Modeling and Simulation of Genetic Regulatory Networks

Hidde de Jong

INRIA Grenoble - Rhône-Alpes

Laboratoire Adaptation Pathogénie des Microorganismes, CNRS UMR 5163, Université Joseph Fourier, Grenoble

Hidde.de-Jong@inria.fr http://ibis.inrialpes.fr

UNIVERSITE JOSEPH FOURIER SCIENCES TECHNOLOGIE MEDECINE

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE INRIA centre de recherche GRENOBLE - RHÔNE-ALPES

INRIA Grenoble - Rhône-Alpes and IBIS

- IBIS: systems biology group of INRIA and Joseph Fourier University/CNRS
  - Analysis of bacterial regulatory networks by means of models and experiments
  - Involves computer scientists, molecular biologists, physicists, …

Overview

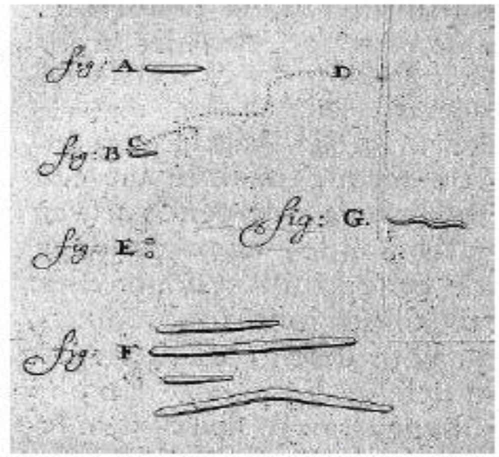
1. Genetic regulatory networks in bacteria
2. Motivations for modeling and simulation
3. Approaches towards modeling and simulation
   - Ordinary differential equations
   - Stochastic master equations

4. Conclusions

Bacteria

- Bacteria were first observed by Antonie van Leeuwenhoek in 1676, using a single-lens microscope of his own design

http://commons.wikimedia.org/



www.euronet.nl/users/warnar/leeuwenhoek.html.

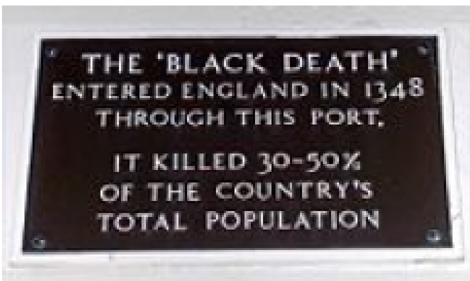
van Leeuwenhoek A (1684), *Philosophical Transactions (1683–1775)* 14: 568–574

"In the morning I used to rub my teeth with salt and rinse my mouth with water and after eating to clean my molars with a toothpick.... I then most always saw, with great wonder, that in the said matter there were many very ***little living animalcules****, very prettily a-moving. The biggest sort had a very strong and swift motion, and shot through the water like a pike does through the water; mostly these were of small numbers."*

Impact of bacteria on humans

- Bacteria as disease agents

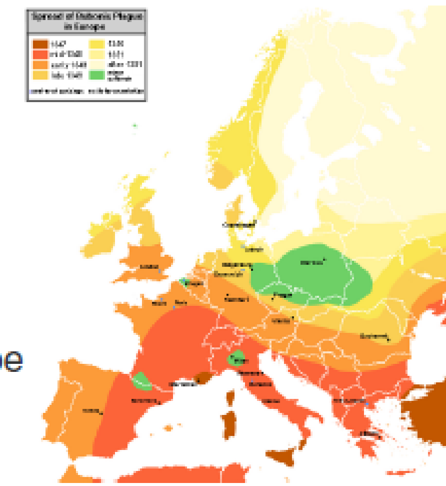
  Tuberculosis, cholera, syphilis, anthrax, leprosy, bubonic plague, …



http://en.wikipedia.org/wiki/Black_Death

Plaque in Weymouth, England

Spread of bulbonic plague over Europe



http://en.wikipedia.org/wiki/Plague_(disease)

- Bacteria in food industry (fermentation)

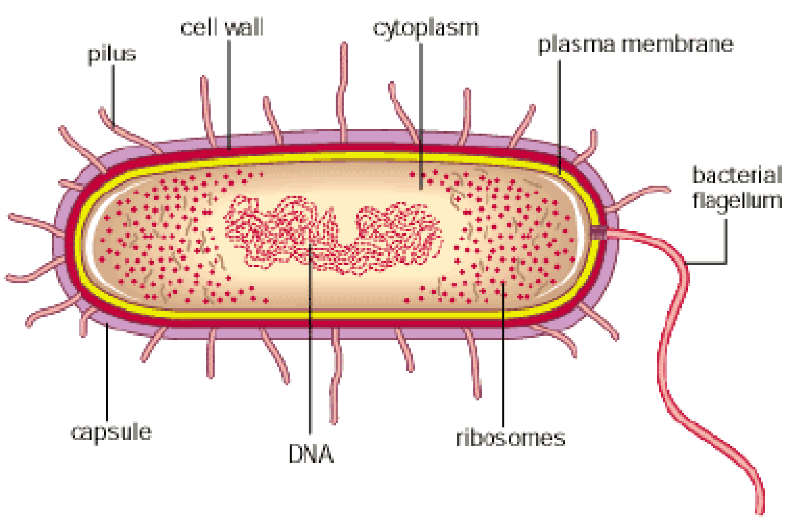
  Cheese, yoghurt, …

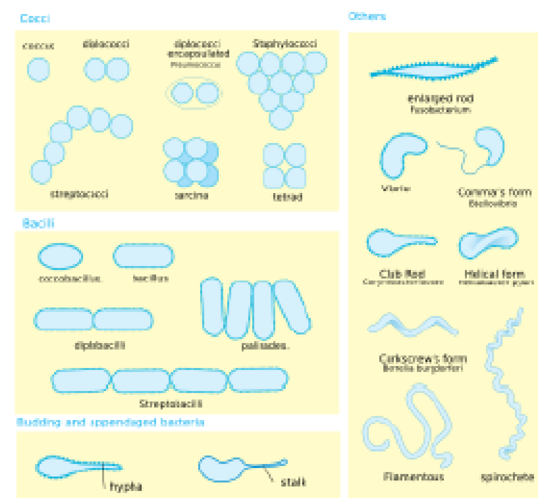
- Bacteria in environmental and biotechnology

  Sewage treatment, bioremediation, synthesis of chemicals, biofuels, …

Bacterial cells

- 40 ·10^6 bacterial cells in 1 g of soil and 10^6 bacterial cells in 1 ml of fresh water
- 10 times as many bacterial cells as human cells in human body
- Wide range of shapes (spheres, rods, spirals, …), typically 0.5–5.0 um in length

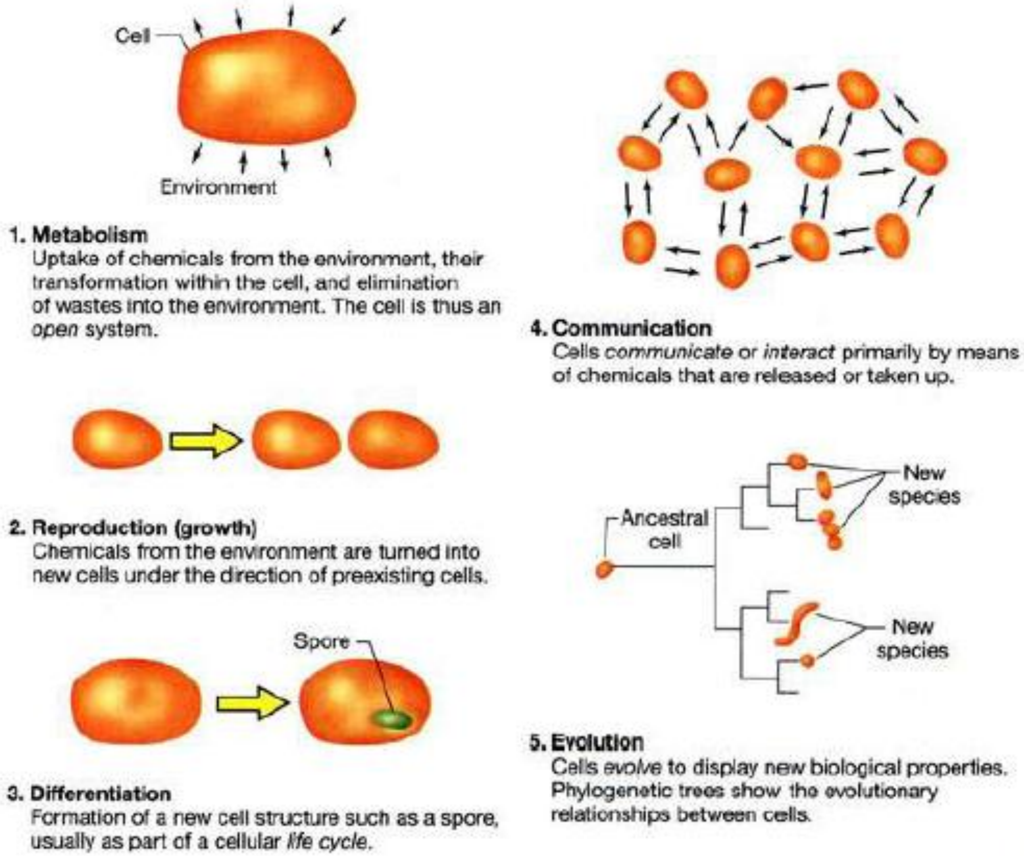




Madigan *et al.* (2003), *Brock Biology of Microorganisms*, Prentice Hall, 10th ed.

Bacteria as living systems

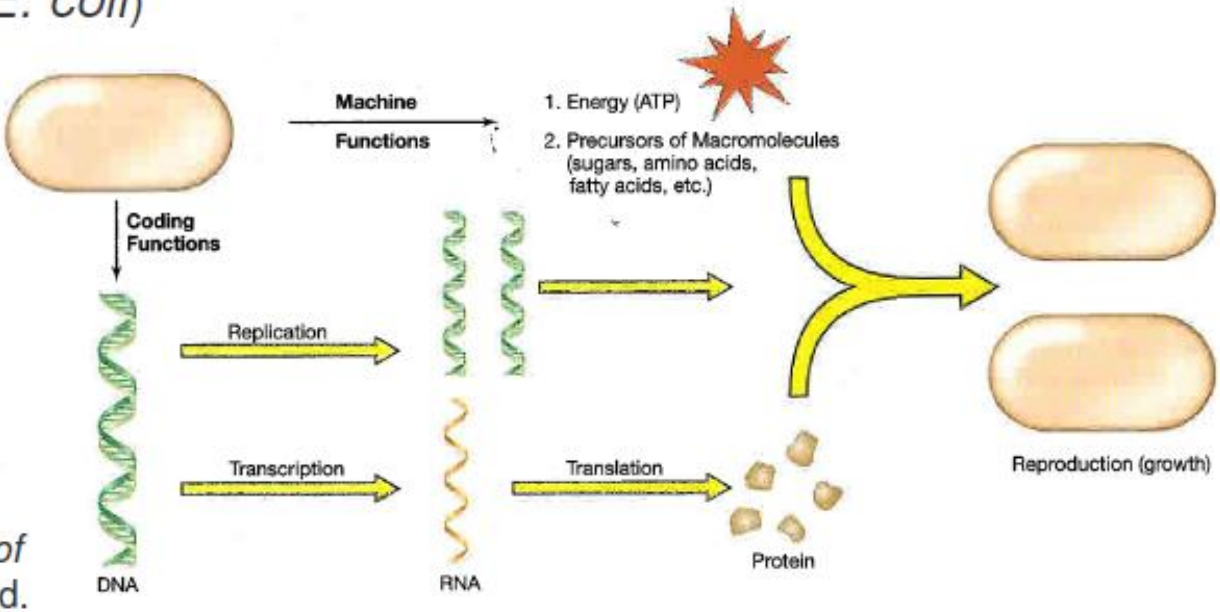
- Bacteria possess characteristics shared by most living systems



Madigan *et al*. (2003), *Brock Biology of Microorganisms*, Prentice Hall, 10th ed.

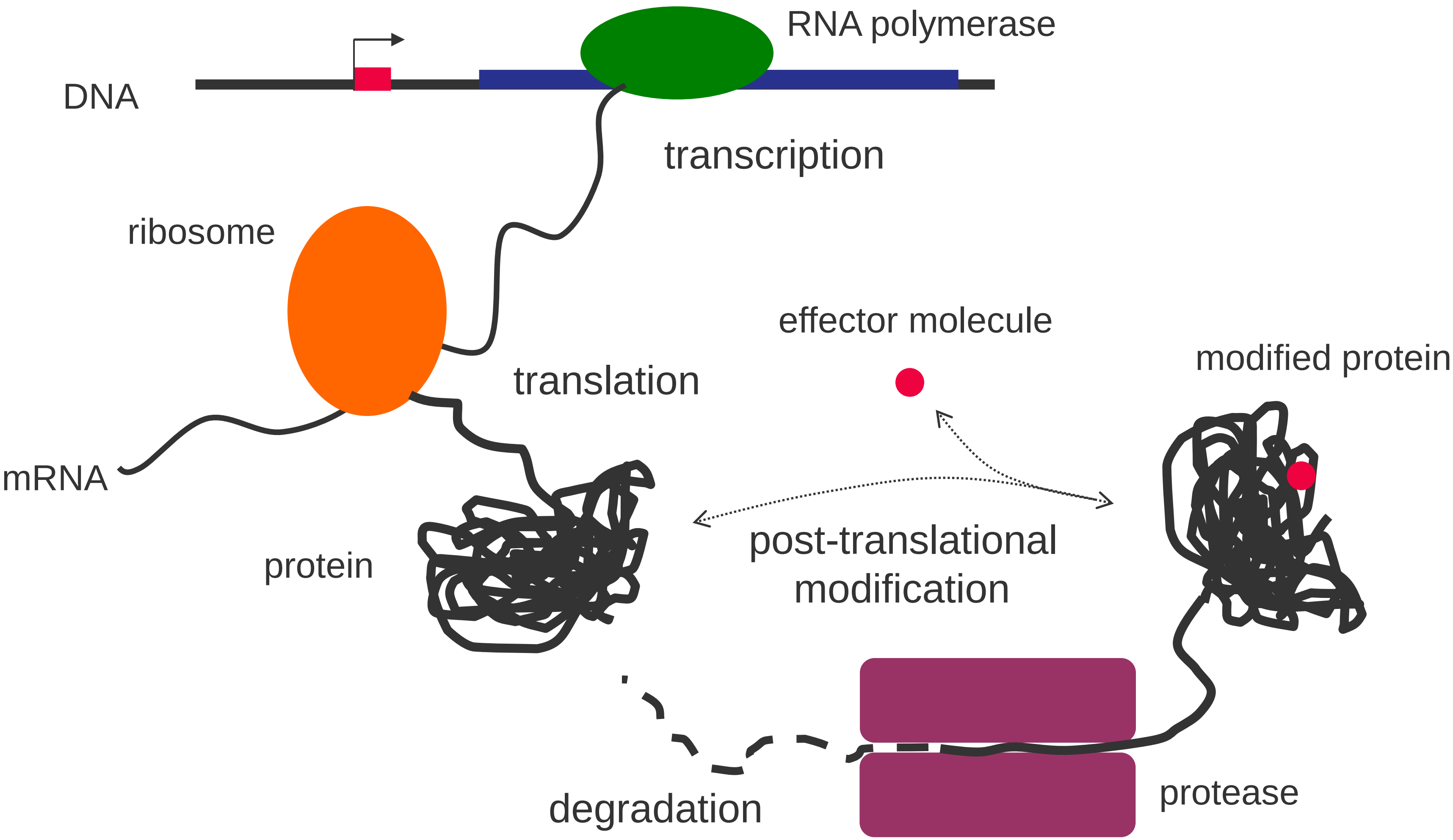
Proteins are building blocks of cell

- Proteins are essential building blocks in machine and coding functions of the cell
  - Cells as **biochemical machines** and as **coding devices**
  - Single cell contains 1900 different kinds of proteins and $2.4 \cdot 10^6$ total protein molecules (*E. coli*)

