Mathematical Modeling of Genetic Regulatory Networks

Hidde de Jong

Projet HELIX

INRIA Rhône-Alpes

655, avenue de l'Europe

Montbonnot, 38334 Saint Ismier CEDEX

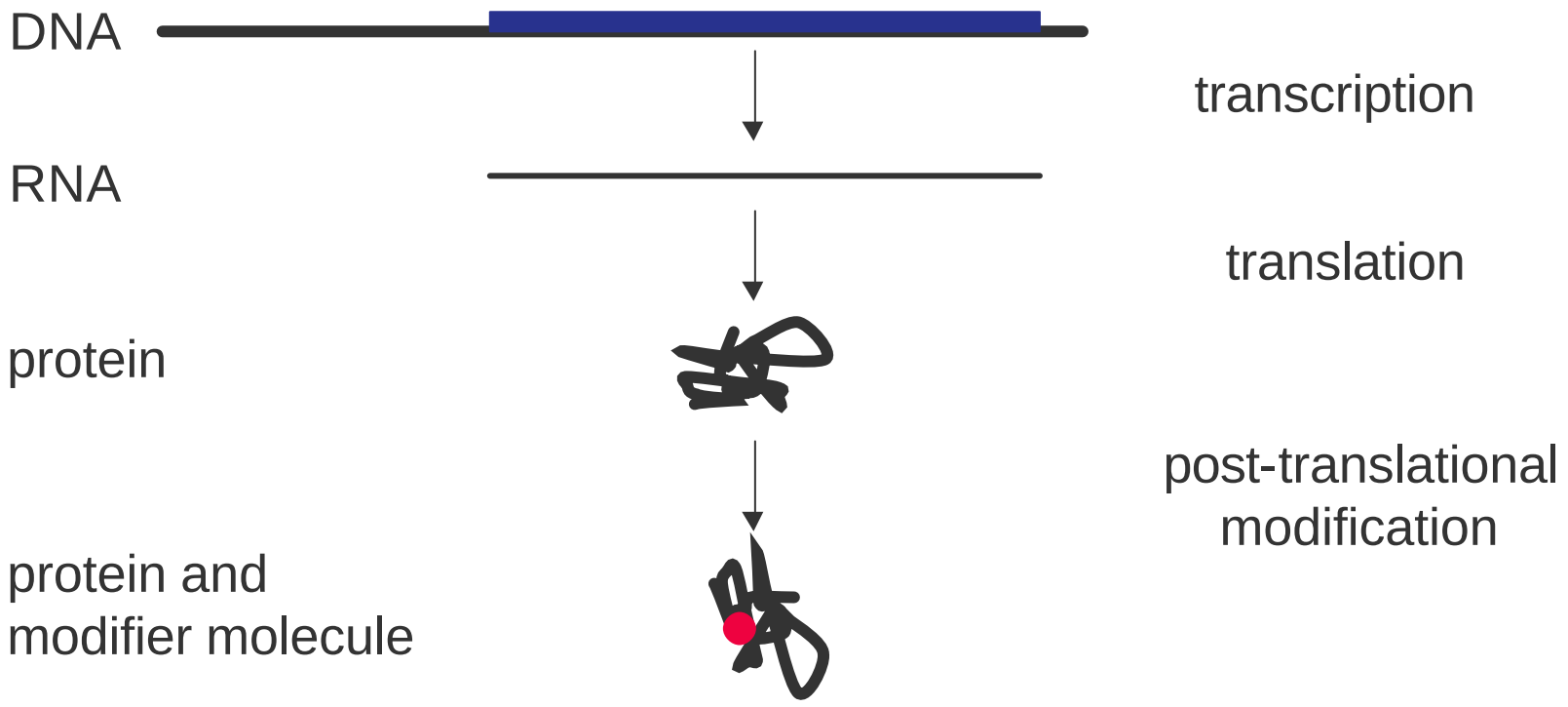
Email: **Hidde.de-Jong@inrialpes.fr**

Overview

1. Genetic regulatory networks
2. Modeling and simulation of genetic regulatory networks
3. Modeling and simulation approaches:
   - differential equations
   - stochastic equations
4. Conclusions

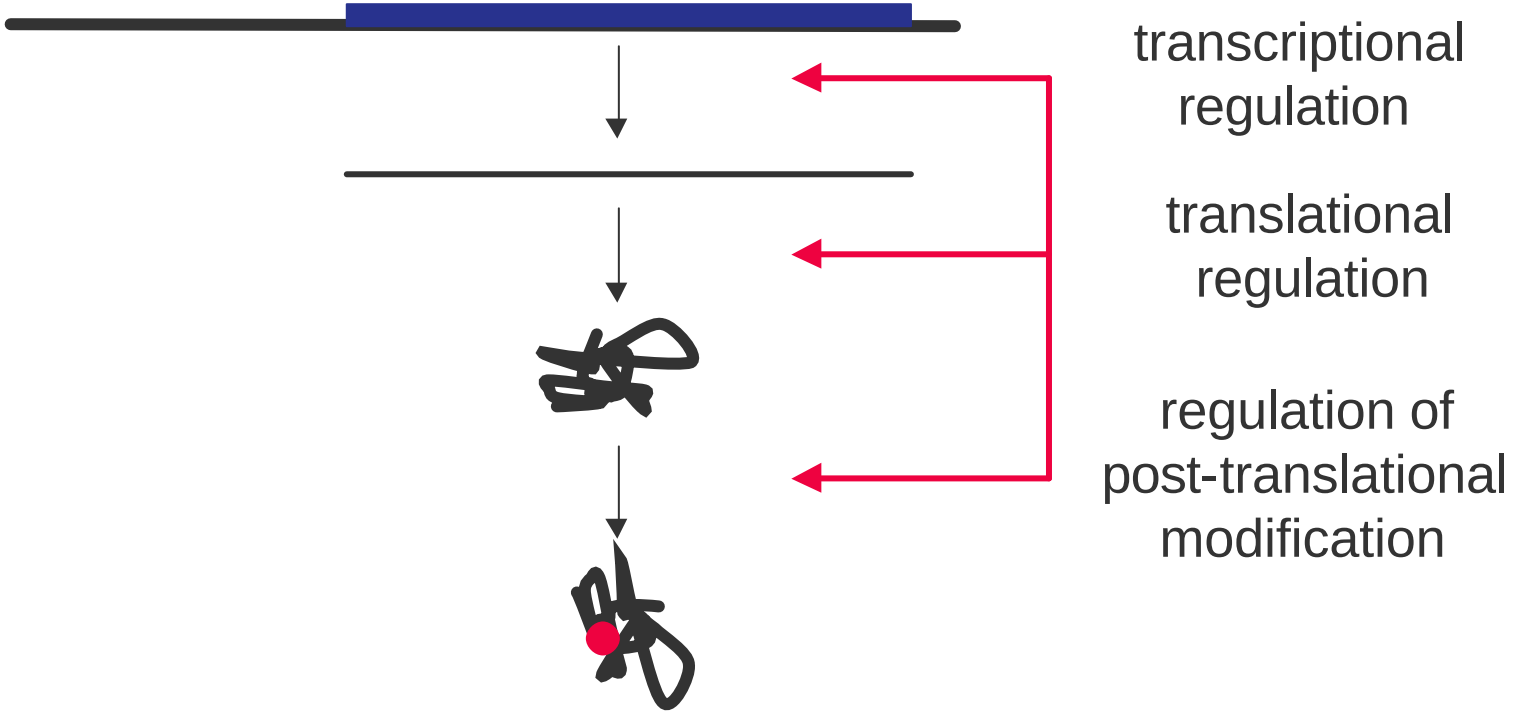
Genes and proteins

- Genes code for proteins that are essential for development and functioning of organism: **gene expression**



Regulation of gene expression

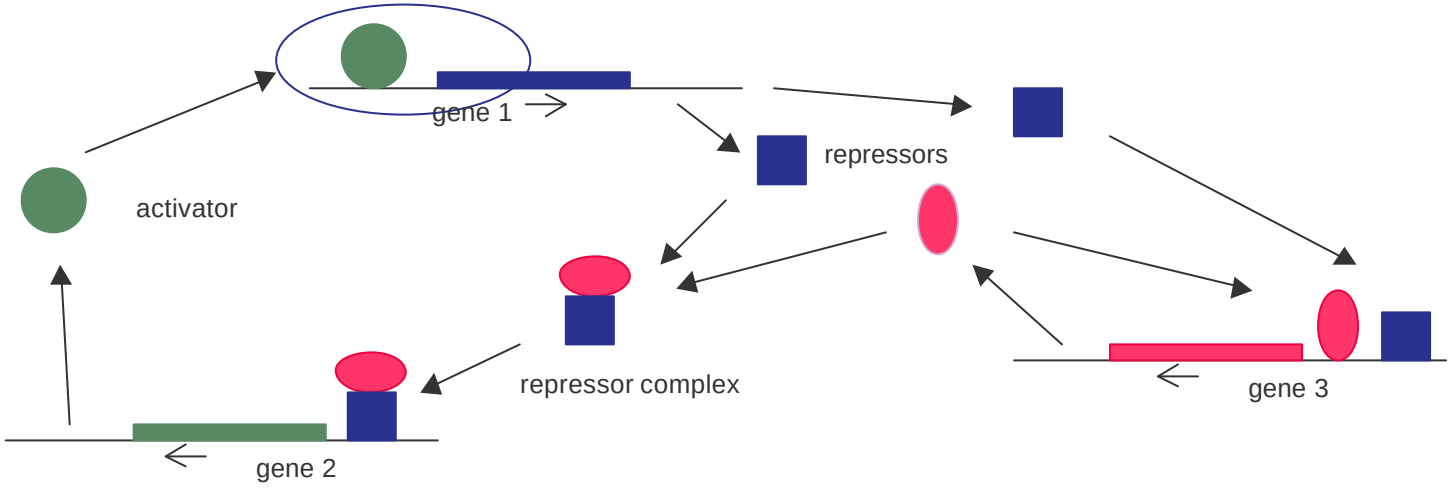
- **Regulation** of gene expression on several levels



