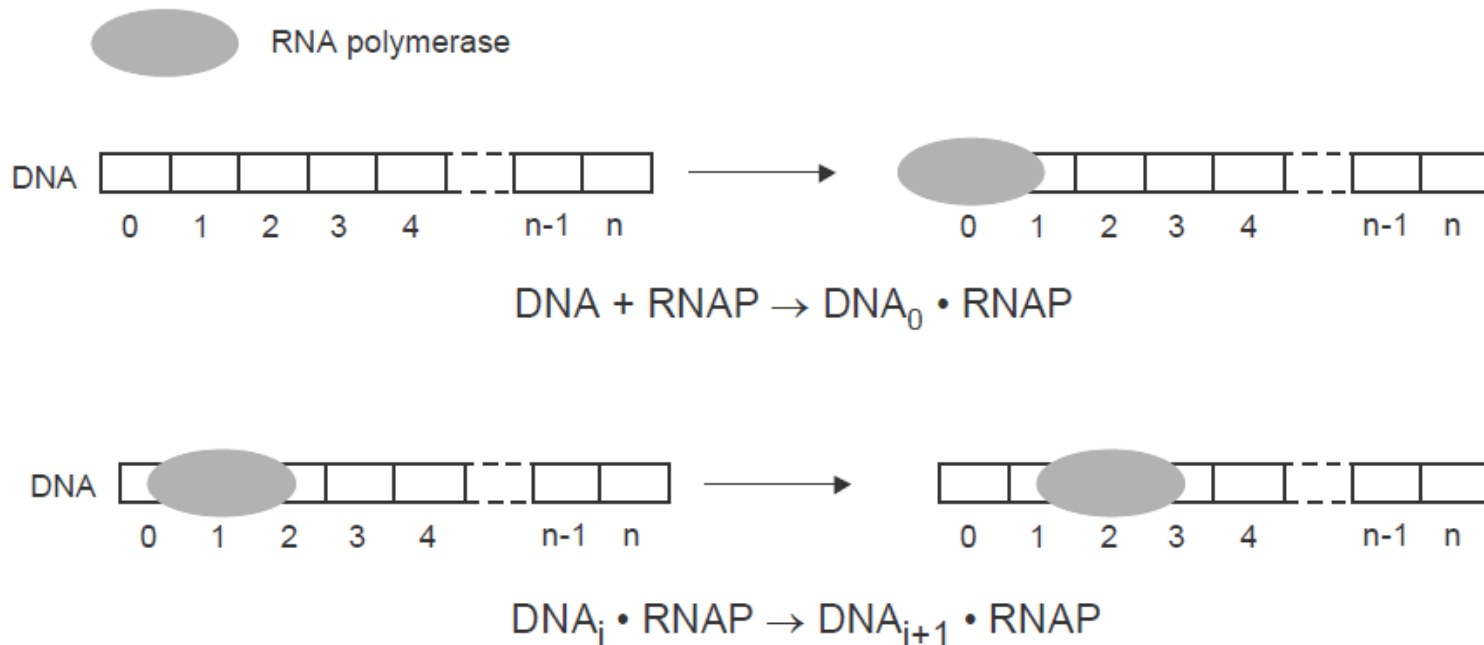
Study GRN dynamics by stochastic models

Gene expression is stochastic process

❖ Gene expression is **stochastic** process: random time intervals τ between occurrence of reactions

RNA polymerase

DNA | 0 1 2 3 4 ... n-1 n → 0 1 2 3 4 ... n-1 n

❖ Time interval τ has probability distribution

$P(\tau)$

$\tau \rightarrow$
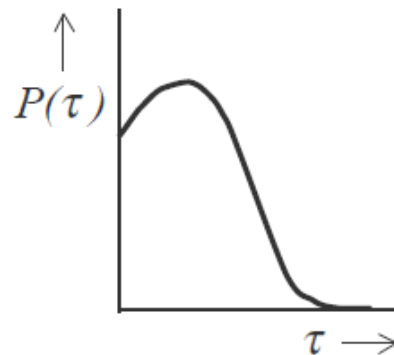
Study GRN dynamics by stochastic models

Differential equations are abstractions

❖ Differential equation models make **continuous** and **deterministic** abstraction of discrete and stochastic process

- $x_i(t) \in \mathbf{R}_{\geq 0}$ is continuous variable

- $\dot{x}_i = f_i(\boldsymbol{x})$ determines change in x_i at t

❖ Abstraction may not be warranted when modeling gene regulation on molecular level: low number of molecules

❖ Therefore, more realistic **stochastic models** of gene regulation

Study GRN dynamics by stochastic models

Stochastic variables

❖ **Stochastic variables** X_i describe number of molecules of proteins, mRNAs, etc.