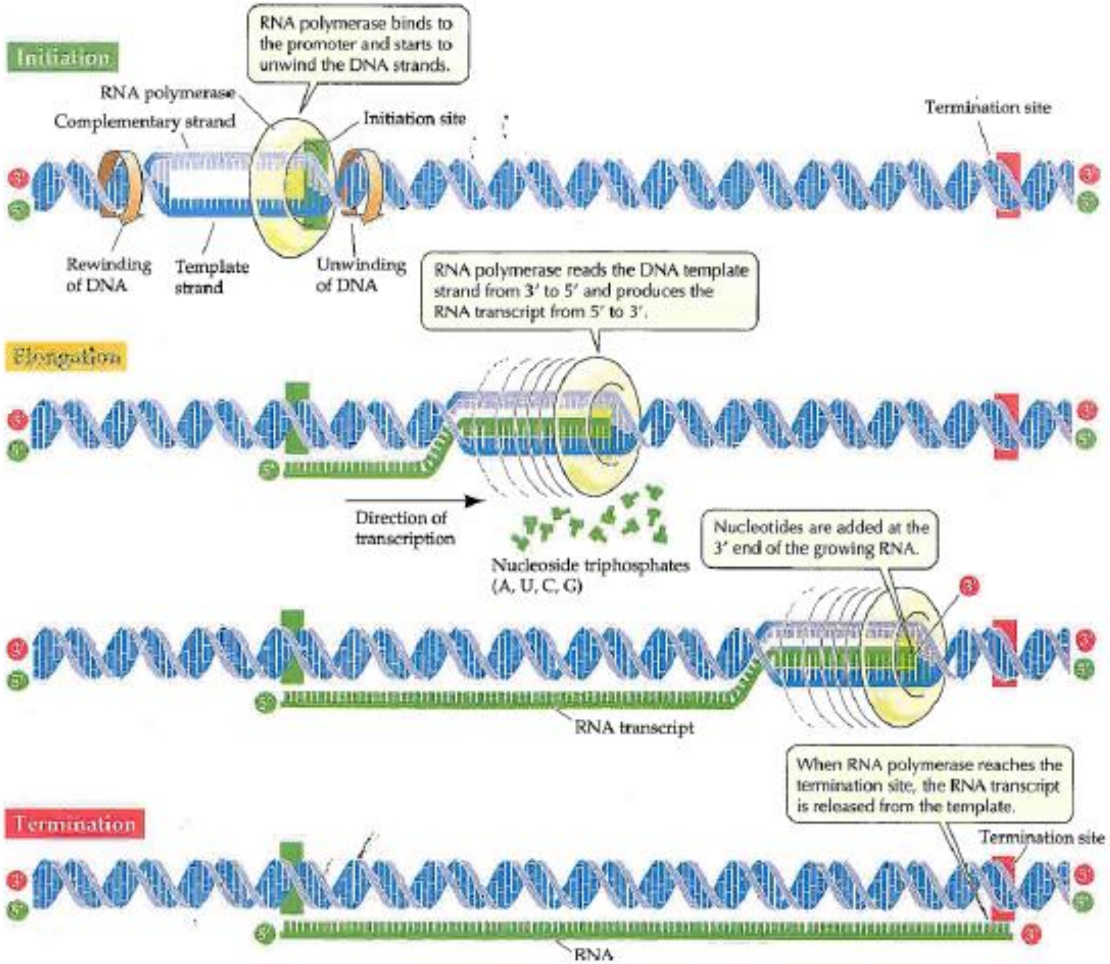


Madigan *et al.* (2003), *Brock Biology of Microorganisms*, Prentice Hall, 10th ed.

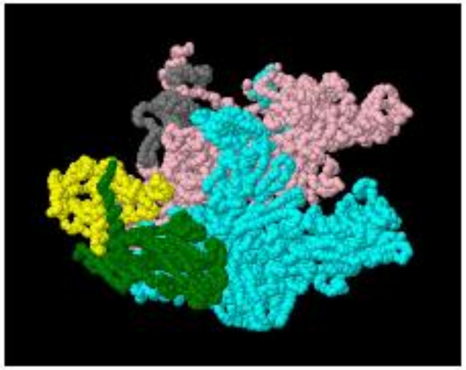
Synthesis and degradation of proteins



Transcription



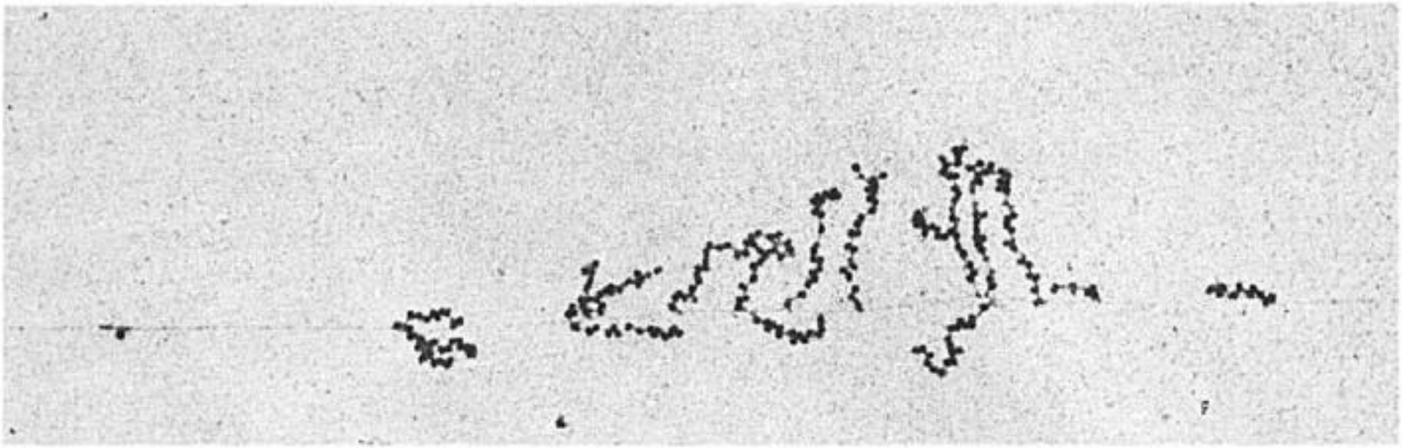
RNA polymerase



http://www.steve.gb.com/science/transcription.html

Purves *et al.* (2003), *Life*

Synthesis and degradation of proteins



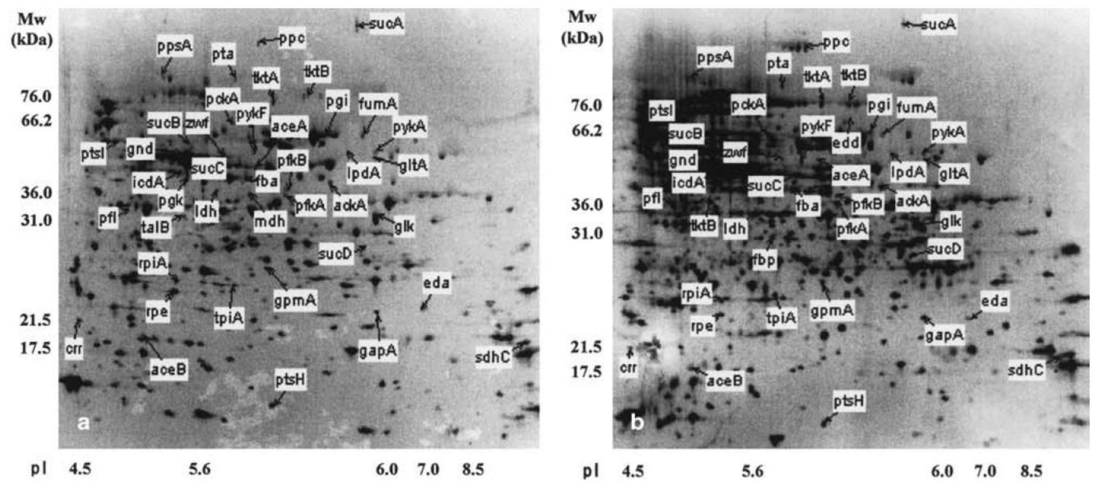
http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/chroms-genes-prots/transcription-translation.html

Variation in protein levels

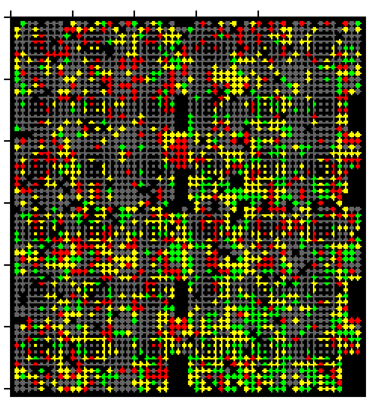
❖ Protein levels in cell are adjusted to specific environmental conditions

Peng, Shimizu (2003),
App. Microbiol. Biotechnol., 61:163-178

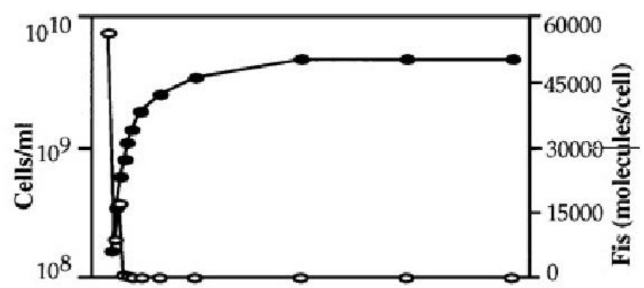
2D gels



DNA microarrays

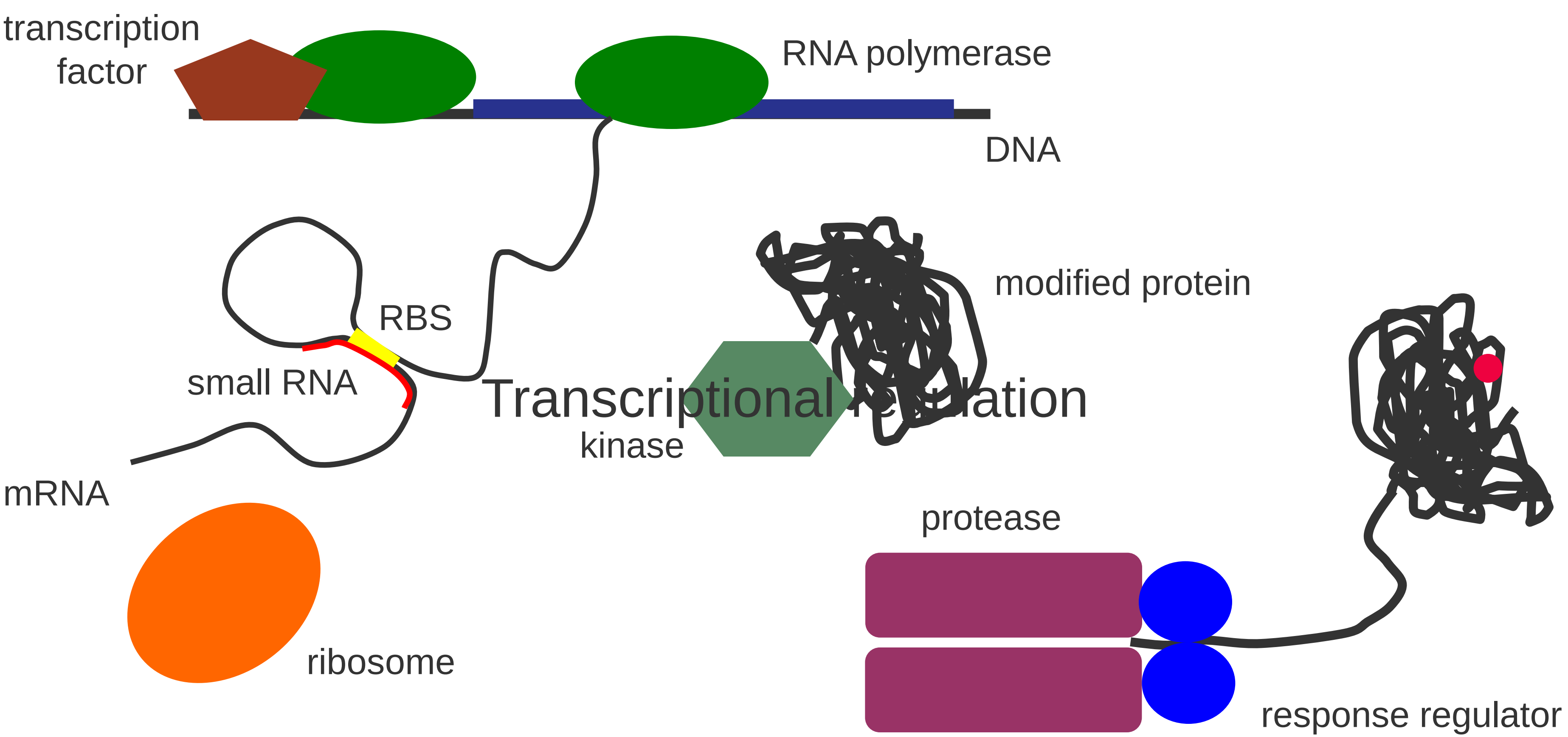


Western blots



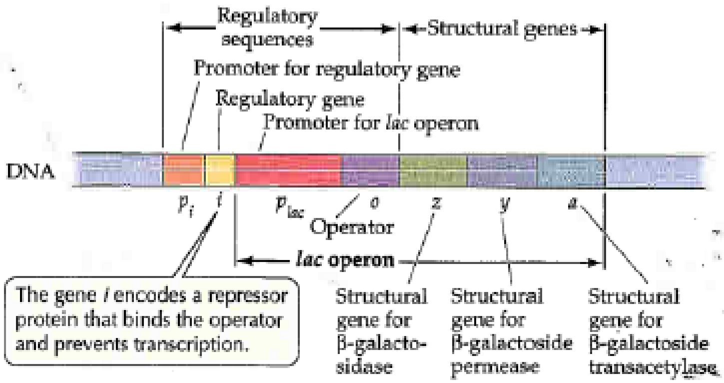
Ali Azam *et al.* (1999), *J. Bacteriol.*, 181(20):6361-6370

Regulation of synthesis and degradation



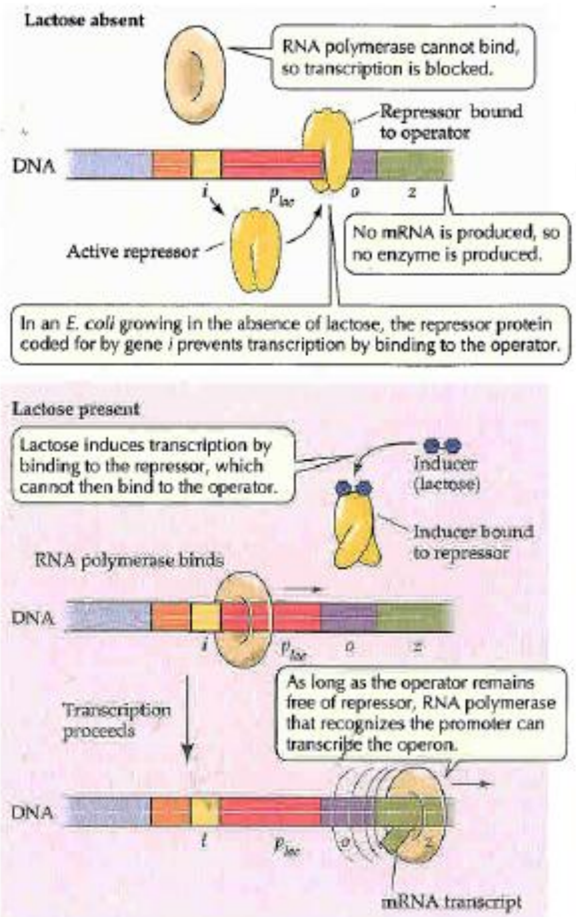
❖ Mostly transcriptional regulation in bacteria, but sometimes regulation on all four levels