- Gene expression controlled by proteins produced by other genes: **regulatory interactions**

Genetic regulatory network

- **Genetic regulatory network** consists of set of genes, proteins, small molecules, and their mutual regulatory interactions

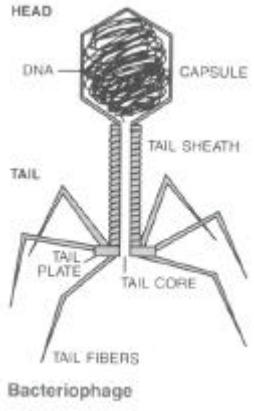


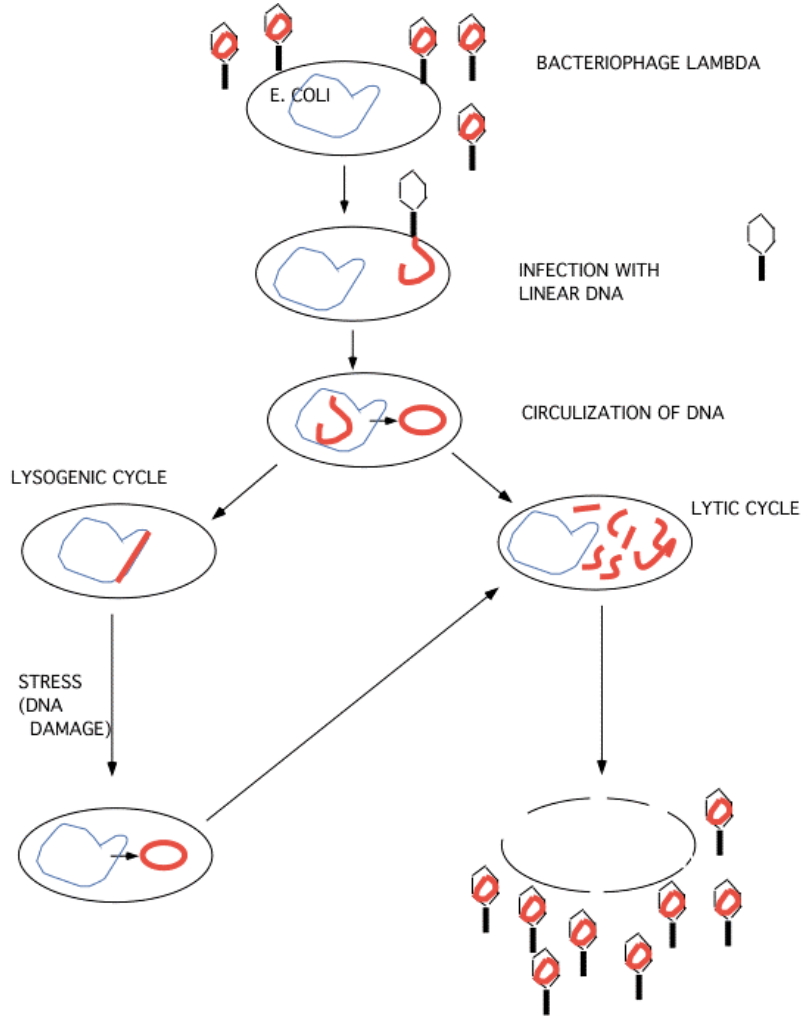
- Development and functioning of organisms cell emerges from interactions in genetic regulatory networks

Bacteriophage λ infection of *E. coli*

- Response of *E. coli* to phage λ infection involves decision between alternative developmental pathways: **lytic cycle** and **lysogeny**

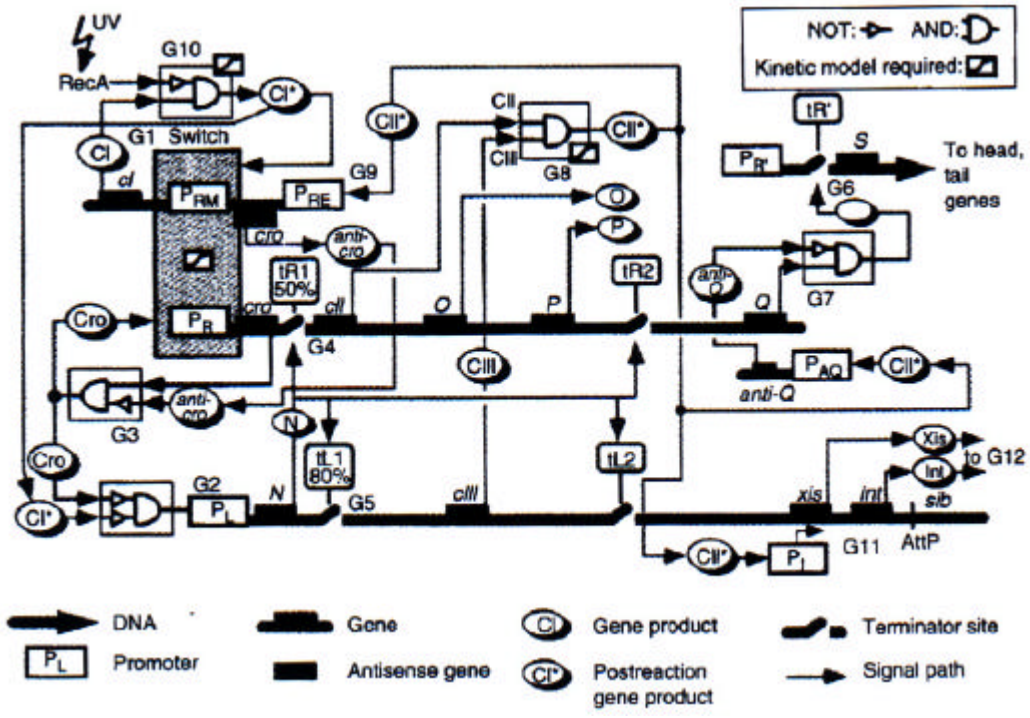
  Ptashne, 1992





Genetic regulatory network phage λ

- Choice between alternative developmental pathways controlled by network of genes, proteins, and mutual regulatory interactions



McAdams & Shapiro, 1995

Computational approaches

- Most genetic regulatory networks are **large** and **complex**

  Cells have many components that can interact in complex ways

- Dynamics of large and complex genetic regulatory processes hard to understand by intuitive approaches alone
- Mathematical methods for **modeling** and **simulation** are required:
  - precise and unambiguous description of network of interactions
  - systematical derivation of behavioral predictions
- Practical application of mathematical methods requires user-friendly **computer tools**

Mathematical modeling approaches

- Mathematical modeling has developed since the 1960s and is currently attracting much attention

  Bower and Bolouri, 2001; Hasty *et al.*, 2001; McAdams and Arkin, 1998; Smolen *et al.*, 2000; de Jong, 2002

- Two **approaches to computer modeling and simulation** discussed in this session:
  - differential equations
  - stochastic equations
- Jean-Luc Gouzé will discuss class of piecewise-linear differential equations central to this project in more detail

Differential equation models

- Cellular concentration of proteins, mRNAs, and other molecules at time-point t represented by continuous variable $x_i(t) \in \mathrm{R}_{\geq 0}$
- Regulatory interactions modeled by **kinetic equations**

$$\dot{x}_i = f_i(\boldsymbol{x}), \quad 1 \leq i \leq n,$$

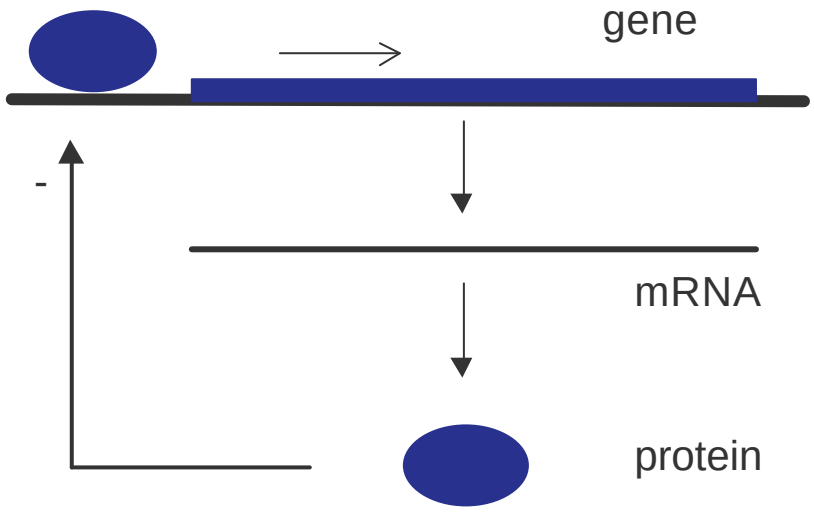
  where $f_i(\boldsymbol{x})$ is **rate law**