- $X_i(t) \in \mathbb{N}_{\geq 0}$ is discrete variable

- $P(X_i(t))$ is probability distribution describing probability that at time-point t cell contains X_i molecules of i
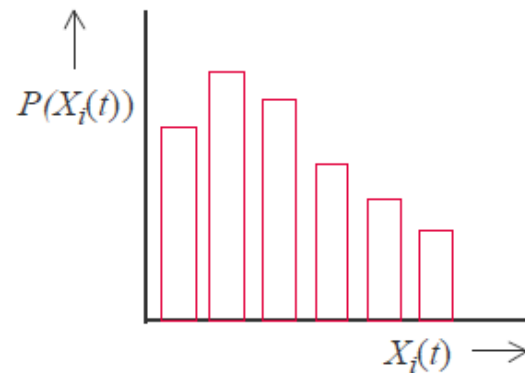
Study GRN dynamics by stochastic models

Stochastic master equations

❖ **Stochastic master equations** describe evolution of state $X = [X_1, \ldots, X_n]'$ of regulatory system

$$P(X(t + \Delta t)) = P(X(t))\left(1 - \sum_{j=1}^{m} \alpha_j \Delta t\right) + \sum_{j=1}^{m} \beta_j \Delta t$$

- m is the number of reactions that can occur in the system

- $\alpha_j \Delta t$ is the probability that reaction j will occur in $[t, t + \Delta t]$ given that the system is in state X at t

- $\beta_j \Delta t$ is the probability that reaction j will bring the system in state X from another state in $[t, t + \Delta t]$

van Kampen, 1997

Study GRN dynamics by stochastic models

Stochastic simulation

❖ For $\Delta t \to 0$ we obtain

$$\frac{\partial}{\partial t} P(X(t)) = \sum_{j=1}^{m} (\beta_j - \alpha_j P(X(t)))$$

❖ Analytical solution of master equations is not possible

❖ Stochastic simulation by predicting a sequence of reactions changing the state of the system, starting from initial state X_0

Stochastic simulation uses stochastic variables τ and ρ

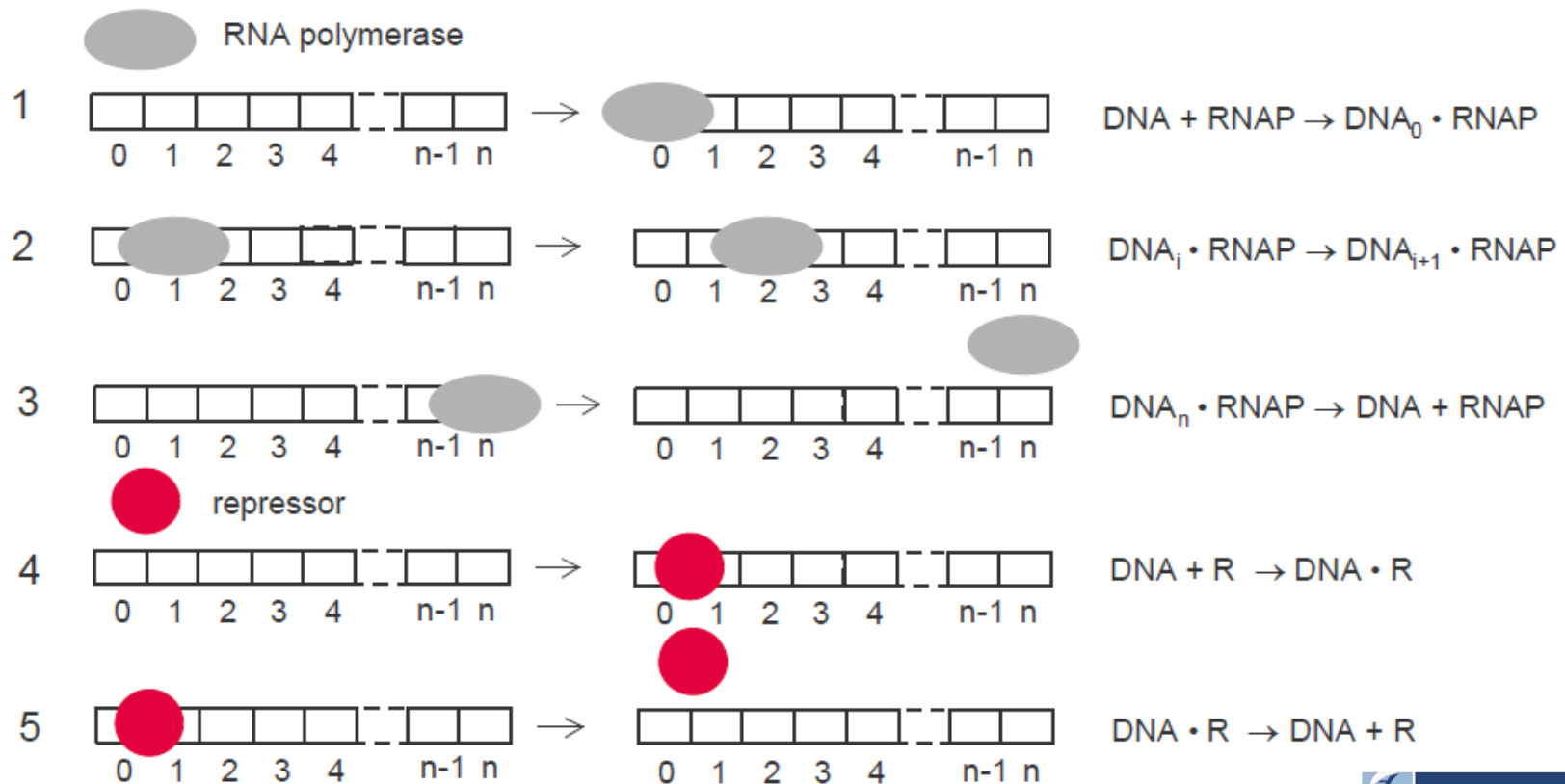τ = time interval until occurrence of next reaction