Transcriptional regulation

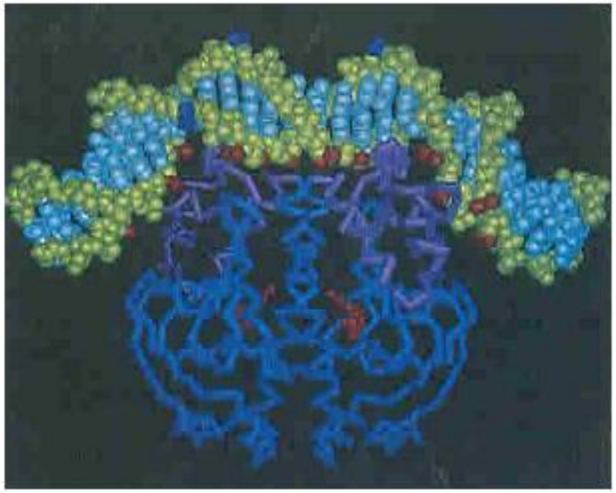


Purves *et al.* (2003), *Life*

Transcriptional regulation



Transcription regulator Crp bound to DNA



Purves *et al.* (2003), *Life*

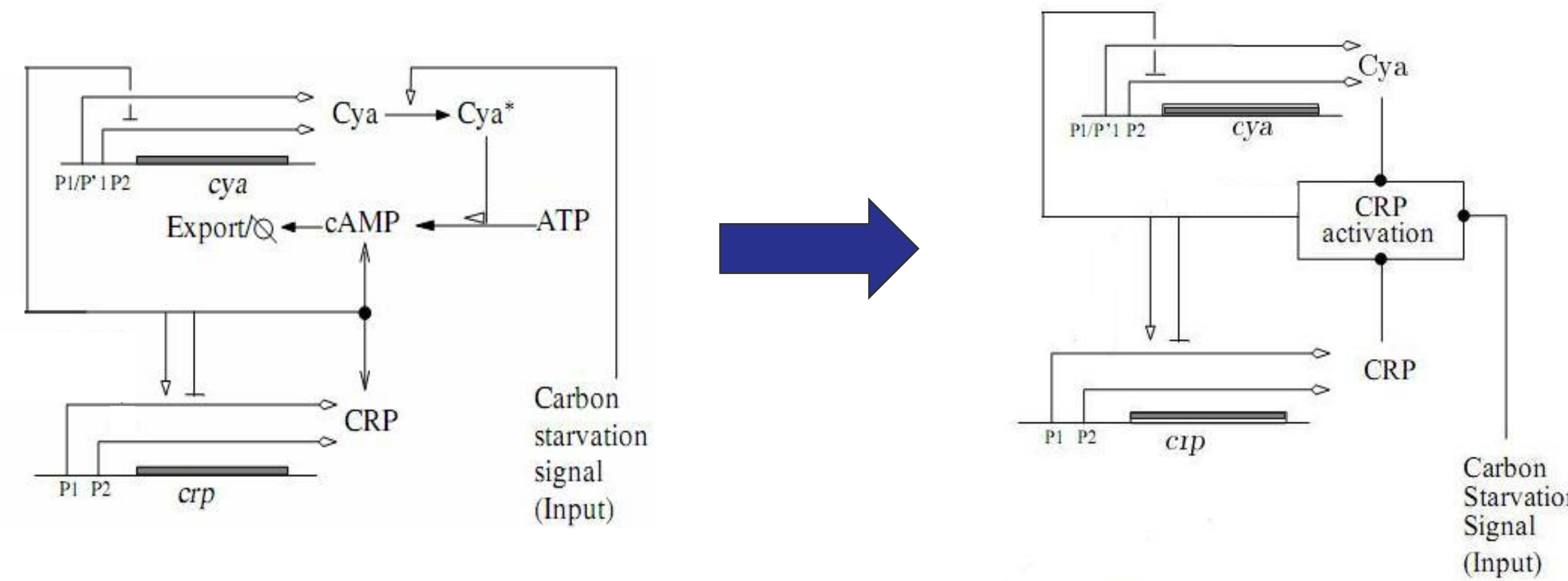
Genetic regulatory networks

- Regulation of synthesis and degradation of proteins is achieved by other proteins/protein complexes

  Transcription regulators, proteases, but also ribozomes, RNA polymerases

- Direct and indirect regulatory interactions give rise to **genetic regulatory network**

Brazhnik et al. (2002), *Trends Biotechnol.*, 20(11):467-72

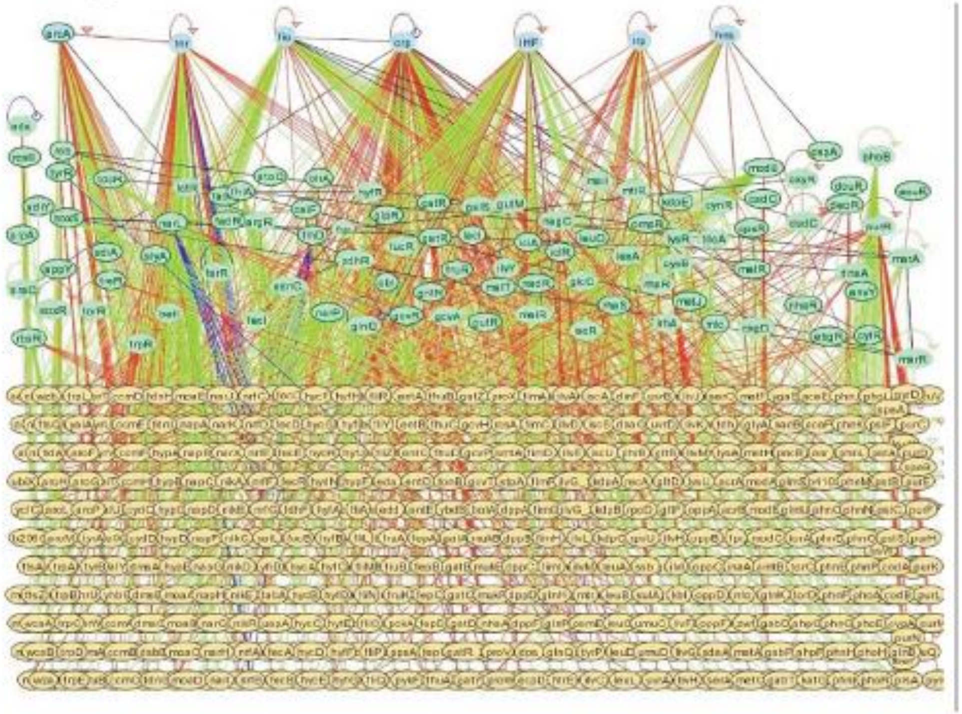


Complexity of genetic regulatory networks

- Most genetic regulatory networks of biological interest are large and complex

  E. coli has 4200 genes coding for several hundreds of transcription factors

Network of transcription regulators in *E. coli*



Martinez-Antonio *et al.* (2003), *Curr. Opin. Microbiol.*, 6(5):482-489

Analysis of genetic regulatory networks

❖ Abundant knowledge on components and interactions of genetic regulatory networks in many bacteria

- Scientific knowledge bases and databases
- Bibliographic databases

❖ Currently little understanding of how global dynamics emerges from local interactions between components

- Response of cell to external perturbation
- Differentiation of cell during development

❖ Shift from **structure** to **dynamics** of networks

« functional genomics », « integrative biology », « systems biology », …

Kitano (2002), *Science*, 295(5560):564

Experimental tools

- Study of dynamics of genetic regulatory networks requires powerful experimental tools

  High-throughput, low-cost, reliable, precise, …

- Different methods for monitoring gene expression, measuring different quantities:
  - Western blots: protein abundance
  - Northern blots: (relative) mRNA abundance
  - DNA microarrays: (relative) mRNA abundance
  - Reporter genes: promoter activity, mRNA abundance
  - ….
- Measurements on population of cells, more recently also measurements on individual cells

Longo and Hasty (2006), *Mol. Syst. Biol.*, msb410011

Mathematical methods and computer tools

- **Modeling** and **simulation** indispensable for dynamic analysis of genetic regulatory networks:
  - understanding role of individual components and interactions
  - suggesting missing components and interactions
- **Mathematical methods** supported by **computer tools** required for modeling and simulation:
  - precise and unambiguous description of network
  - systematic derivation of behavior predictions
- First models of genetic regulatory networks date back to early days of molecular biology

Regulation of *lac* operon

Goodwin (1963), *Temporal Organization in Cells*

Hierarchy of modeling formalisms

- Variety of modeling formalisms exist, describing system on different levels of detail

de Jong (2002), *J. Comput. Biol.*, 9(1): 69-105

Hasty *et al.* (2001), *Nat. Rev. Genet.*, 2(4):268-279

Smolen *et al.* (2000), *Bull. Math. Biol.*, 62(2):247-292

Szallassi et al. (2006), *System Modeling in Cellular Biology*, MIT Press

Boolean networks

↑

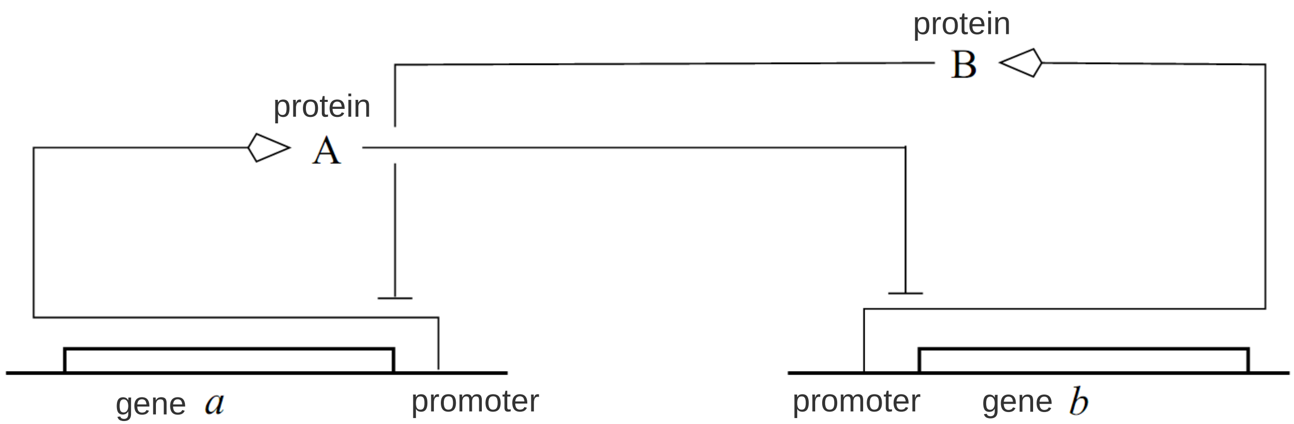
precision ↓ Ordinary differential equations ↑ **abstraction**

↑

Stochastic master equations

Cross-inhibition network

- **Cross-inhibition** network consists of two genes, each coding for transcription regulator inhibiting expression of other gene



- Cross-inhibition network is example of **positive feedback**, important for phenotypic differentiation (multi-stability)

Thomas and d'Ari (1990), *Biological Feedback*

Ordinary 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, controlling synthesis and degradation, modeled by **ordinary differential equations**

$$\frac{d\boldsymbol{x}}{dt} = \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}),$$

where $\boldsymbol{x} = [x_1, \ldots, x_n]'$ and $\boldsymbol{f}(\boldsymbol{x})$ is **rate law**
- Kinetic theory of biochemical reactions provides basis for specification of rate law

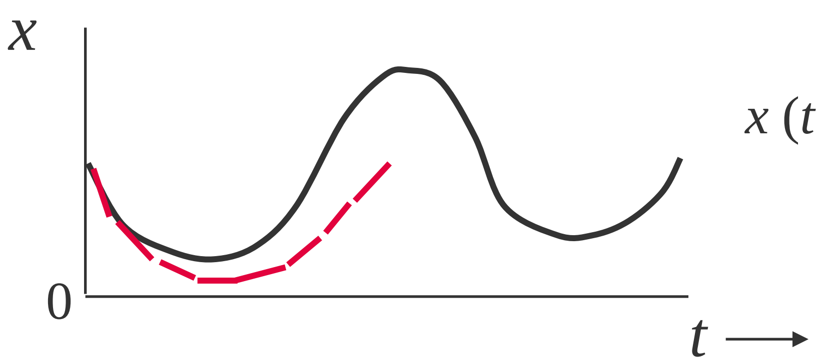
Heinrich and Schuster (1996), *The Regulation of Cellular Systems*
Cornish-Bowden (1995), *Fundamentals of Enzyme Kinetics*

Analysis and numerical simulation

- No analytical solution for most nonlinear differential equations
- **Dynamic systems theory** provides techniques for analysis of nonlinear differential equations, but usually not scalable
  - Phase portrait
  - Bifurcation analysis

Kaplan and Glass (1995), *Understanding Nonlinear Dynamics*

- Approximation of solution obtained by **numerical simulation**, given parameter values and initial conditions $\boldsymbol{x}(0) = \boldsymbol{x}^0$

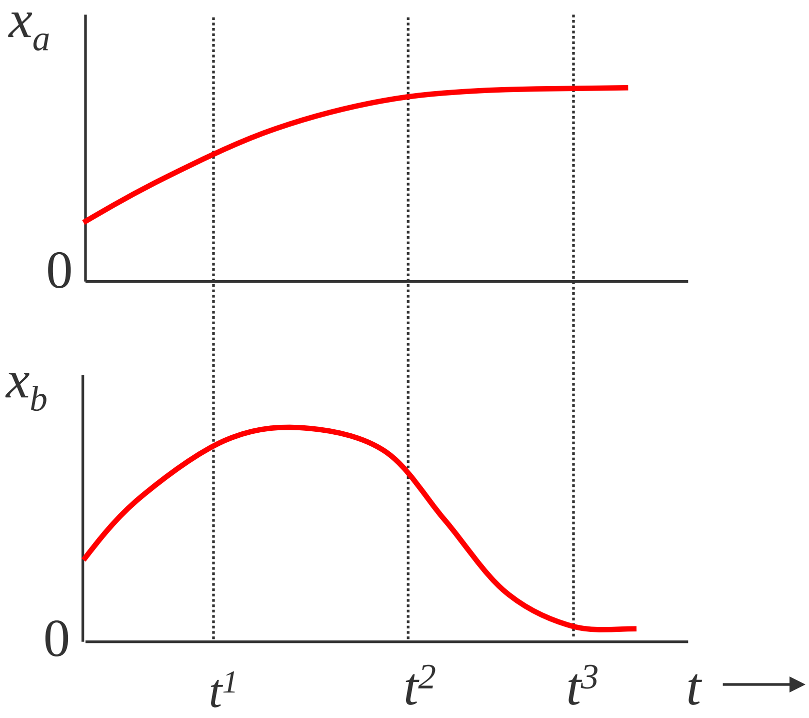


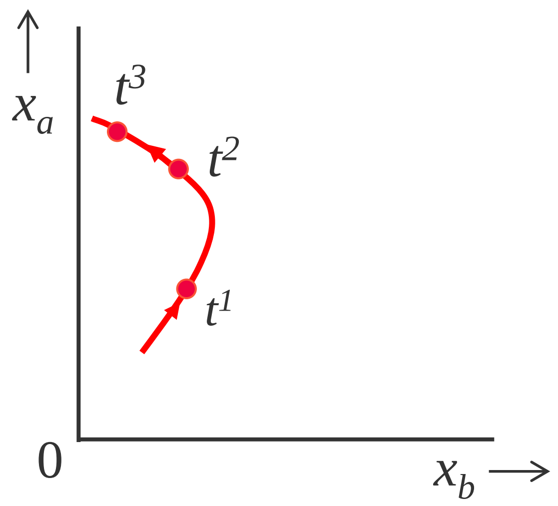
$$x(t+\Delta t) = x(t) + \int_t^{t+\Delta t} f(x)\, dt \approx x(t) + f(x)\,\Delta t$$