- Rate of change of variable x_i is function of other concentration variables $\boldsymbol{x} = [x_1, \ldots, x_n]'$
- Differential equations are major modeling formalism in mathematical biology

Segel, 1984; Kaplan and Glass, 1995; Murray, 2002

Negative feedback system

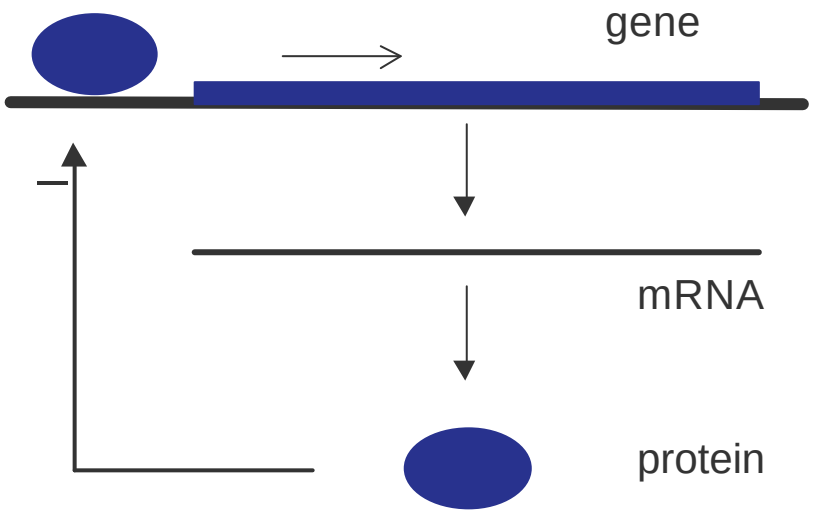
- Gene encodes a protein inhibiting its own expression: **negative feedback**



- Negative feedback important for **homeostasis**, maintenance of system near a desired state

Thomas and d'Ari, 1990

Model of negative feedback system



x_1 = mRNA concentration

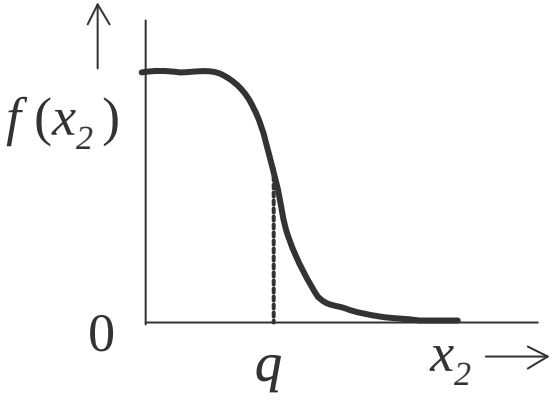
x_2 = protein concentration

$$\dot{x}_1 = k_1 f(x_2) - g_1 x_1$$

$$\dot{x}_2 = k_2 x_1 - g_2 x_2$$

$k_1, k_2 > 0$, production rate constants

$g_1, g_2 > 0$, degradation rate constants

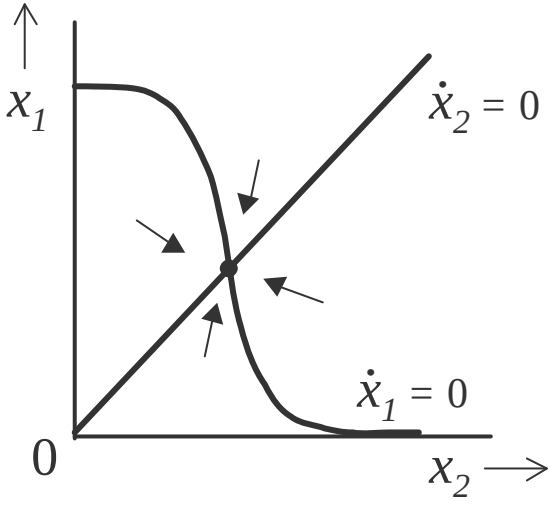


$$f(x_2) = \frac{q^n}{q^n + x_2^n}, \quad q > 0 \text{ threshold}$$

Steady state analysis

- No analytical solution of nonlinear differential equations describing feedback system

- System has single **steady state** at $\dot{x} = \mathbf{0}$



$$\dot{x}_1 = 0 : \; x_1 = \frac{k_1}{g_1} f(x_2)$$

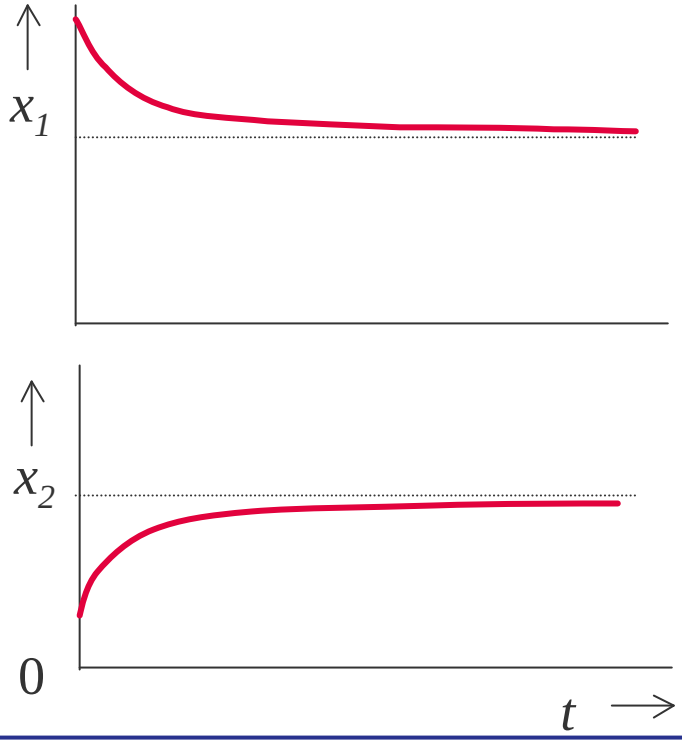
$$\dot{x}_2 = 0 : \; x_1 = \frac{g_2}{k_2} x_2$$

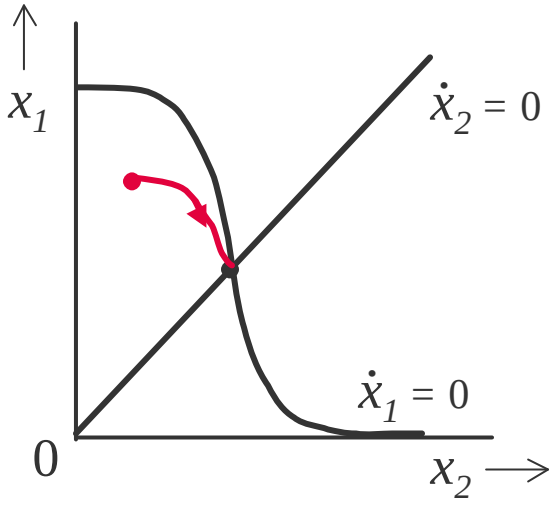
- Steady state is **stable**, that is, after perturbation system will return to steady state (homeostasis)

Transient behavior after pertubation

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

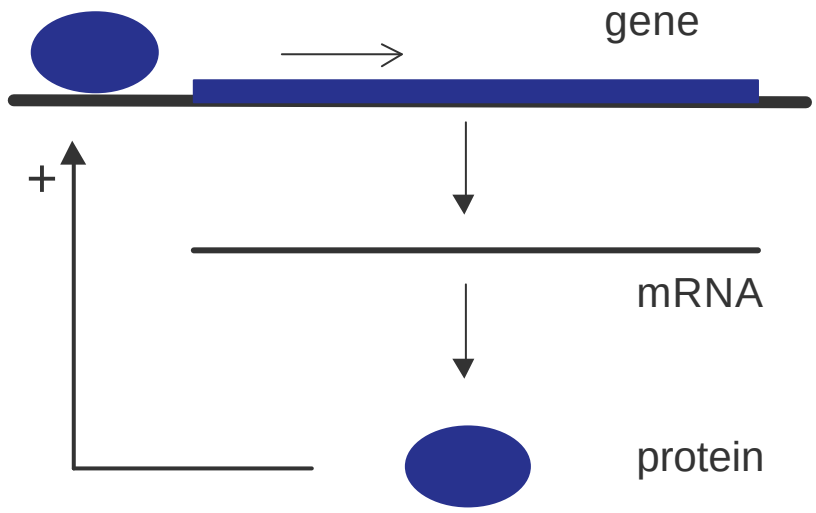
Initial values $x_1(0)$, $x_2(0)$ correspond to perturbation