ρ = type of reaction

Gillespie, 1977

Study GRN dynamics by stochastic models
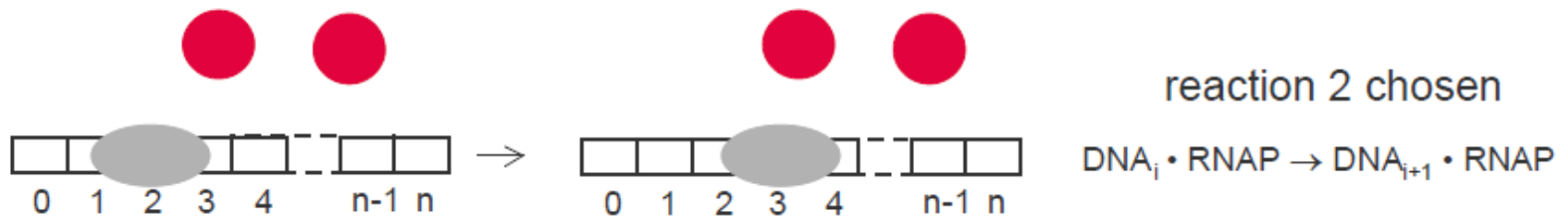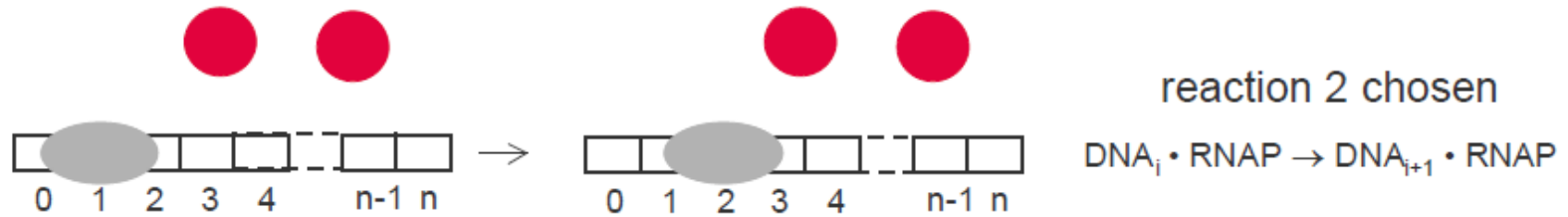