Lambert (1991), *Numerical Methods for Ordinary Differential Equations*

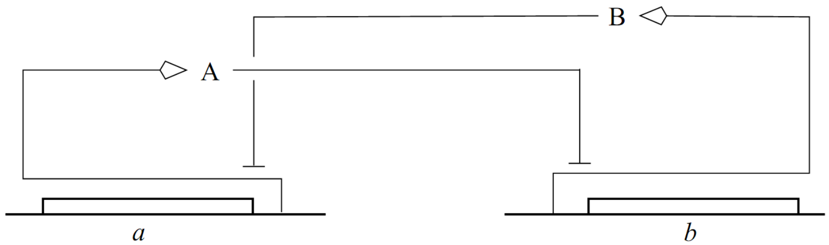
Solution trajectories in phase plane

- Representation of solutions in phase plane yields **solution trajectories**





ODE model of cross-inhibition network



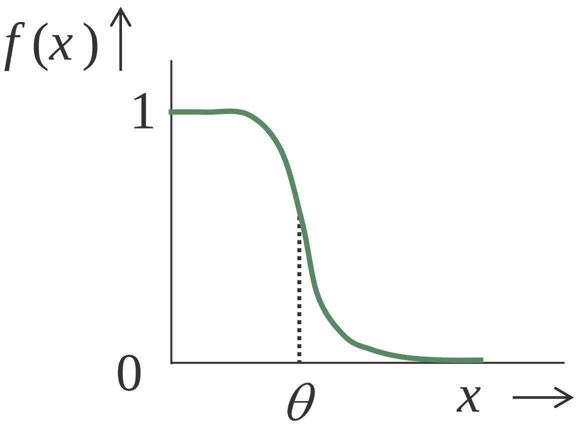
$$\dot{x}_a = \kappa_a f(x_b) - \gamma_a x_a$$
$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$

x_a = concentration protein A

x_b = concentration protein B

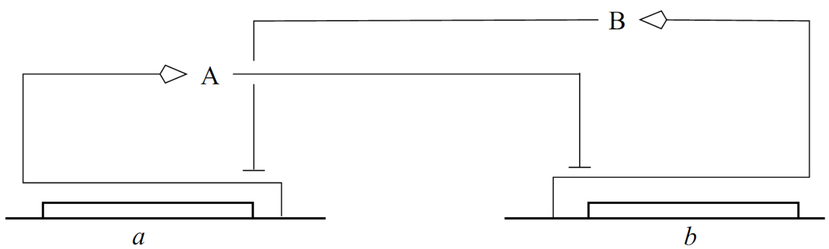
$\kappa_a, \kappa_b > 0$, production rate constants

$\gamma_a, \gamma_b > 0$, degradation rate constants



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

ODE model of cross-inhibition network



$$\dot{x}_a = \kappa_a f(x_b) - \gamma_a x_a$$
$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$

x_a = concentration protein A

x_b = concentration protein B

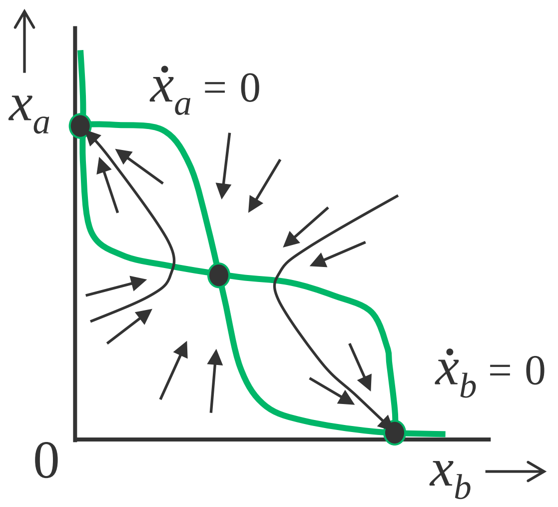
$\kappa_a, \kappa_b > 0$, production rate constants

$\gamma_a, \gamma_b > 0$, degradation rate constants

- Implicit modeling assumptions:
  - Ignore intermediate gene products (mRNA)
  - Ignore gene expression machinery (RNA polymerase, ribosome)
  - Simplification of complex interactions of regulators with DNA to single response function

Bistability of cross-inhibition network

❖ Analysis of **steady states** in phase plane



$$\dot{x}_a = 0: \; x_a = \frac{\kappa_a}{\gamma_a} f(x_b)$$

$$\dot{x}_b = 0: \; x_b = \frac{\kappa_b}{\gamma_b} f(x_a)$$

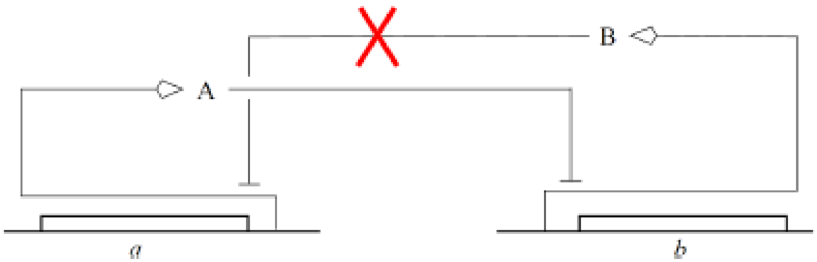
❖ System is **bistable**: two stable and one unstable steady state.

❖ For almost all initial conditions, system will converge to one of two stable steady states (**differentiation**)

❖ System returns to steady state after small perturbation

Switching in cross-inhibition network

- Transient perturbation may cause irreversible **switch** from one steady state to the other
  - Temporary disable one of the inhibitors



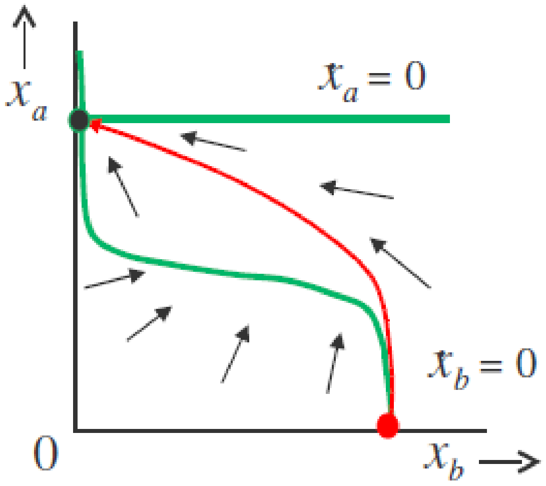
$$\dot{x}_a = \kappa_a - \gamma_a x_a$$

$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$

Switching in cross-inhibition network

- Transient perturbation may cause irreversible **switch** from one steady state to the other

System evolves to new steady state

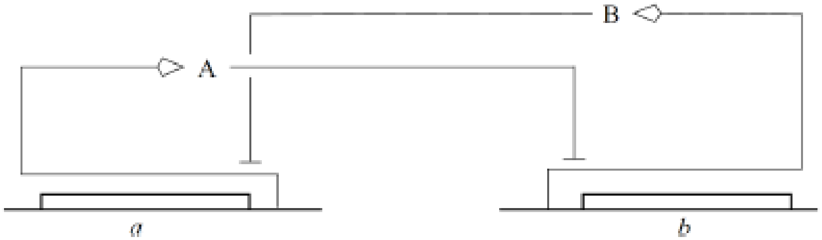


$$\dot{x}_a = 0: \ x_a = \frac{\kappa_a}{\gamma_a}$$

$$\dot{x}_b = 0: \ x_b = \frac{\kappa_b}{\gamma_b} f(x_a)$$

Switching in cross-inhibition network

- Transient perturbation may cause irreversible **switch** from one steady state to the other
  - Enable again inhibitor



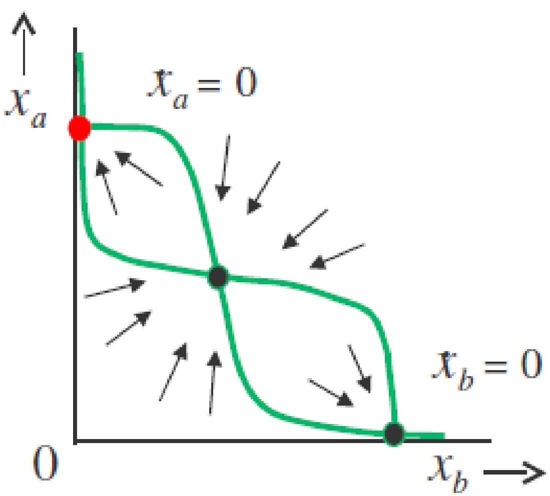
$$x_a = \kappa_a f(x_b) - \gamma_a x_a$$

$$x_b = \kappa_b f(x_a) - \gamma_b x_b$$

Switching in cross-inhibition network

- Transient perturbation may cause irreversible **switch** from one steady state to the other

System remains in new steady state



$$\dot{x}_a = 0 : \; x_a = \frac{\kappa_a}{\gamma_a} f(x_b)$$

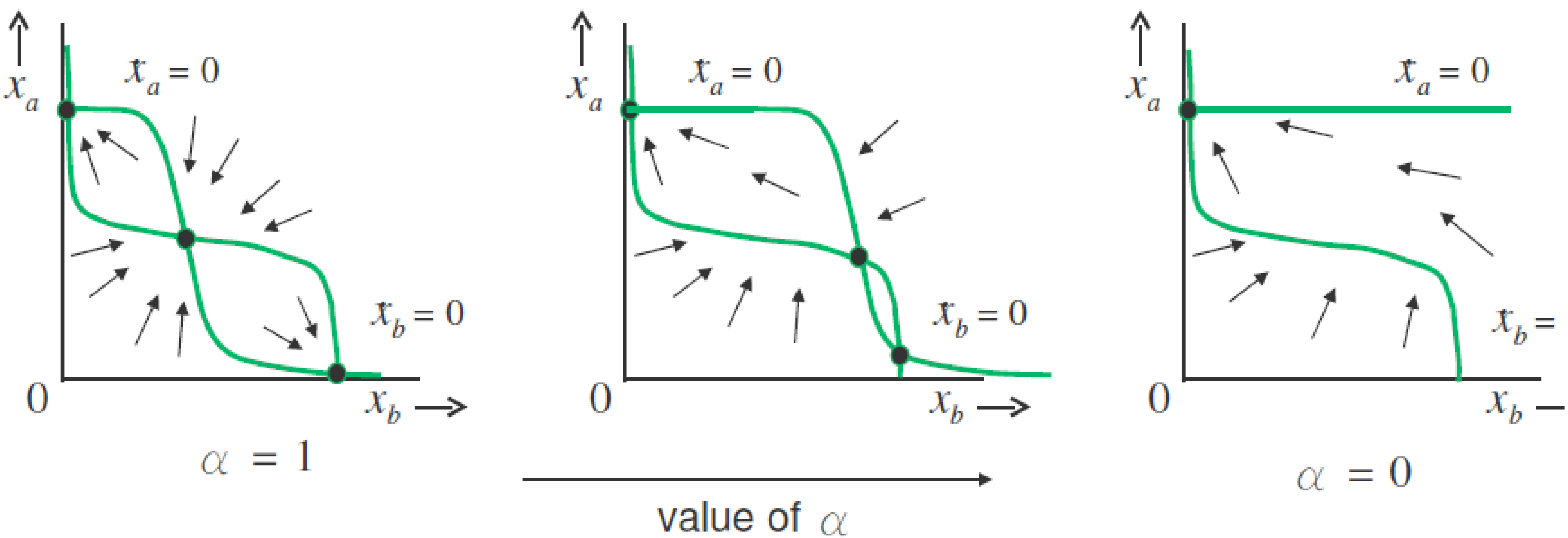
$$\dot{x}_b = 0 : \; x_b = \frac{\kappa_b}{\gamma_b} f(x_a)$$

Bifurcation in cross-inhibition network

- Switching of cross-inhibition network can be interpreted as sequence of **bifurcations**, induced by change in parameter

$$x_a = \kappa_a f(\alpha\ x_b) - \gamma_a x_a$$

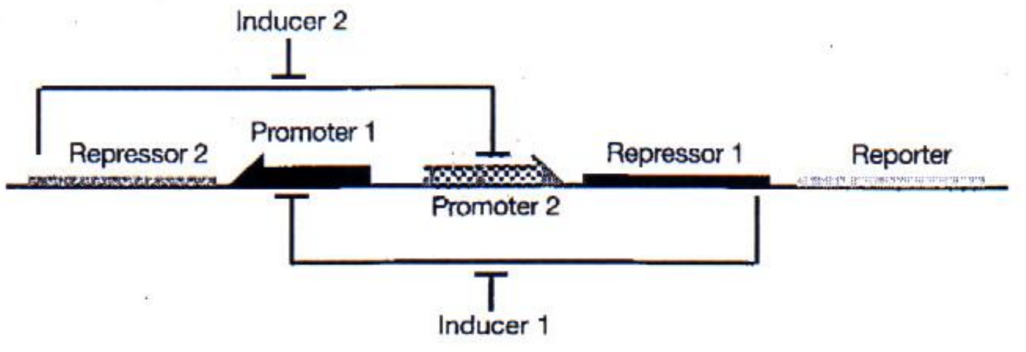
$$x_b = \kappa_b f(x_a) - \gamma_b x_b$$

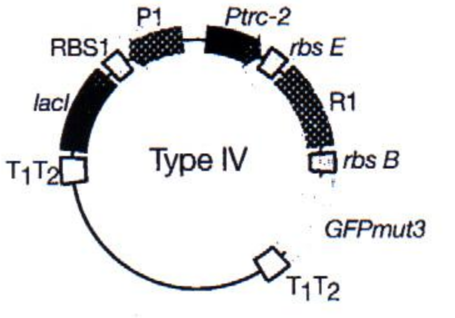


Construction of cross inhibition network

- Construction of cross inhibition network *in vivo*

Gardner *et al.* (2000), *Nature*, 403(6786): 339-342





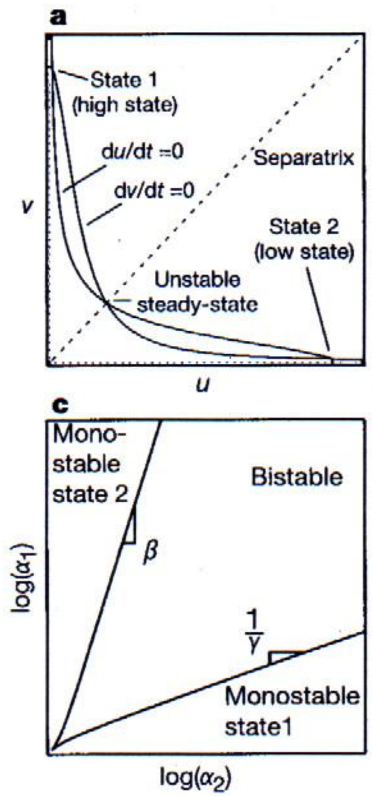
- Differential equation model of network

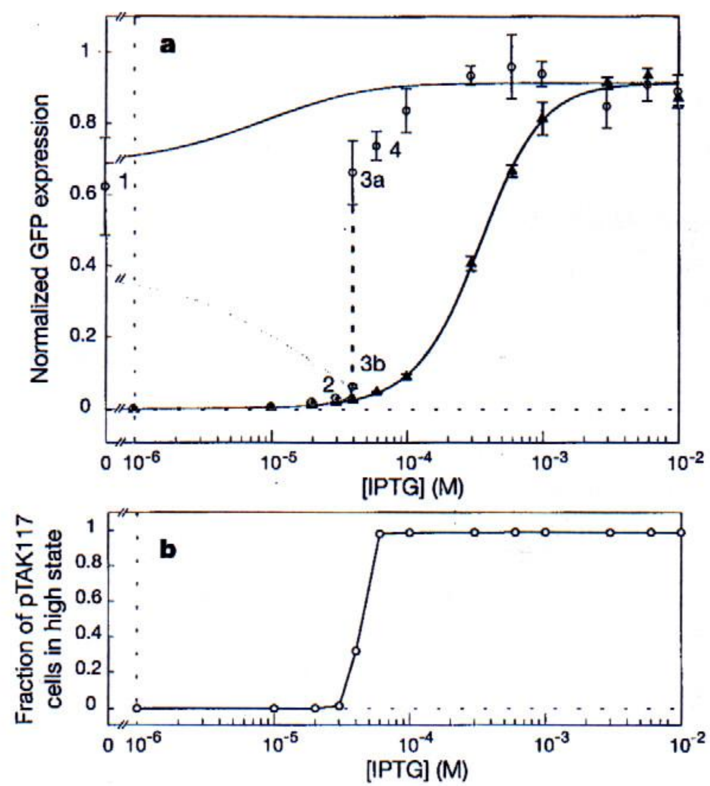
$$\dot{u} = \frac{\alpha_1}{1 + v^{\beta}} - u \qquad \dot{v} = \frac{\alpha_2}{1 + u^{\gamma}} - v$$