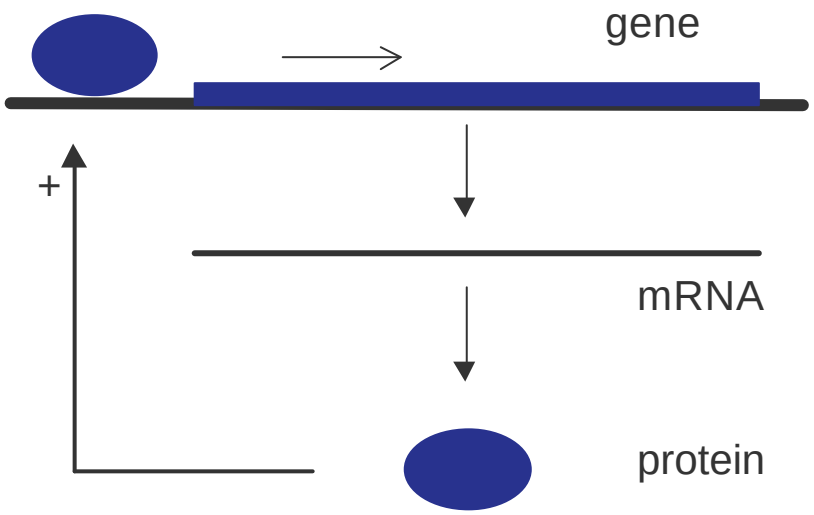
Positive feedback system

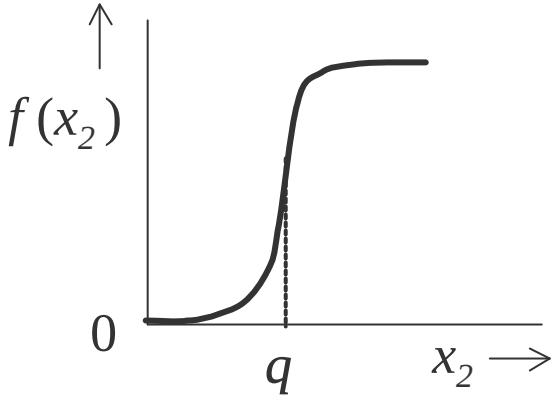
- Gene encodes a protein activating its own expression: **positive feedback**



- Positive feedback important for **differentiation**, evolution towards one of two alternative states of system

Model of positive feedback system





x_1 = mRNA concentration

x_2 = protein concentration

$$\dot{x}_1 = k_1 f(x_2) - g_1 x_1$$

$$\dot{x}_2 = k_2 x_1 - g_2 x_2$$

$k_1, k_2 > 0$, production rate constants

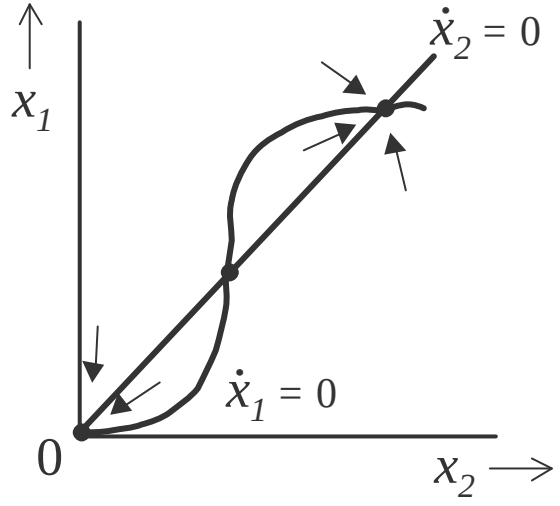
$g_1, g_2 > 0$, degradation rate constants

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

Steady state analysis

- No analytical solution of nonlinear differential equations describing feedback system

- System has three **steady states**



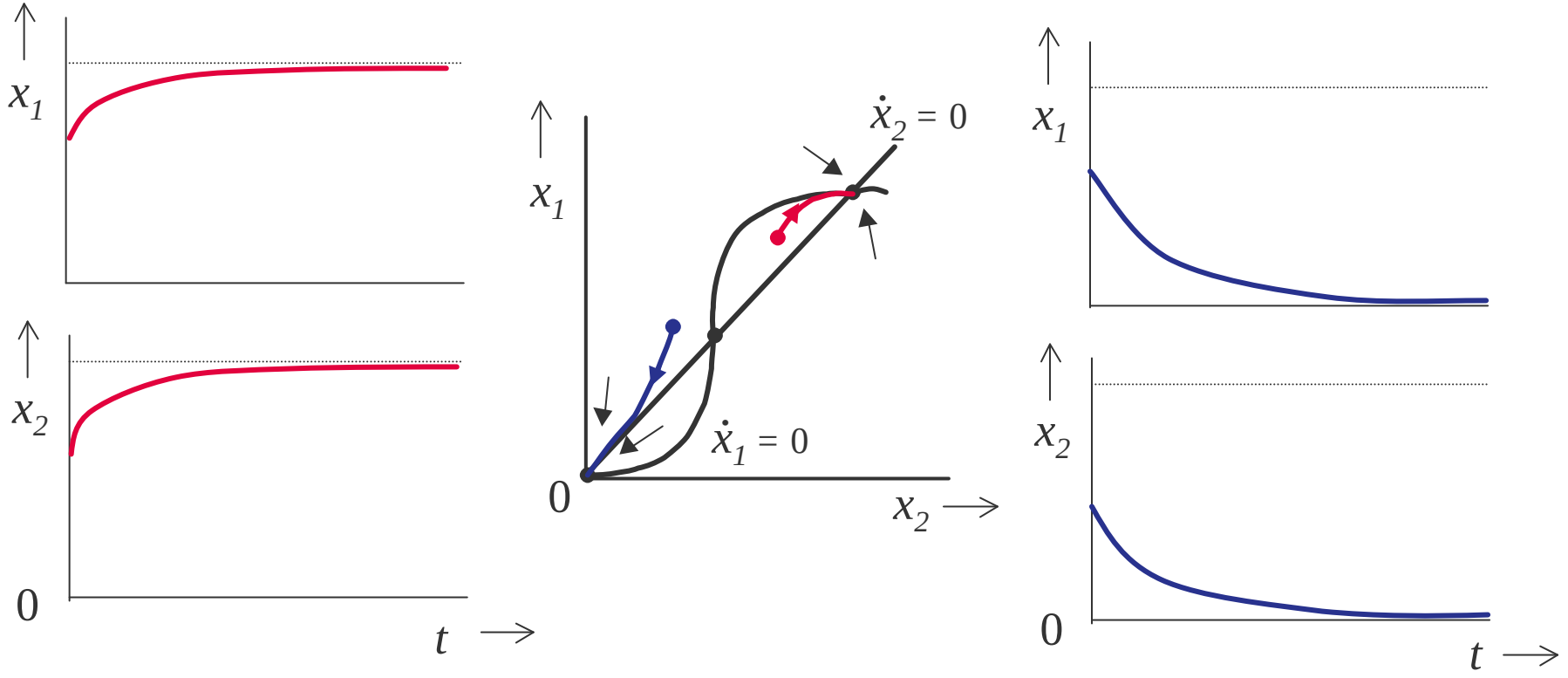
$$\dot{x}_1 = 0 : \; x_1 = \frac{k_1}{g_1} f(x_2)$$

$$\dot{x}_2 = 0 : \; x_1 = \frac{g_2}{k_2} x_2$$

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

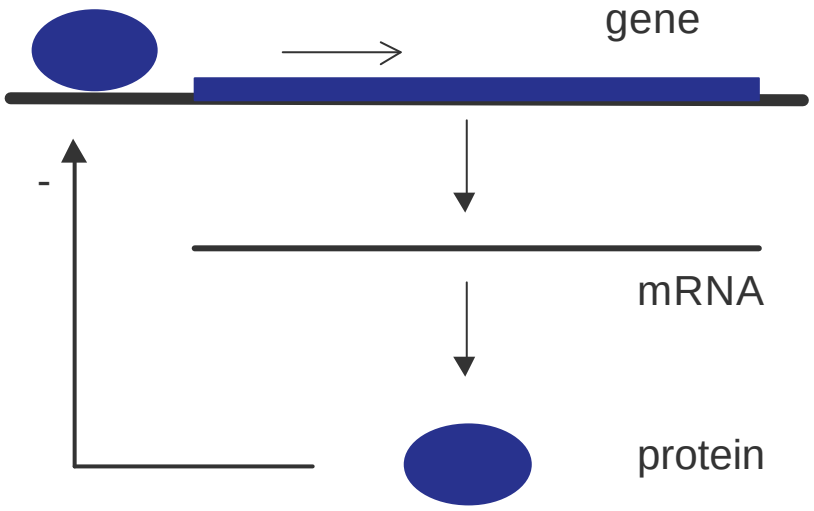
Transient behavior after pertubation

- Depending on strength of perturbation, transient behavior towards different steady states



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

$$\dot{x}_1 = k_1 f(x_2^t) - g_1 x_1$$

$$\dot{x}_2 = k_2 x_1^t - g_2 x_2$$

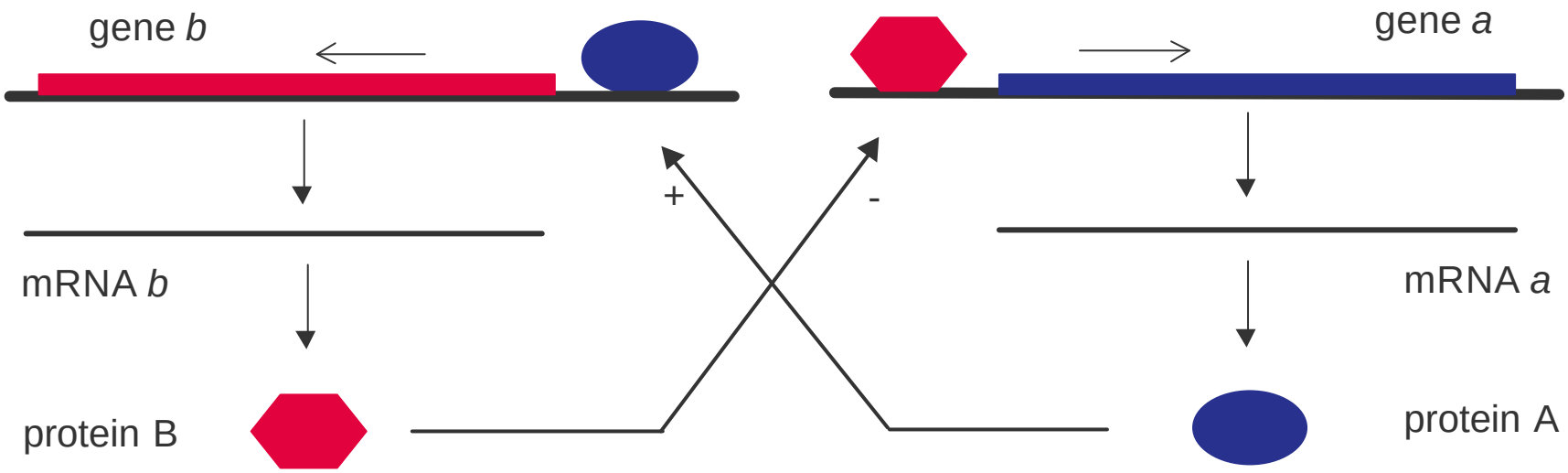
$x_1^t(t) = x_1(t - t_1)$, $t_1 > 0$ time-delay