Reactions in gene expression

❖ Five possible reactions in gene expression are considered

RNA polymerase

1. $DNA + RNAP \rightarrow DNA_0 \cdot RNAP$

2. $DNA_i \cdot RNAP \rightarrow DNA_{i+1} \cdot RNAP$

3. $DNA_n \cdot RNAP \rightarrow DNA + RNAP$

repressor

4. $DNA + R \rightarrow DNA \cdot R$

5. $DNA \cdot R \rightarrow DNA + R$

Study GRN dynamics by stochastic models

Simulation of gene expression

❖ Stochastic simulation from initial state

reaction 1 chosen

$$DNA + RNAP \rightarrow DNA_0 \cdot RNAP$$

reaction 2 chosen

$$DNA_i \cdot RNAP \rightarrow DNA_{i+1} \cdot RNAP$$

reaction 2 chosen

$$DNA_i \cdot RNAP \rightarrow DNA_{i+1} \cdot RNAP$$
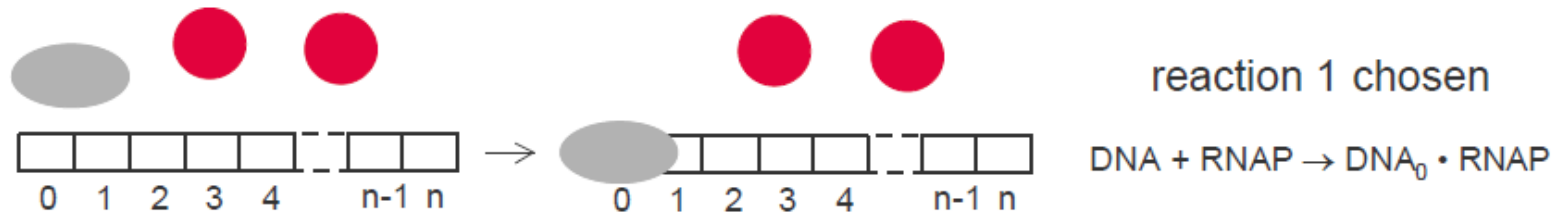
Study GRN dynamics by stochastic models

Stochastic outcome of simulation

- Simulation starting from same initial state will generally lead to different results

reaction 4 chosen

$DNA + R \rightarrow DNA \cdot R$

reaction 5 chosen

$DNA \cdot R \rightarrow DNA + R$

reaction 1 chosen

$DNA + RNAP \rightarrow DNA_0 \cdot RNAP$
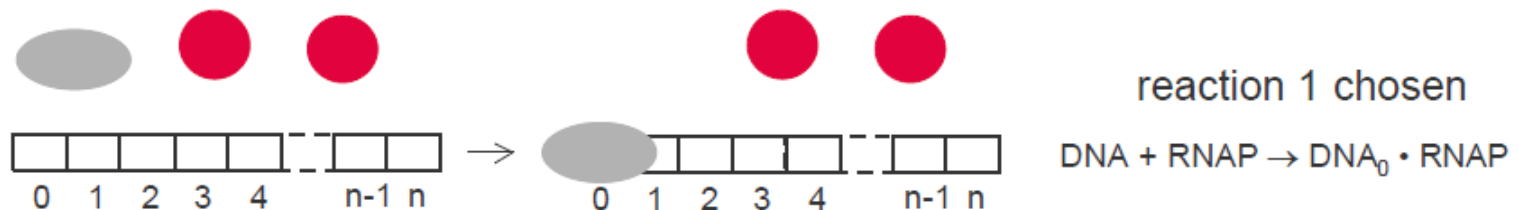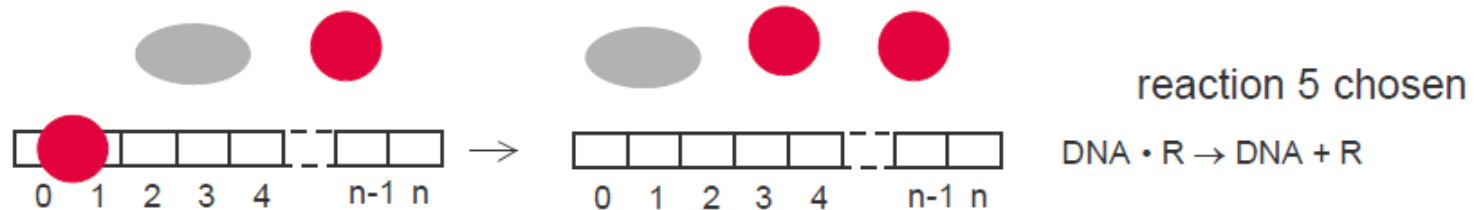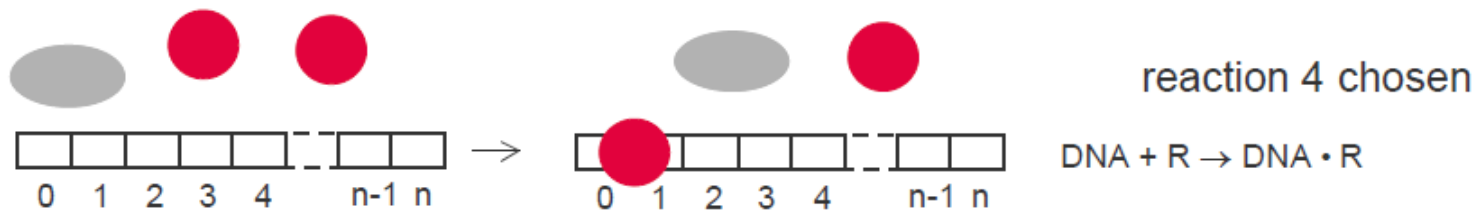
Study GRN dynamics by stochastic models