Experimental test of model

- Experimental test of mathematical model (bistability and hysteresis)

Gardner *et al.* (2000), *Nature*, 403(6786): 339-342

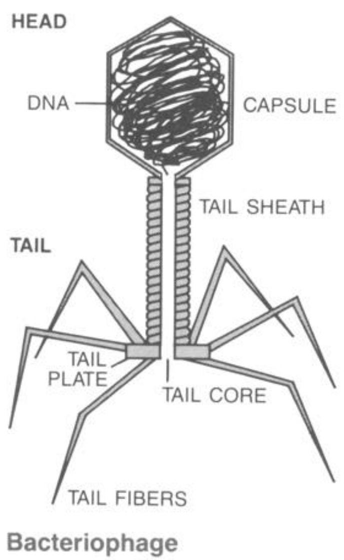


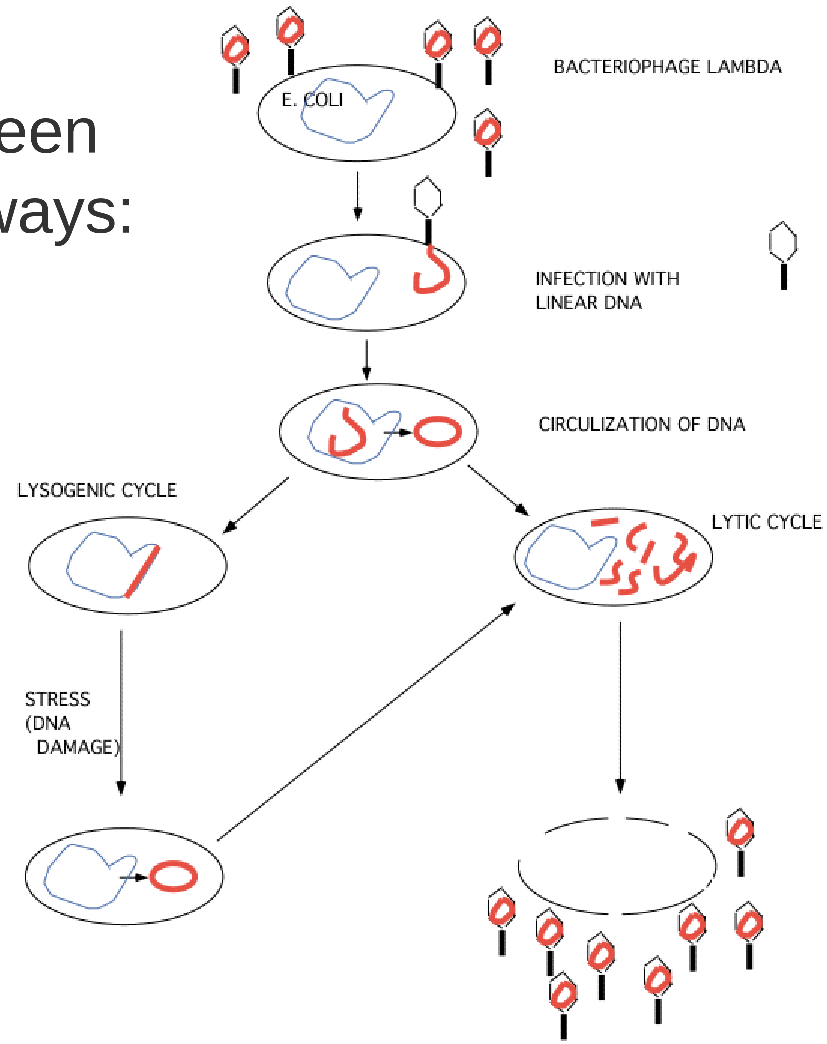


Bacteriophage λ infection of *E. coli*

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

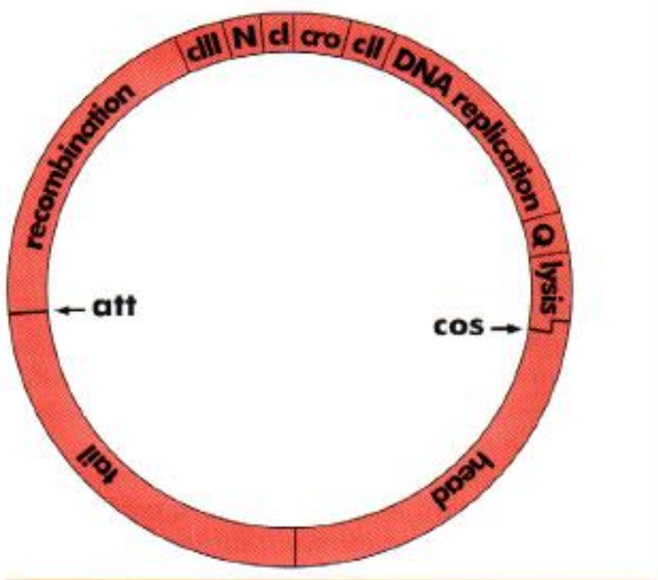
Ptashne, *A Genetic Switch*, Cell Press,1992



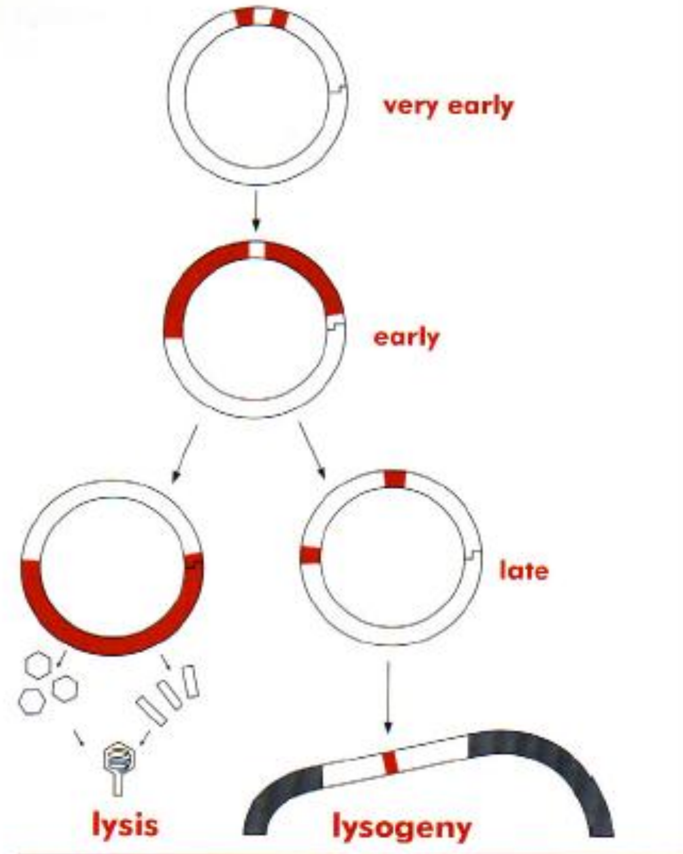


Bistability in phage λ

- Lytic and lysogenic pathways involve different patterns of gene expression

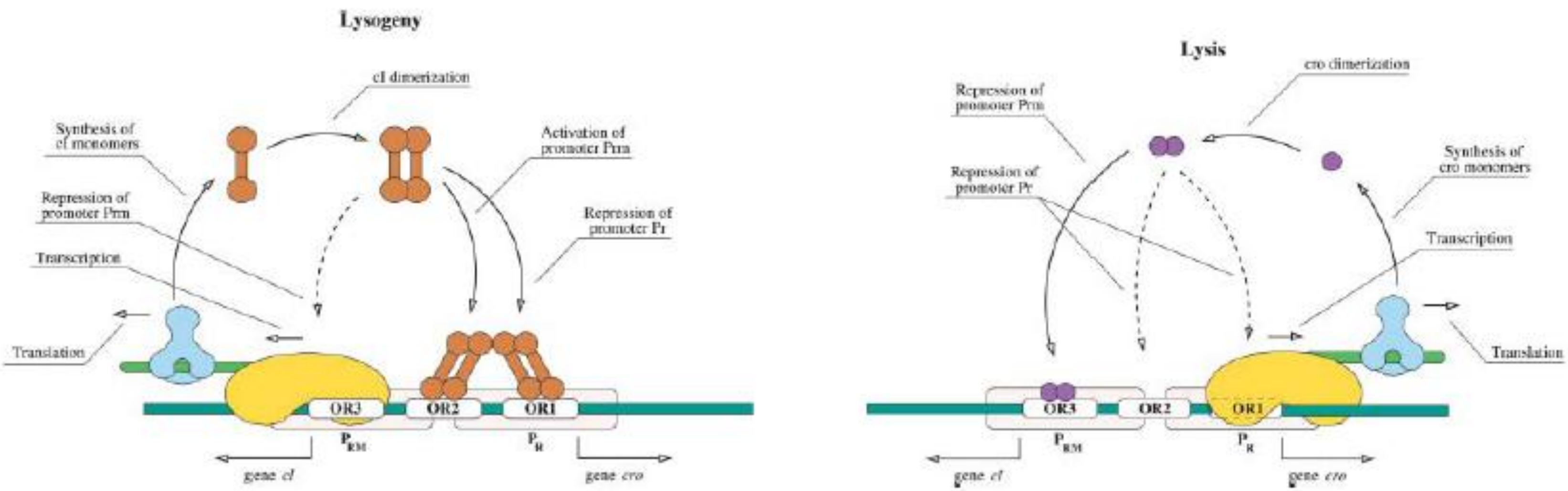


Ptashne, *A Genetic Switch*, Cell Press,1992



Control of phage λ fate decision

- Cross-inhibition feedback plays key role in establishment of lysis or lysogeny, as well as in induction of lysis after DNA damage



Santillán, Mackey (2004), *Biophys. J.*, 86(1): 75-84

Simple model of phage λ fate decision

- Differential equation model of cross-inhibition feedback network involved in phage λ fate decision

  mRNA and protein, delays, thermodynamic description of gene regulation

$$\frac{d[M_{\mathrm{cI}}]}{dt} = k^q_{\mathrm{cI}}[O_{\mathrm{R}}]f^q_{\mathrm{RM}}([CI_2]_{\tau_{\mathrm{M}}}, [CI_2]_{\tau_{\mathrm{M}}}) + k^s_{\mathrm{cI}}[O_{\mathrm{R}}]f^s_{\mathrm{RM}}([CI_2]_{\tau_{\mathrm{M}}}, [Cro_2]_{\tau_{\mathrm{M}}}) - (\gamma_{\mathrm{M}} + \mu)[M_{\mathrm{cI}}],$$

$$\frac{d[M_{\mathrm{cro}}]}{dt} = k_{\mathrm{cro}}[O_{\mathrm{R}}]f_{\mathrm{R}}([CI_2]_{\tau_{\mathrm{M}}}) - (\gamma_{\mathrm{M}} + \mu)[M_{\mathrm{cro}}],$$

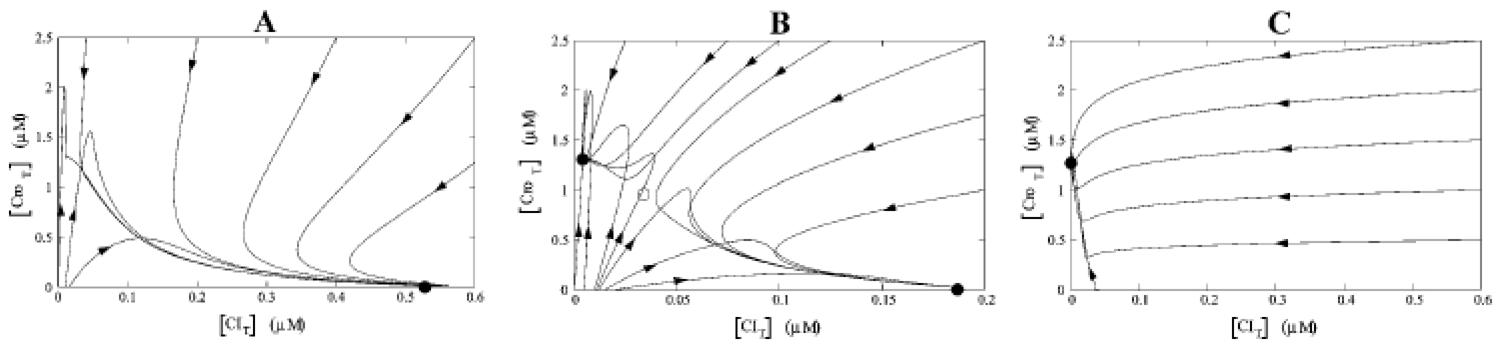
$$\frac{d[CI_{\mathrm{T}}]}{dt} = \upsilon_{\mathrm{cI}}[M_{\mathrm{cI}}]_{\tau_{\mathrm{cI}}} - (\gamma_{\mathrm{cI}} + \mu)[CI_{\mathrm{T}}],$$

$$\frac{d[Cro_{\mathrm{T}}]}{dt} = \upsilon_{\mathrm{cro}}[M_{\mathrm{cro}}]_{\tau_{\mathrm{cro}}} - (\gamma_{\mathrm{cro}} + \mu)[Cro_{\mathrm{T}}].$$

Santillán, Mackey (2004), *Biophys. J.*, 86(1): 75-84

Analysis of phage λ model

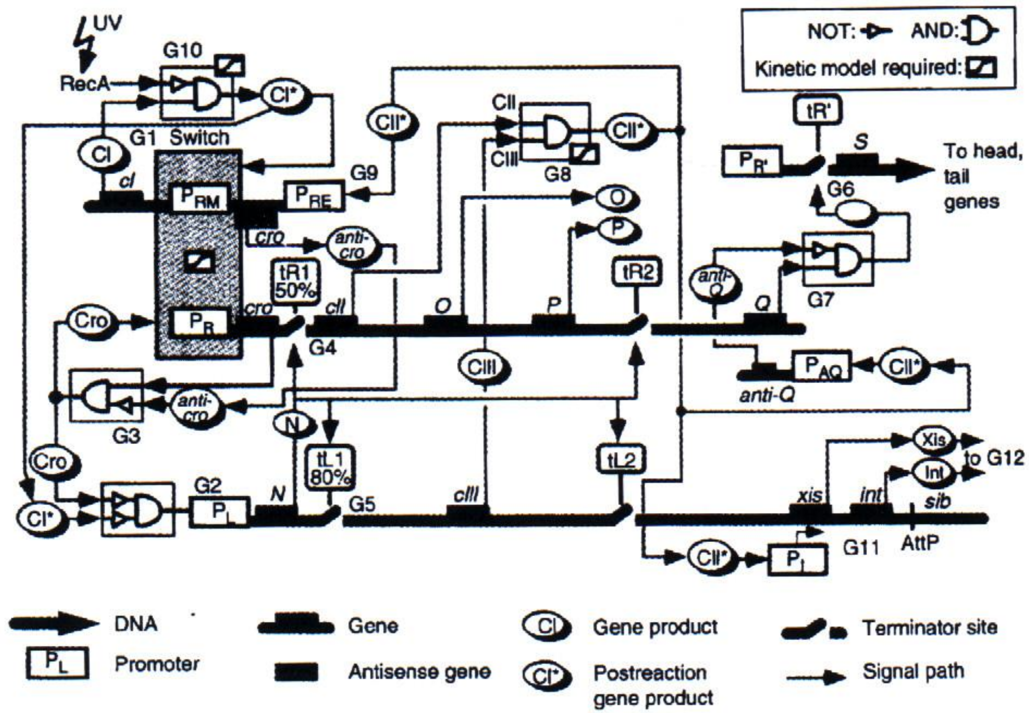
- Bistability (lysis and lysogeny) only occurs for certain parameter values
- Switch from lysogeny to lysis involves bifurcation from one monostable regime to another, due to change in degradation constant