$x_2^t(t) = x_2(t - t_2)$, $t_2 > 0$ time-delay

- Time-delay feedback systems may exhibit oscillatory behavior

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

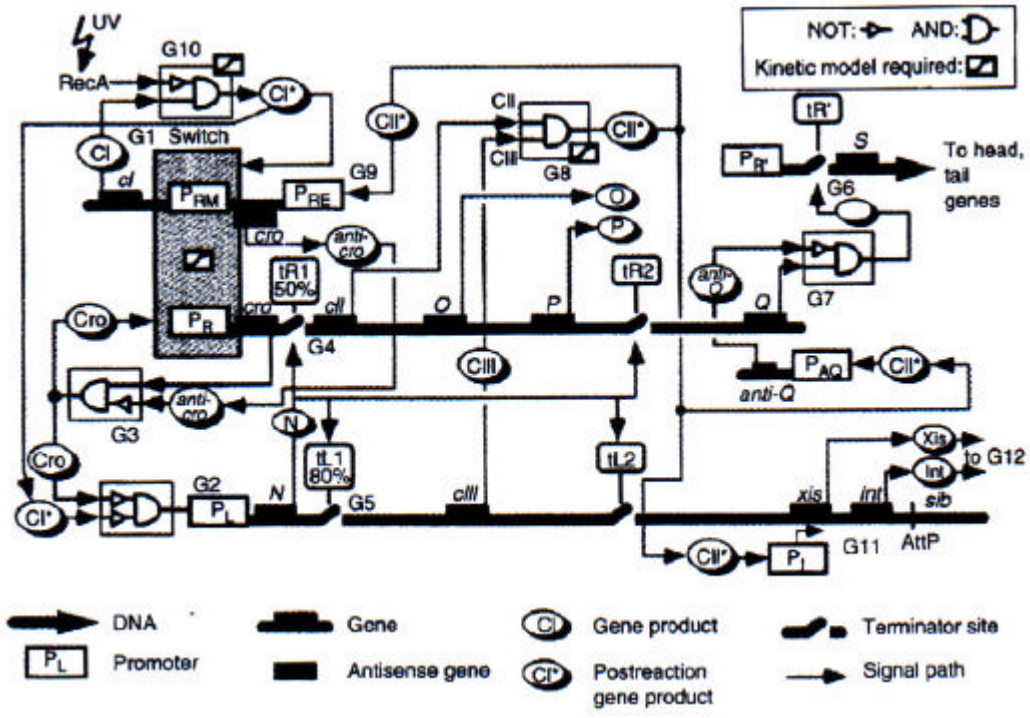
Application of differential equations

- 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)
  - ...

Simulaton of phage ? infection

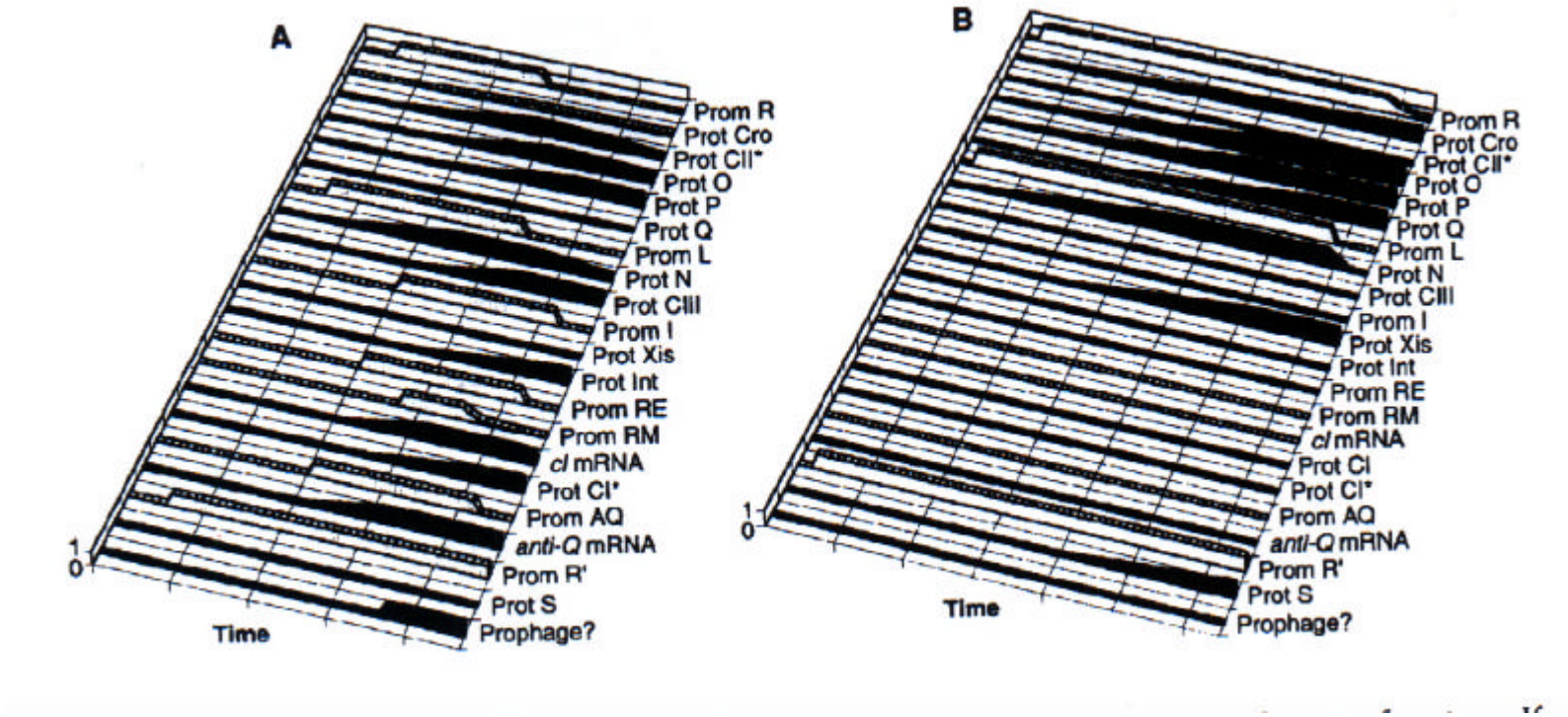
- Kinetic model of the phage ? network underlying decision between lytic cycle and lysogeny

  McAdams & Shapiro, 1995



Simulaton of phage ? infection

- Time evolution of promoter activity and protein concentrations in (a) lysogenic and (b) lytic pathways



McAdams & Shapiro, 1995

Evaluation of differential equations

- **Pro**: general formalism for which powerful analysis and simulation techniques exist

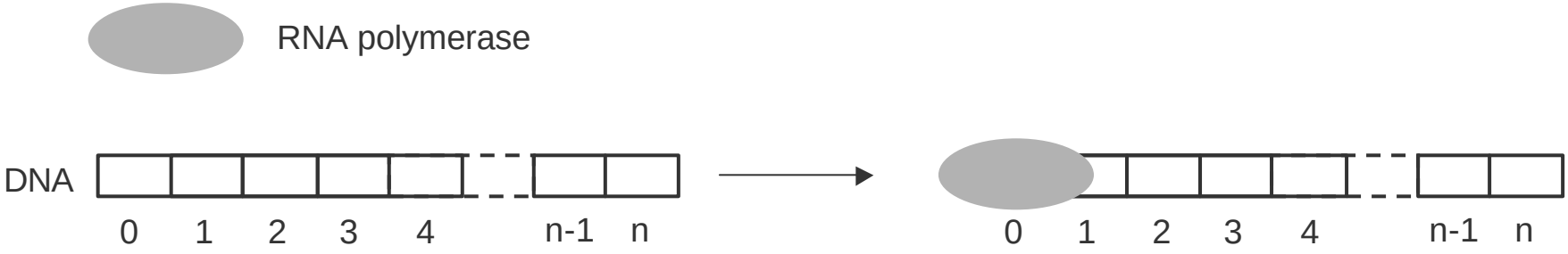
- **Contra**: numerical techniques are often not appropriate due to lack of quantitative knowledge

  value of parameters and evolution of concentrations are not known

- **Contra**: implicit assumptions of continuous and deterministic change of concentrations may not be valid on molecular level