Stochastic simulation and master equation

❖ Repeating stochastic simulations allows approximation of $P(X(t))$ in master equation to be given
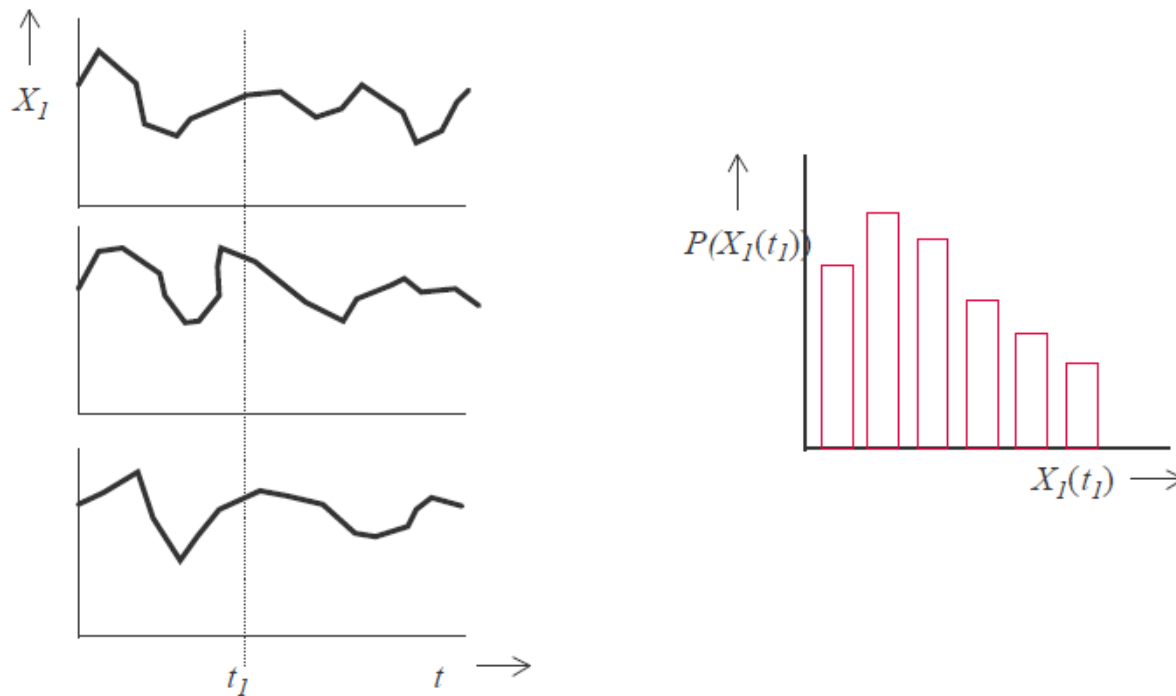
Study GRN dynamics by stochastic models

Application of stochastic equations

❖ Stochastic equations have been used to model genetic and other regulatory systems:

- λ phage infection of *E. coli* (Arkin *et al.*, 1998)

- chemotactic signalling in *E. coli* (Morton-Firth and Bray, 1998)

- ...

Study GRN dynamics by stochastic models

❖ **Pro**: more realistic models of gene regulation

❖ **Contra**: required information on regulatory mechanisms on molecular level usually not available

    reaction schemas and values of parameters τ and ρ are not or incompletely known

❖ **Contra**: stochastic simulation is computationally expensive

    large networks cannot currently be handled

Take Home Messages

- Introduction to gene regulation

- Construction of GRN

  - Unsupervised

  - Supervised

- Modeling the dynamics of GRN

  - Discrete Models (Boolean Network)

  - Differential and Stochastic Equations