Santillán, Mackey (2004), *Biophys. J.*, 86(1): 75-84

Extended model of phage λ infection

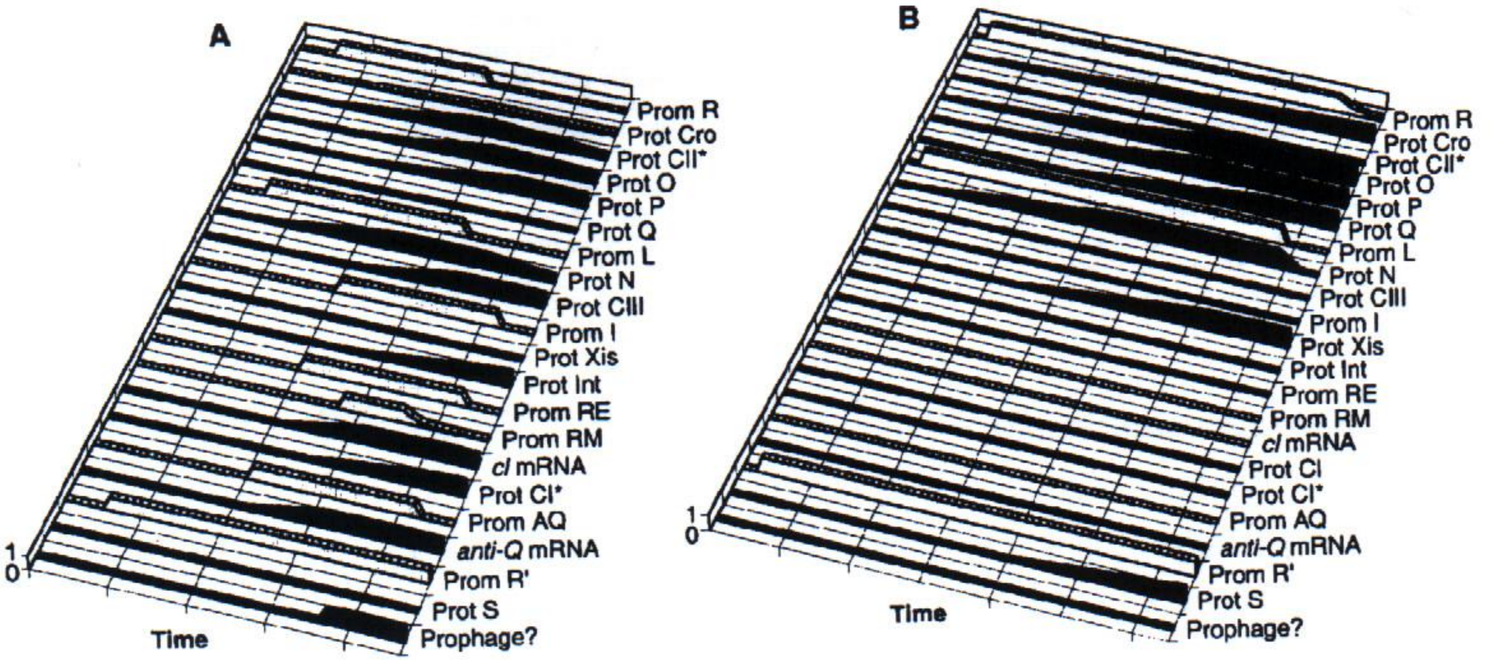
- Differential equation model of the extended network underlying decision between lysis and lysogeny



McAdams, Shapiro (1995), *Science*, 269(5524): 650-656

Simulation of phage λ infection

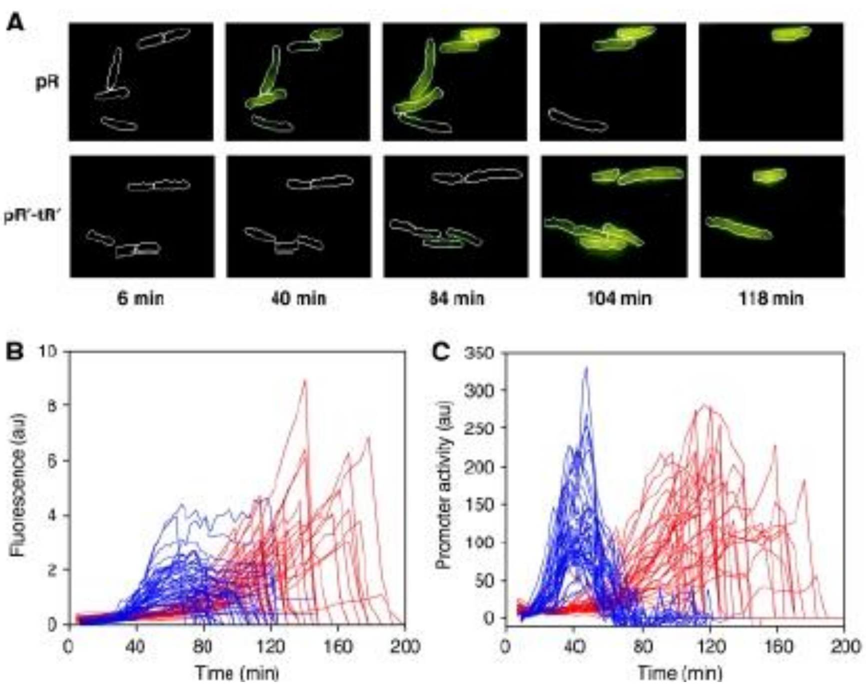
- Numerical simulation of promoter activity and protein concentrations in (a) lysogenic and (b) lytic pathways

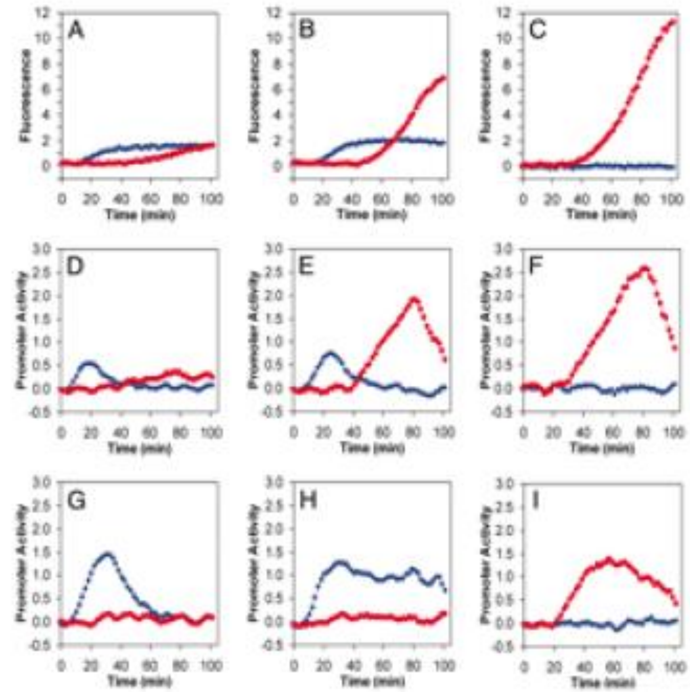


- Cell follows one of two pathways for different initial conditions

Measurements of phage λ infection

- Use of fluorescent reporter genes to follow phage promoters over time, both in populations and in individual cells





Kobiler *et al.* (2005), *Proc. Natl. Acad. Sci. USA*, 102(12): 4470-75

Amir *et al.* (2007), *Mol. Syst. Biol.*, 3:71

Necessary criteria for bistability

- Many other examples of bistability exist in bacteria, such as the *lac* operon
  Dubnau, Losick (2006), *Mol. Microbiol.*, 61 (3):564–72
- Necessary criterion for bistability, or multistability, is the occurrence of **positive feedback** loops in the regulatory network
  Thomas and d'Ari (1990), *Biological Feedback*
- Criterion is not sufficient, as the actual occurrence of bistability depends on parameter values
- **Oscillations** also occur in bacteria, for instance cell cycle or circadian rhythms in photosynthetic bacteria
- Necessary criterion for oscillations is the occurrence of **negative feedback** loops in the regulatory network

Other ODE models

- Circadian clock in mammals

  Leloup and Goldbeter (2003), *Proc. Natl. Acad. Sci. USA*, 100(12):7051-7056

- Cell cycle in yeast

  Chen *et al.* (2004), *Mol. Biol. Cell*, 15(8):3841-3862

- Carbon starvation in bacteria

  Bettenbrock (2005), *J. Biol. Chem.*, 281(5):2578-2584

- Signal transduction cascades and developmental decisions

  Ferrell and Machleder (1998), *Science*, 280(5365):852-853

- Pattern formation in fruit fly embryon

  Jaeger *et al.* (2004), *Nature*, 430(6997):368-371

Evaluation of differential equations

- **Pro**: general formalism for which powerful analysis and simulation techniques exist
- **Pro**: well-developed theoretical framework for application to genetic regulatory networks
- **Contra**: numerical techniques are often not appropriate due to lack of quantitative data on model parameters
- **Contra**: assumptions of continuous and deterministic change of concentrations may not be valid on molecular level

Lack of quantitative information: strategies

- Three main strategies to deal with lack of quantitative data:
  - **Parameter sensitivity and robustness**
  - Parameter estimation from time-series data
  - Model reduction

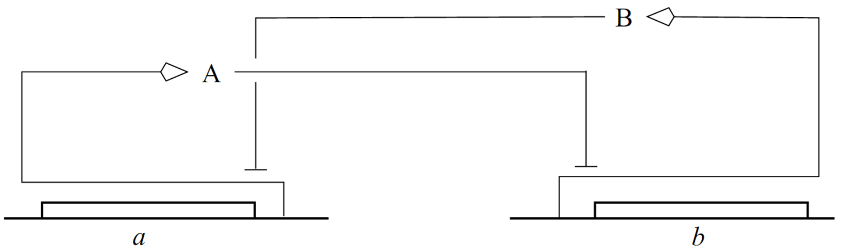
De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

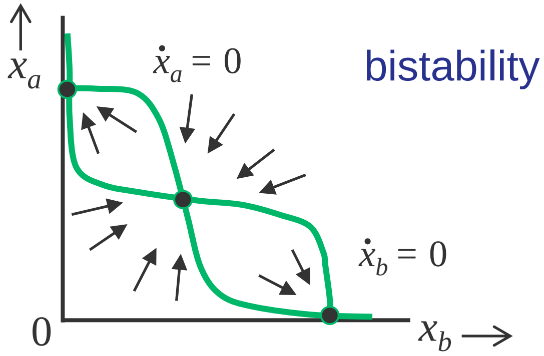
Lack of quantitative data: robustness

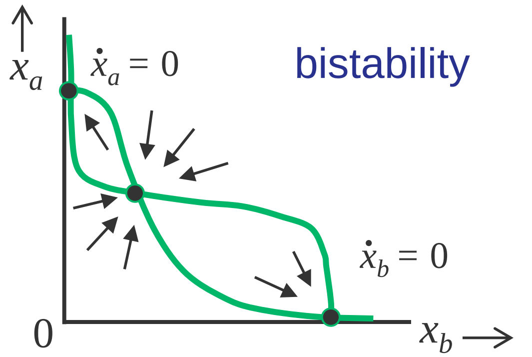
- Important dynamic properties are expected to be **robust** over large ranges of parameter values

    Important dynamic properties should be insensitive to moderate variations in parameter values

Stelling *et al.* (2004), *Cell*, 118(6):675-685

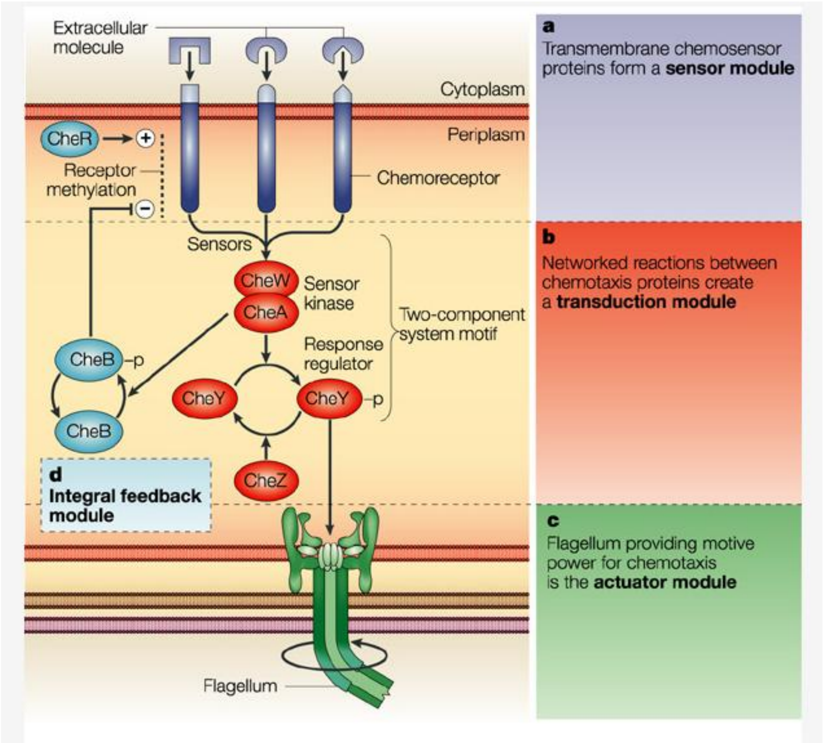


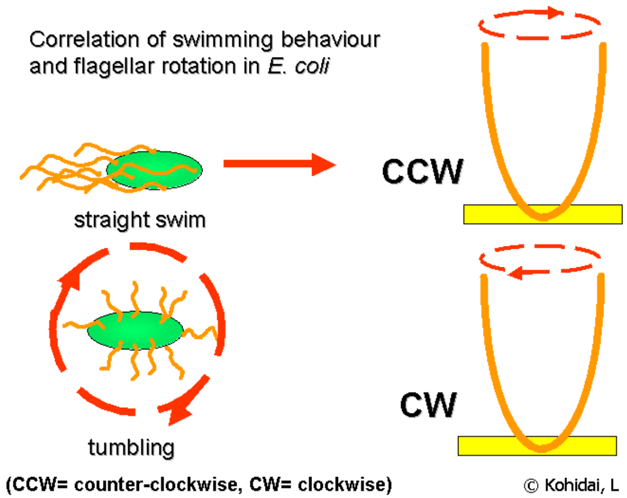




Robustness in *E. coli* chemotaxis

- **Chemotaxis** in bacteria is ability to sense gradient of chemical ligands in environment
  - Adjustment of tumbling frequency of molecular motor

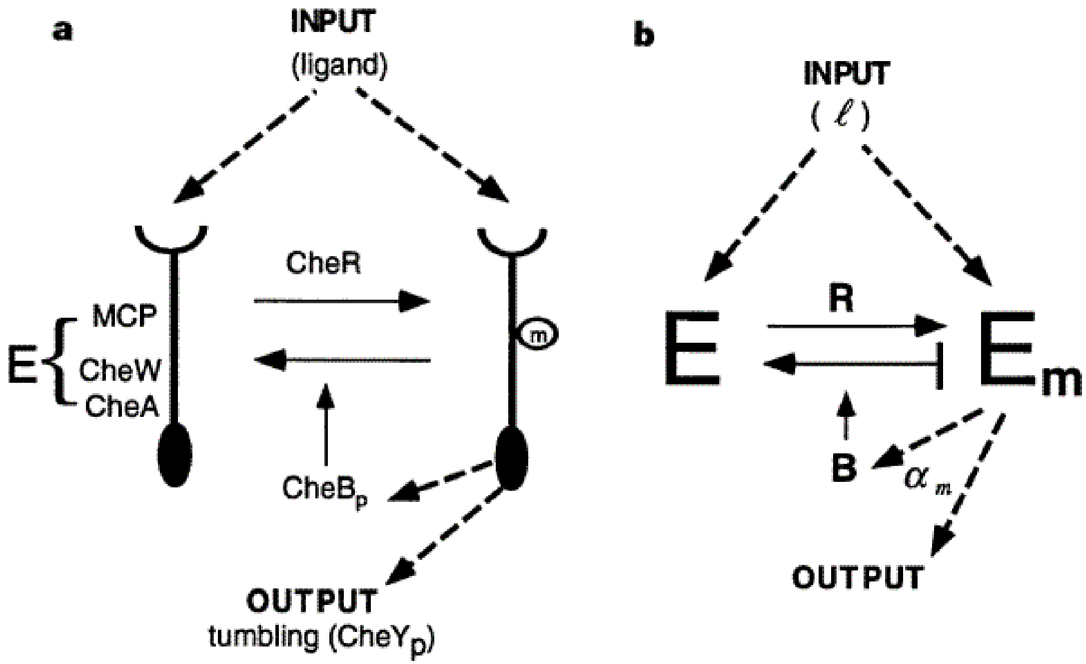




McAdams *et al.* (2004), *Nat. Rev. Genet.*, 5:169-178

Robustness in *E. coli* chemotaxis

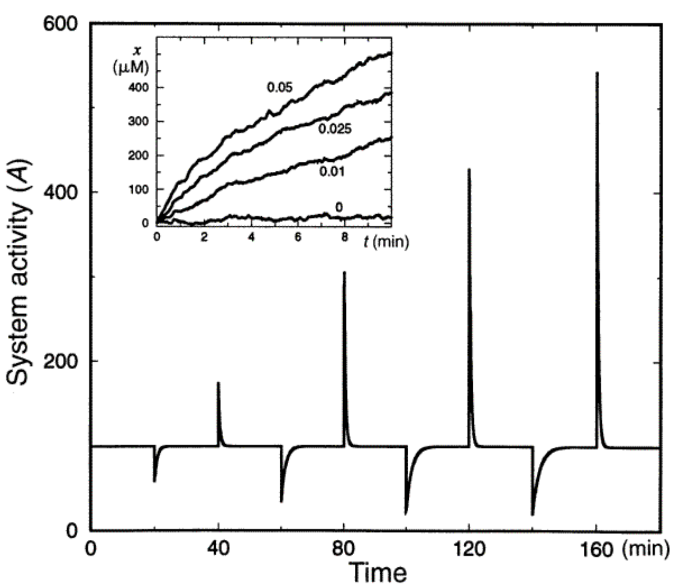
- Differential equation model of signal transduction network underlying bacterial chemotaxis