Gene expression is discrete process

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

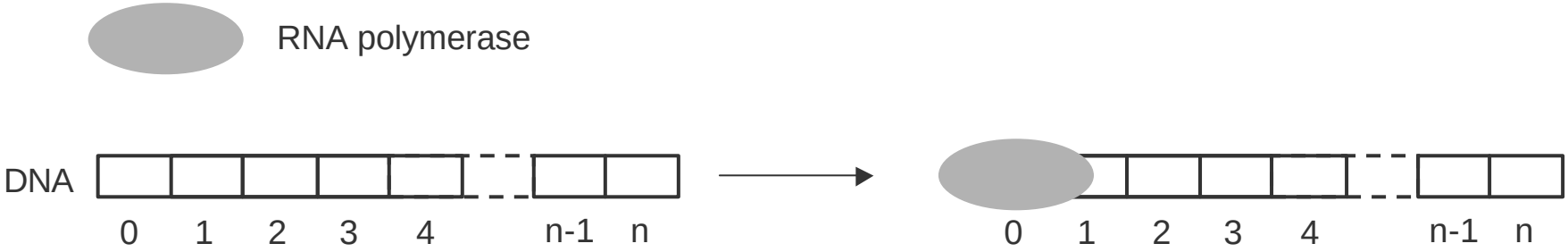


$$DNA + RNAP \rightarrow DNA_0 \bullet RNAP$$

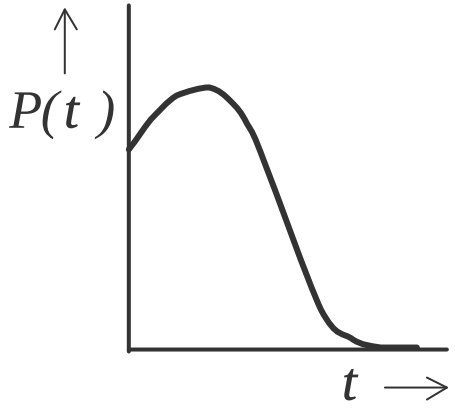
$$DNA_i \bullet RNAP \rightarrow DNA_{i+1} \bullet RNAP$$

Gene expression is stochastic process

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



- Time interval τ has probability distribution

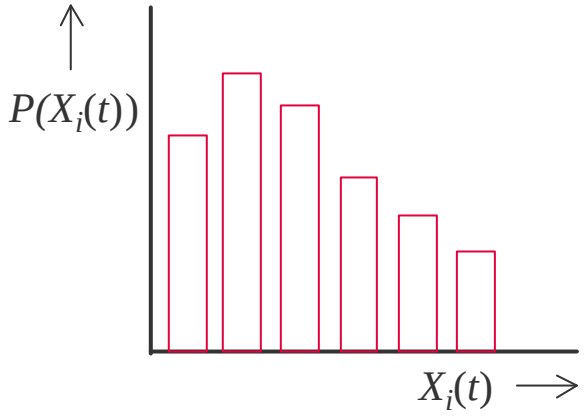


Differential equations are abstractions

- Differential equation models make **continuous** and **deterministic** abstraction of discrete and stochastic process
  - $x_i(t) \in \mathrm{R}_{\geq 0}$ 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

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



Stochastic master equations

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

$$P(\boldsymbol{X}(t + Dt)) = P(\boldsymbol{X}(t)) \left(1 - \sum_{j=1}^{m} a_j \, Dt\right) + \sum_{j=1}^{m} b_j \, Dt$$

  - m is the number of reactions that can occur in the system
  - $a_j \, Dt$ is the probability that reaction j will occur in $[t, t + Dt]$ given that the system is in state $\boldsymbol{X}$ at t
  - $b_j \, Dt$ is the probability that reaction j will bring the system in state $\boldsymbol{X}$ from another state in $[t, t + Dt]$

van Kampen, 1997

Stochastic simulation

- For $Dt \rightarrow 0$ we obtain

$$\frac{\partial}{\partial t} P(\boldsymbol{X}(t)) = \sum_{j=1}^{m} (b_j - a_j P(\boldsymbol{X}(t)))$$

- 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 $\boldsymbol{X_0}$

  Stochastic simulation uses stochastic variables t and r

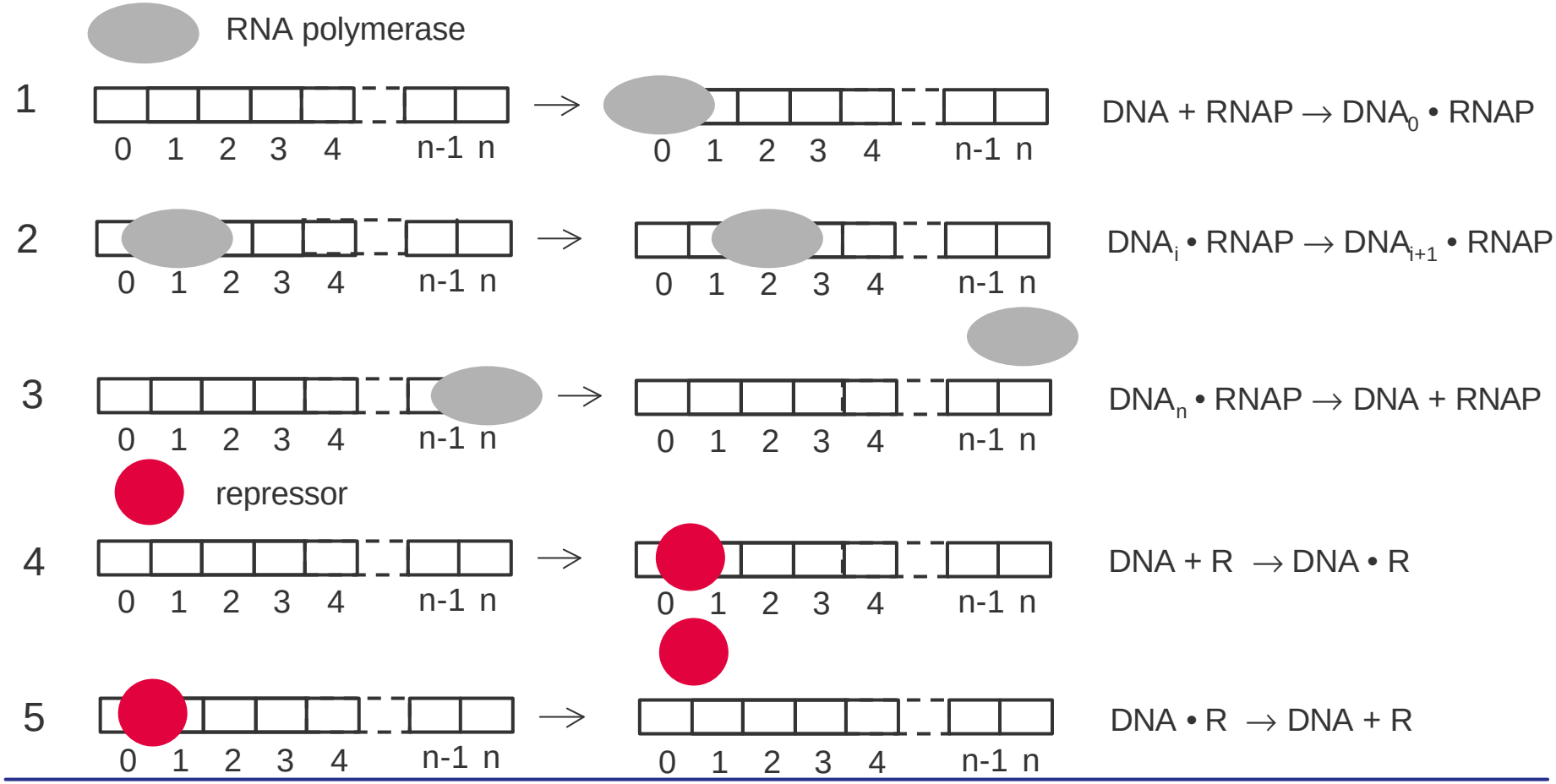
  t = time interval until occurrence of next reaction