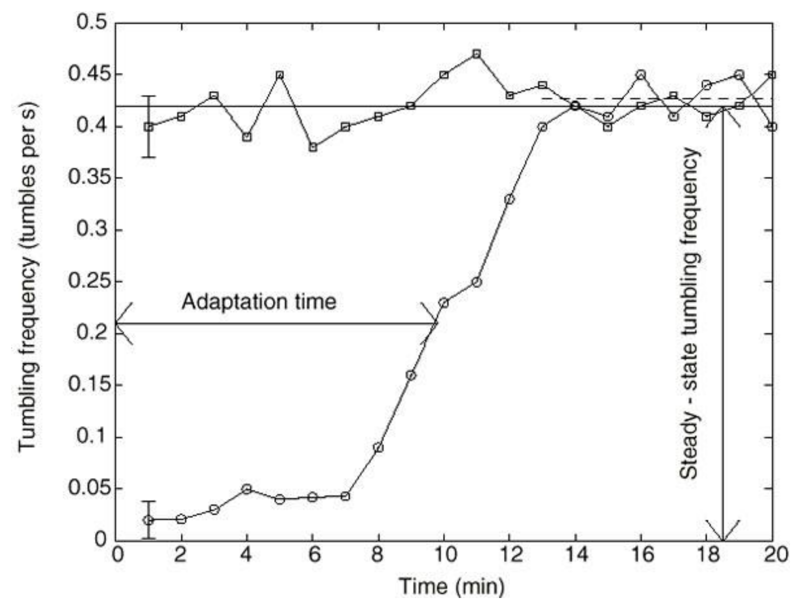
Barkai and Leibler (1997), *Nature*, 387(6636):913-917

Robustness in *E. coli* chemotaxis

- **Adaptation** property is insensitivity of steady-state tumbling frequency to ligand concentration
- Robustness of adaptation property over wide range of parameter values (model and experiments)



Barkai and Leibler (1997), *Nature*, 387:913-917



Alon *et al.* (1999), *Nature*, 397:168-171

Lack of quantitative information: strategies

- Three main strategies to deal with lack of quantitative data:
  - Parameter sensitivity and robustness
  - **Parameter estimation from time-series data**
  - Model reduction

De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

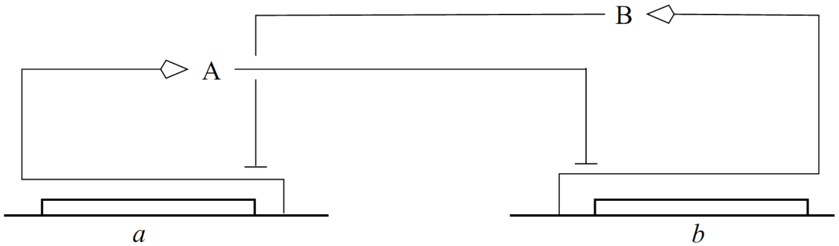
Lack of quantitative data: estimation

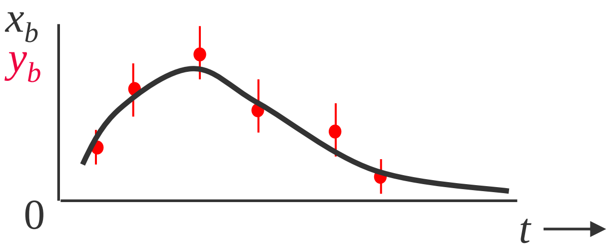
- **Estimate** parameter values from experimental time-series data

    Systems identification in control and engineering

    Ljung (1999), *System Identification: Theory for the User*

- Given model structure, search parameter values for which model predictions best fit experimental data





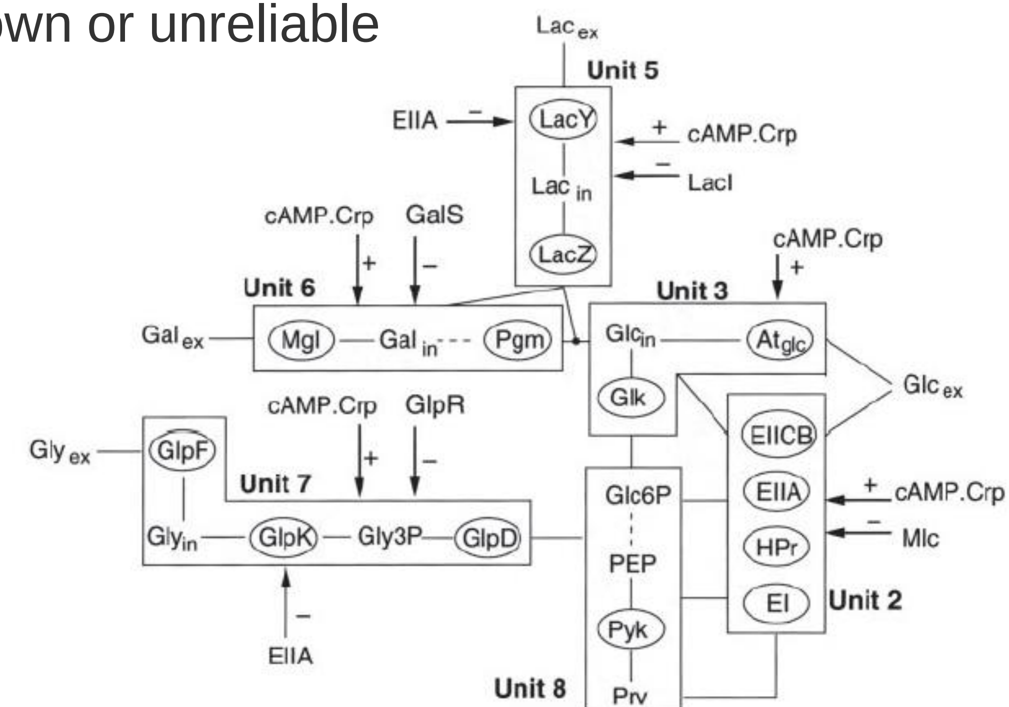
- Minimization of objective function, for instance sum of squared errors:

$$\sum_t (x(t,\theta) - y(t))^2$$

    Possibility to add constraint or penalty terms to restrict parameter space

Estimation of parameter values

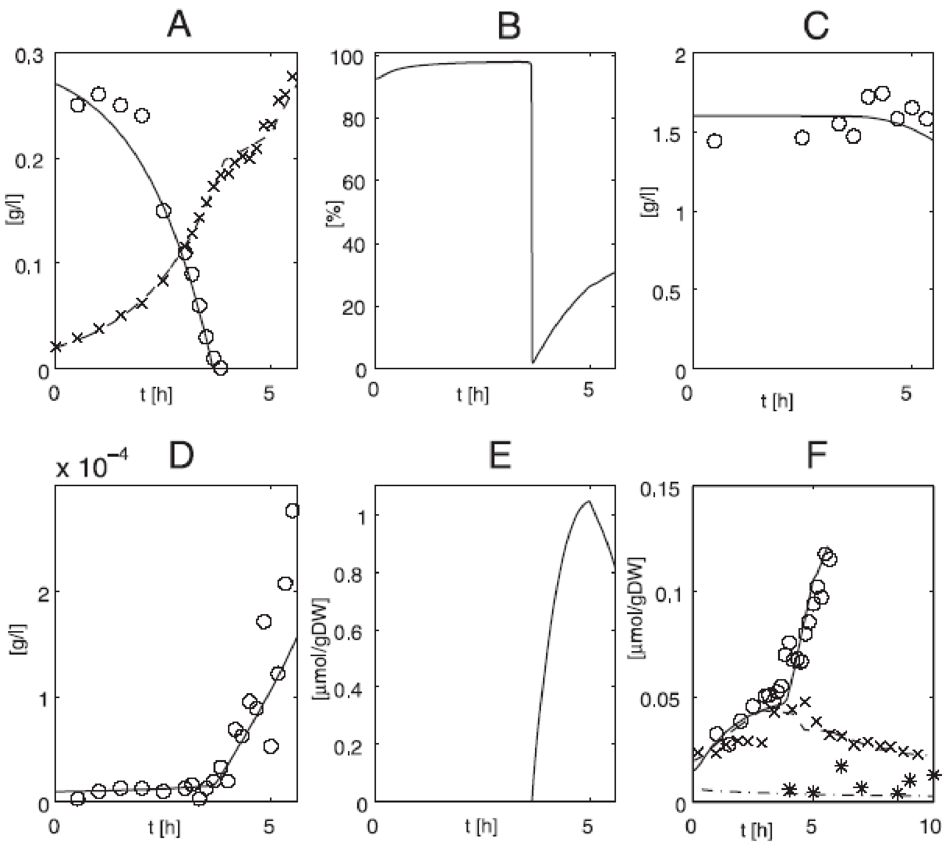
- Nonlinear differential equation model of uptake of carbon sources (glucose, lactose, glycerol, …) by *E. coli*
  - Several dozens of equations and more than a hundred parameters, many of them unknown or unreliable



Bettenbrock *et al.* (2005), *J. Biol. Chem.*, 281(5): 2578-2584

Estimation of parameter values

- Estimation of parameter values from time-series measurements of metabolite concentrations on wild-type and mutant strains



Bettenbrock *et al.* (2005), *J. Biol. Chem.*, 281(5): 2578-2584

Limitations of system identification

- No algorithms that guarantee globally optimal solution for parameter estimation in nonlinear models

  Evolutionary algorithms, simulated annealing, genetic algorithms, …

- Model identifiability demands experimental data of sufficient quantity and quality

  Common problems: noise, sampling density, unobserved variables, …

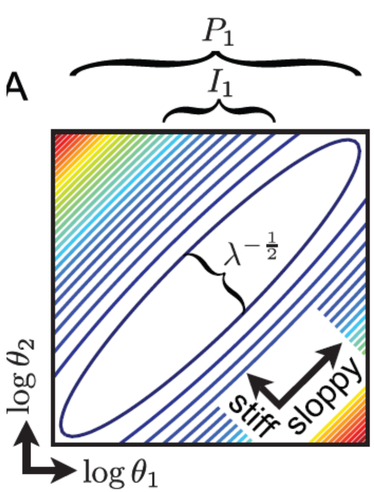
  Van Riel (2006), *Brief. Bioinform.*, 7(4):364-374

- However, models of cellular regulatory networks may be non-identifiable by principle, and …

  … even partially identifiable models may yield interesting results

Sloppy parameter sensitivities

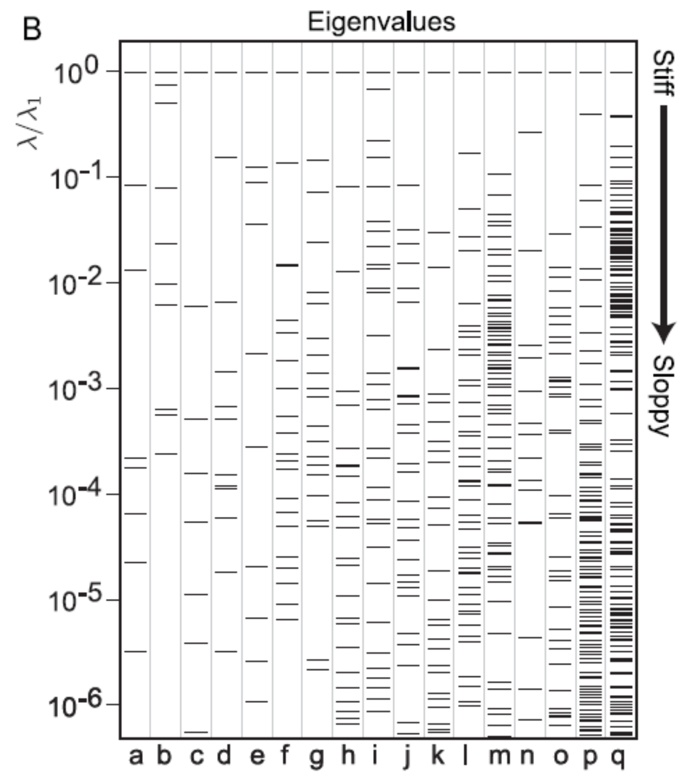
- Sensitivity of model predictions to variation of individual parameters may be limited, though certain combinations of parameters may be tightly constrained
  - Diagrams showing ellipsoids of constant model behavior (error)
  - Skewedness of ellipsoid measured by eigenvalues λ of Hessian matrix accounting for sensitivity of model behavior to changes in parameters



Gutenkunst *et al.* (2007), *PLoS Comput. Biol.*, 3(10): e189

Sloppy parameter sensitivities

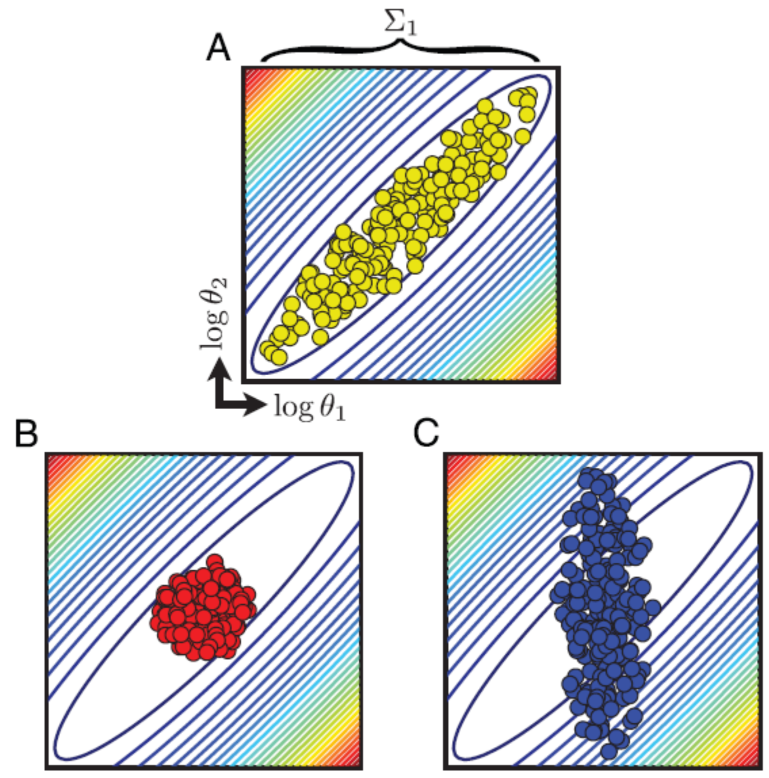
- Sensitivity of model predictions to variation of individual parameters may be limited, though certain combinations of parameters may be tightly constrained
  - Most models have skewed ellipsoids, as indicated by relative eigenvalues far from 1
  - Moreover, ratios of eigenvalues spread over several orders of magnitude: **sloppy** parameter sensitivities