  r = type of reaction

  Gillespie, 1977

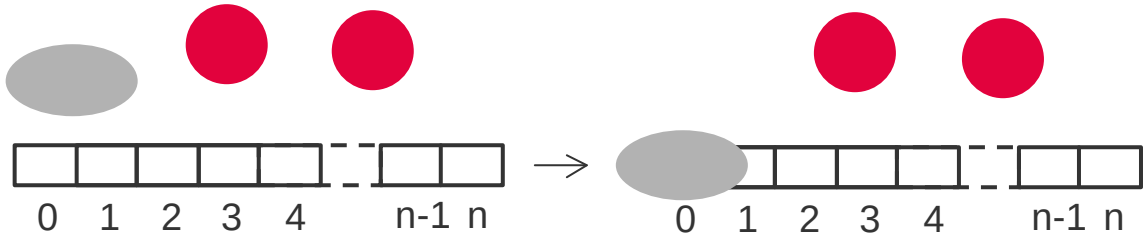
Reactions in gene expression

- Five possible reactions in gene expression are considered



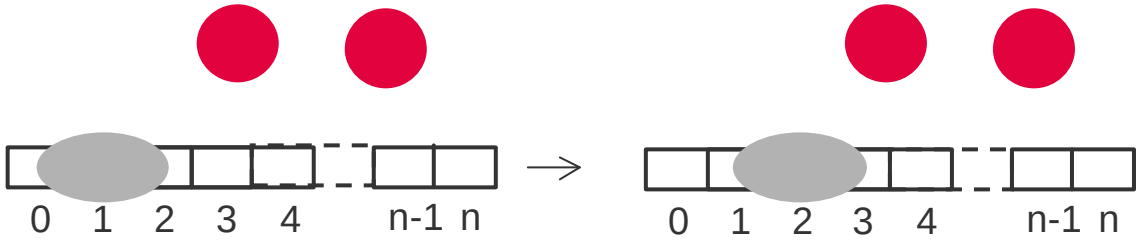
Simulation of gene expression

- Stochastic simulation from initial state



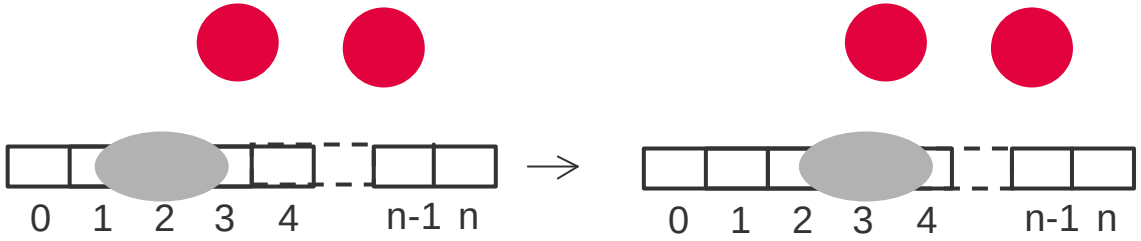
reaction 1 chosen

$DNA + RNAP \rightarrow DNA_0 \bullet RNAP$



reaction 2 chosen

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

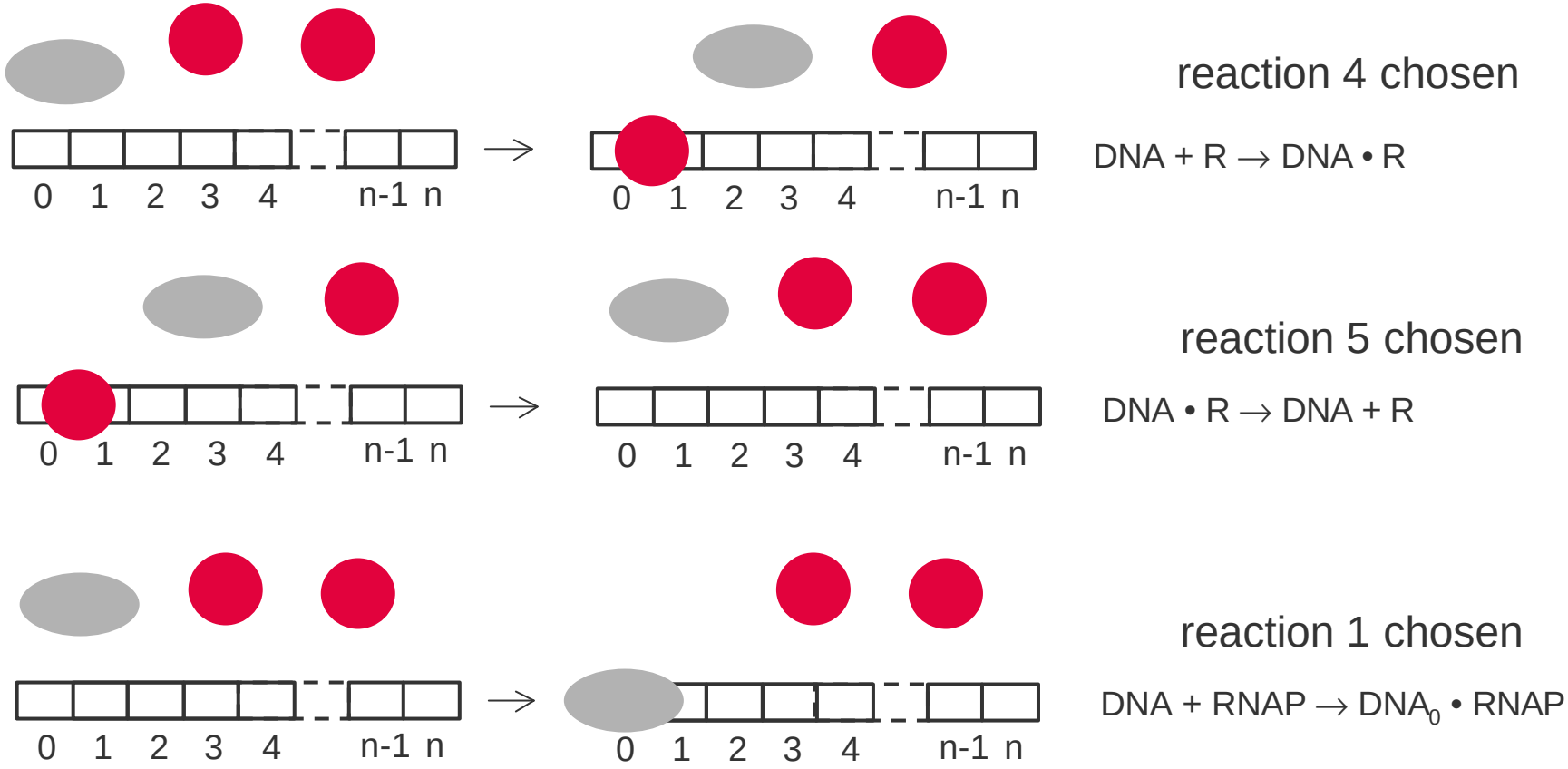


reaction 2 chosen

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

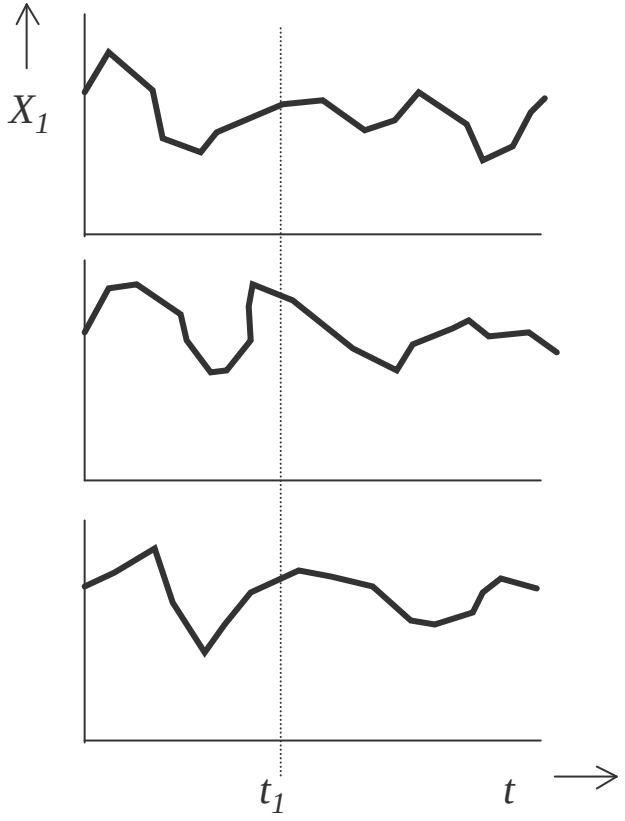
Stochastic outcome of simulation

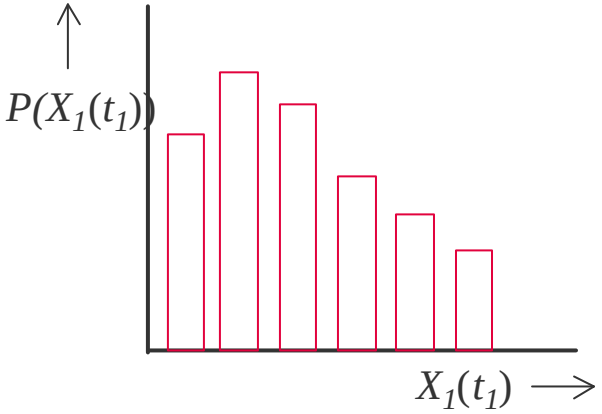
- Simulation starting from same initial state will generally lead to different results



Stochastic simulation and master equation

- Repeating stochastic simulations allows approximation of $P(\boldsymbol{X}(t))$ in master equation to be given





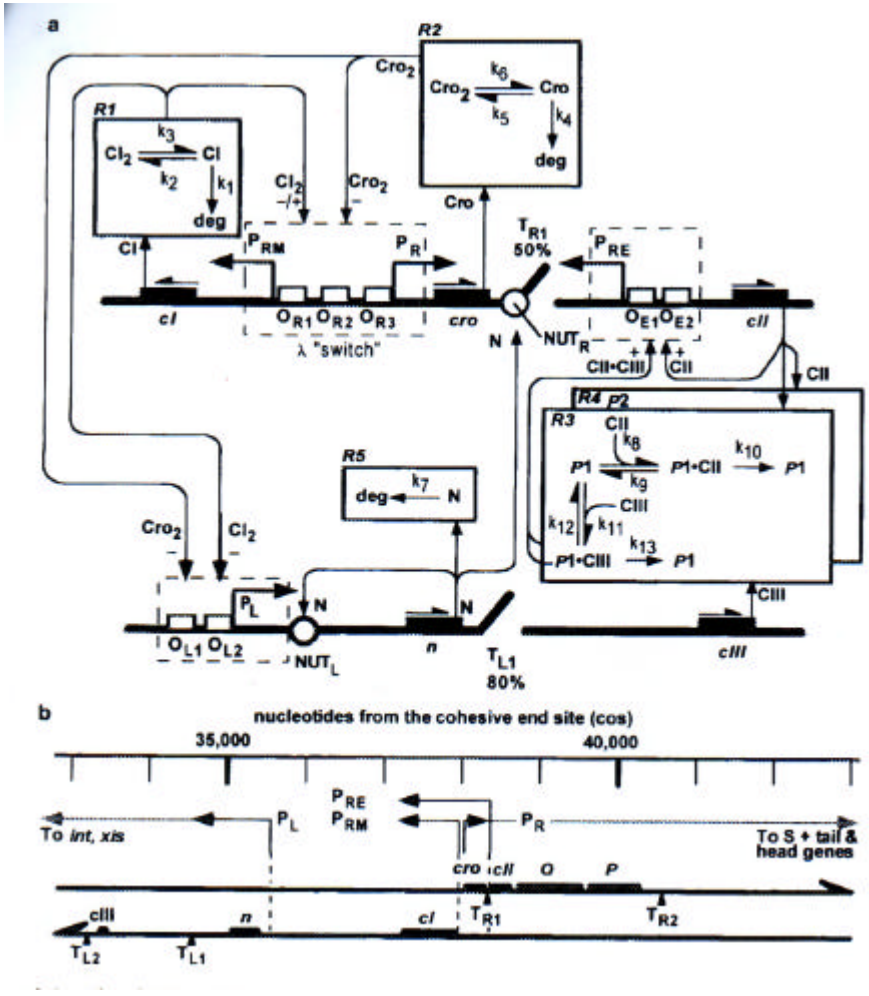
Application of stochastic equations

- 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)
  - ...

Stochastic analysis of phage ? infection

- Stochastic model of ? lysis-lysogeny decision network

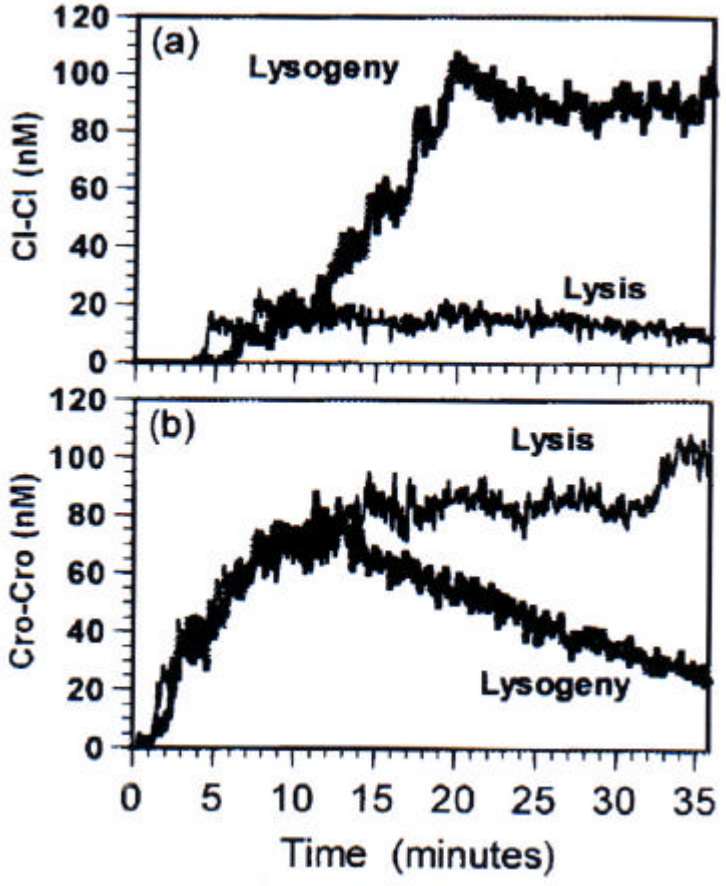
  Arkin *et al.*, 1998



Stochastic analysis of phage ? infection

- Time evolution of Cro and CI dimer concentrations

- Due to stochastic fluctuations, under identical conditions cells follow one or other pathway with some probability



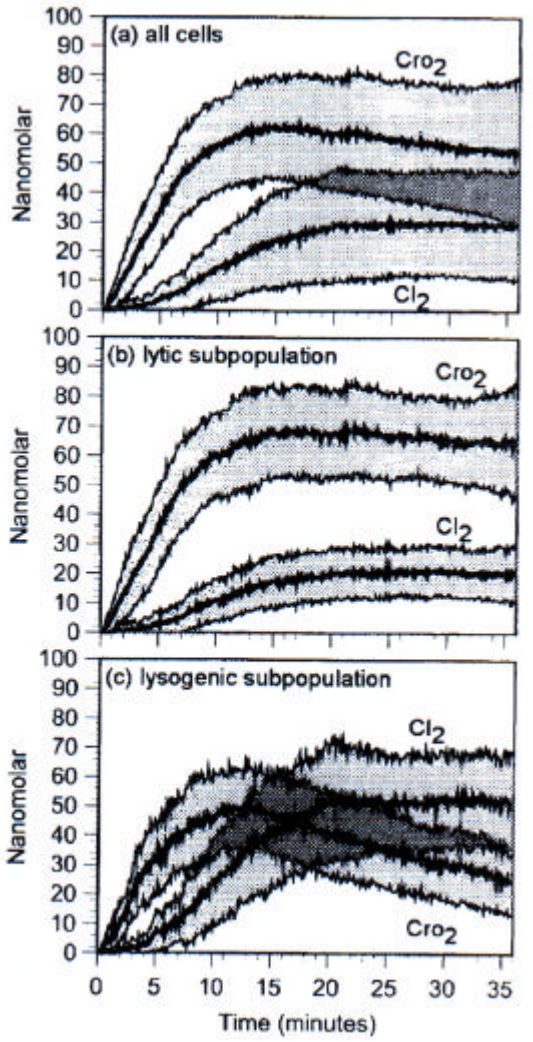
Arkin *et al.*, 1998

Comparison with deterministic approach

- Deterministic models can be seen as predicting **average behavior** of cell population

  Gillespie, 2000

- However, analysis of average behavior may obscure that one part of population chooses one pathway rather than another



Arkin *et al.*, 1998

Evaluation of stochastic equations

- **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 t and r are not or incompletely known

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

  large networks cannot currently be handled

Conclusions

- Computer tools for modeling and simulation will be necessary to understand genetic regulatory processes
- Variety of approaches available, representing genetic regulatory systems on different levels of abstraction
- Choice of approach depends on aim of analysis and on available information:
  - knowledge on reaction mechanisms
  - quantitative data on model parameters and gene expression levels
- Serious applications are beginning to emerge

Literature

A. Arkin *et al.*, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells, *Genetics*, **149**:1633-1648, 1998

J.M. Bower and H. Bolouri, *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, 2001

T.A. Carrier and J.D. Keasling, Investigating autocatalytic gene expression systems through mechanistic modeling, *J. Theor. Biol.*, 201:25-36, 1999

J.L. Cherry and F.R. Adler, How to make a biological switch, *J. Theor. Biol.*, 203:117-133, 2000

Literature

D. Endy *et al.*, Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes, *P. Nat. Acad. Sc. USA*, 97(10):5375-5380, 2000

J. Hasty *et al.*, Computational studies of gene regulatory networks: *in numero* molecular biology, *Nat. Rev. Genet.*, 2(4):268-279, 2001

H. de Jong, Modeling and simulation of genetic regulatory systems: A literature review, *J. Comput. Biol*., 9(1): 69-105, 2002

N. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, 1997

Literature

D. Kaplan and L. Glass, *Understanding Nonlinear Dynamics*, Springer Verlag, 1995

A.D. Keller, Model genetic circuits encoding autoregulatory transcription factors, *J. Theor. Biol.*, 172:169-185, 1995

J.-C. Leloup and A. Goldbeter, A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins, *J. Biol. Rhythms*, **13**(1):70-87, 1998

B. Lewin, *Genes VI*, Oxford University Press, 199

Literature

H. McAdams and A. Arkin, Simulation of prokaryotic genetic circuits*, Annu. Rev. Biophys. Biomol. Struct.*, **27**:199-224, 1998

C. Morton-Firth and D. Bray, Predicting temporal fluctuations in an intracellular signalling pathway, *J. Theor. Biol.*, **192**:117-128, 1998

H. McAdams and L. Shapiro, Circuit simulation of genetic networks, *Science*, **269**:650-656

B. Novak and J. Tyson, Modeling the cell-division cycle: M-phase trigger, oscillations and size control, *J. Theor. Biol.*, **165**:101-134, 1993

M. Ptashne, *A Genetic Switch: Phage ? and Higher Organisms*,

J. Reinitz and D. Sharp, Gene circuits and their uses, in: J. Collado-Vides *et al.* (eds), *Integrative Approaches to Molecular Biology*, MIT Press, 253-272,1996

Literature

M. Santillán and M.C. Mackey, Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data, *P. Nat. Acad. Sc. USA*, 98(4):1364-1369, 2001

L. Segel, *Modeling Dynamic Phenomena in Molecular and Cellular Biology*, Cambridge University Press, 1984

P. Smolen *et al.*, Modeling transcription control in gene networks: Methods, recent results, and future directions, *Bull. Math. Biol.*, **62**:247-292, 2000

S.H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Perseus Books, 1994

R. Thomas and R. d'Ari, *Biological Feedback*, CRC Press, 1990