Gutenkunst *et al.* (2007), *PLoS Comput. Biol.*, 3(10): e189

Sloppy parameter sensitivities

- Consequence: uncertainty in individual parameters estimated from data may be large, but model predictions nevertheless tightly constrained
- Also: direct measurements of parameters may need to be extremely precise to obtain good predictions



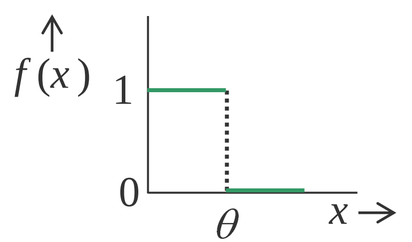
Gutenkunst *et al.* (2007), *PLoS Comput. Biol.*, 3(10): e189

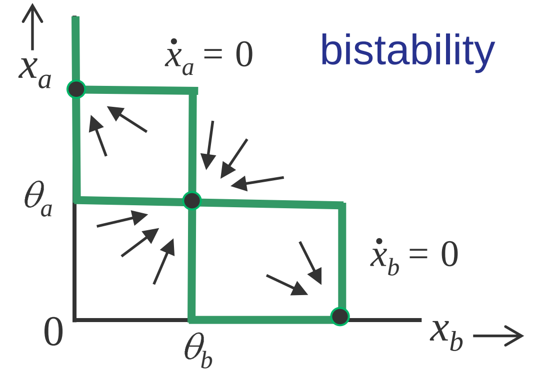
Lack of quantitative data: reduction

- Use **model reduction** to obtain simpler models that can be analyzed with less information on parameter values

  Piecewise-linear instead of nonlinear models

$$\dot{x}_a = \kappa_a f(x_b) - \gamma_a x_a$$
$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$





Glass and Kauffman (1973), *J. Theor. Biol.*, 39(1):103-29

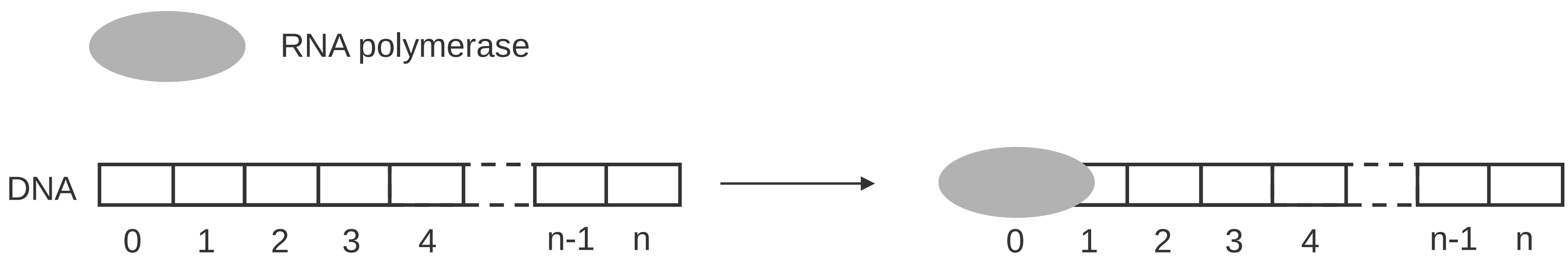
de Jong *et al.* (2004), *Bull. Math. Biol.*, 66(2):301-340

- Other example of model reduction: quasi-steady state assumption

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*

Gene expression is discrete process

- Gene expression is result of large number of **discrete** events: chemical reactions involved in protein synthesis (and degradation)

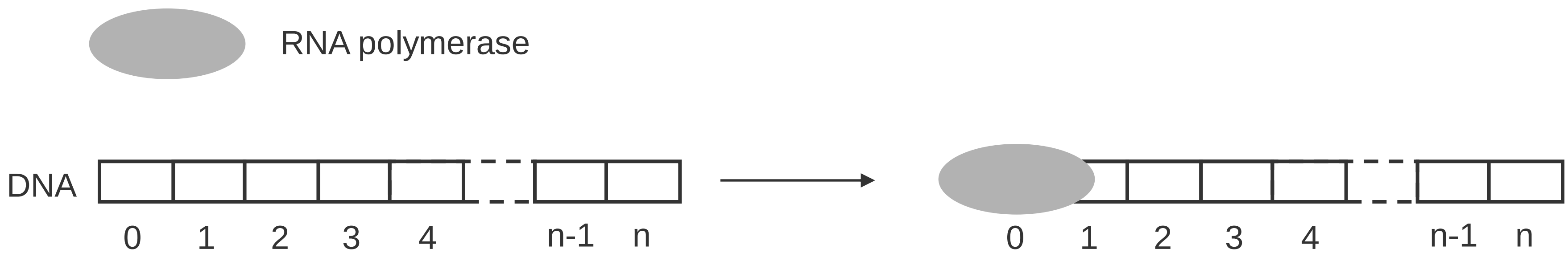


$$\mathrm{DNA} + \mathrm{RNAP} \rightarrow \mathrm{DNA}_0 \bullet \mathrm{RNAP}$$

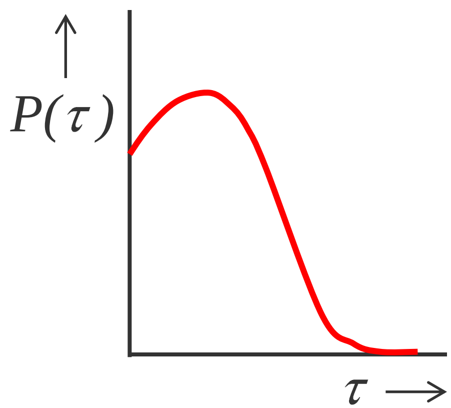
$$\mathrm{DNA}_i \bullet \mathrm{RNAP} \rightarrow \mathrm{DNA}_{i+1} \bullet \mathrm{RNAP}$$

Gene expression is stochastic process

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

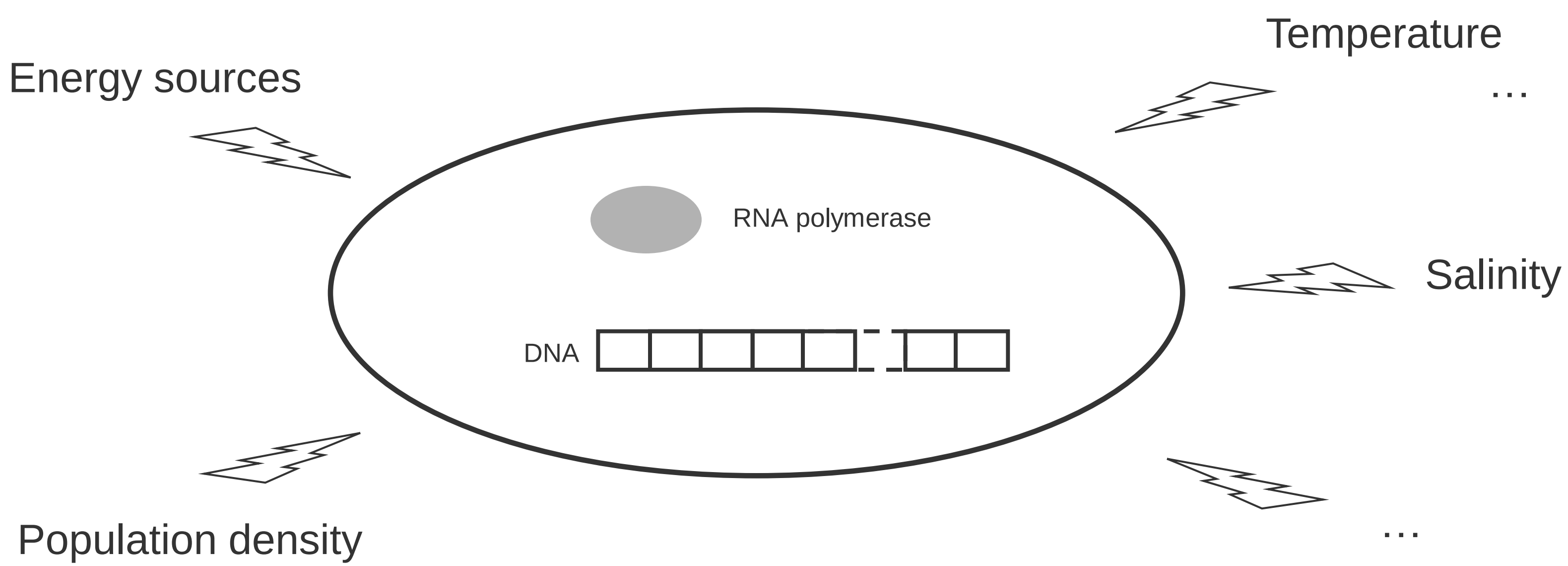


- Time interval τ has probability distribution



Gene expression is stochastic process

- Gene expression is **stochastic** process: reactions in cell occur in presence of external fluctuations



Differential equations are abstractions

- Differential equation models make **continuous** and **deterministic** abstraction of discrete and stochastic process
  - $x_i(t) \in R_{\geq 0}$ is continuous variable
  - $\dot{x}_i = f_i(\boldsymbol{x})$ means deterministic change of x_i at t
- Abstraction may not be warranted when modeling gene regulation on molecular level

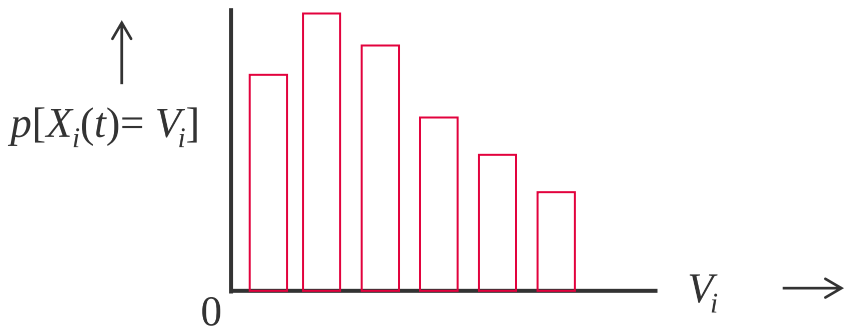
  Stochasticity gives rise to (internal and external) **noise**
- Noise effects strengthened by low number of molecules of each species

Rao *et al.* (2002), *Nature*, 420(6912): 231-237

Kaern *et al.* (2005), *Nat. Rev. Genet.*, 6(6):451-464

Stochastic models

- **Stochastic models** of gene regulation are more realistic
- Number of molecules of each species i at time-point t represented by discrete variable $X_i(t) \in$ N
- Reactions between molecular species lead to change in state of system from $\boldsymbol{X}(t)$ to $\boldsymbol{X}(t+\Delta t)$ over time-interval Δt, where $\boldsymbol{X} = [X_1,\ldots, X_n]'$
- Probability distribution $p[X_i(t)=V_i]$ describes probability that at time-point t there are V_i molecules of species i



Rao *et al.* (2002), *Nature*, 420(6912): 231-237

Stochastic master equation

❖ Equation describes evolution of state $\boldsymbol{X}$ of regulatory system

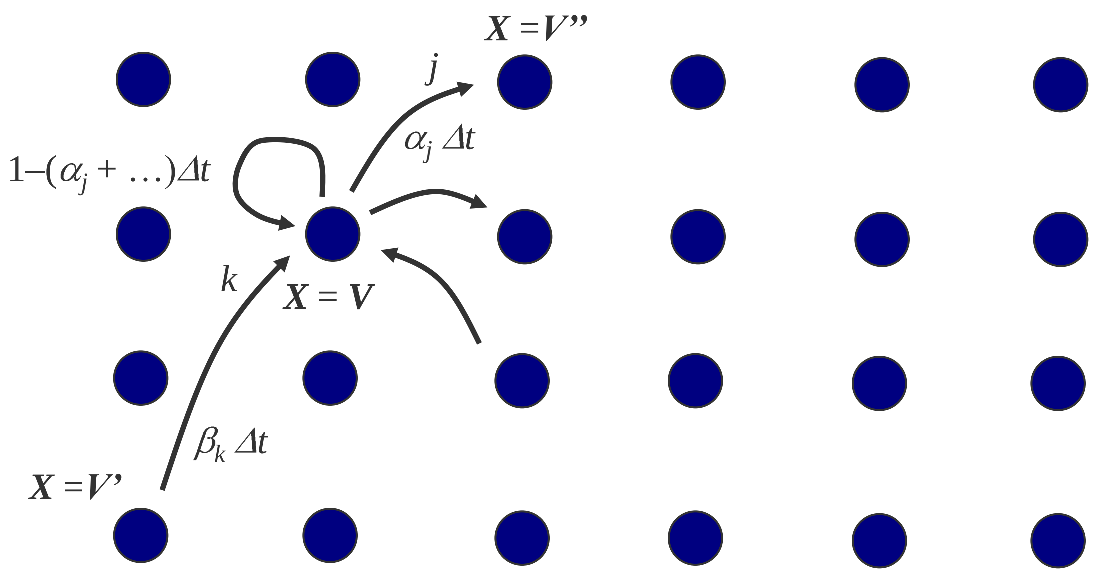
$$p[\boldsymbol{X}(t+\Delta t)=\boldsymbol{V}] = p[\boldsymbol{X}(t)=\boldsymbol{V}]\,(1-\sum_{j=1}^{m}\alpha_j\,\Delta t) + \sum_{\boldsymbol{V'}\in N^n,\,\boldsymbol{V'}\neq\boldsymbol{V}} p[\boldsymbol{X}(t)=\boldsymbol{V'}]\sum_{k=1}^{m}\beta_k\,\Delta t$$

- 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 $\boldsymbol{X}(t)=\boldsymbol{V}$
- $\beta_k\,\Delta t$ is the probability that reaction k will bring the system from $\boldsymbol{X}(t)=\boldsymbol{V'}$ to $\boldsymbol{X}(t+\Delta t)=\boldsymbol{V}$ in $[t, t+\Delta t]$

Van Kampen (1997), *Stochastic Processes in Physics and Chemistry*

Stochastic view on dynamics

- Reactions between molecular species lead to state change
  - $\alpha_j \, \Delta t$ is the probability that reaction j will occur in interval of length Δt given that $\boldsymbol{X}=\boldsymbol{V}$
  - $\beta_k \, \Delta t$ is the probability that reaction k will bring the system from $\boldsymbol{X}=\boldsymbol{V'}$ to $\boldsymbol{X}=\boldsymbol{V}$ in interval of length Δt

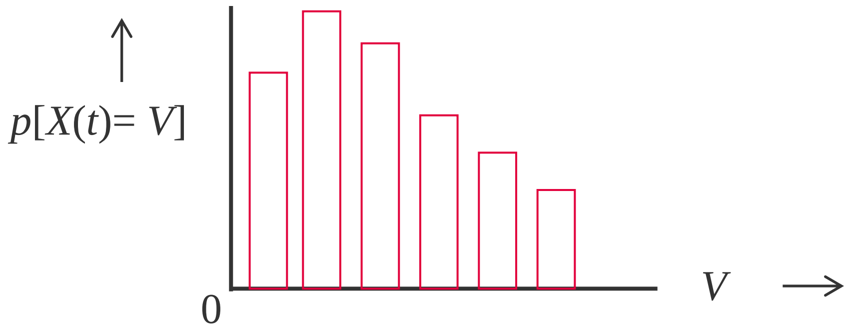


Stochastic master equation

- For $\Delta t \to 0$ we obtain **stochastic master equation**

$$\partial p[\boldsymbol{X}(t)=\boldsymbol{V}] \,/\, \partial t = \sum_{j=1}^{m} p[\boldsymbol{X}(t)=\boldsymbol{V}-\boldsymbol{\nu}_{\mathrm{j}}]\, \beta_j - p[\boldsymbol{X}(t)=\boldsymbol{V}]\, \alpha_j$$

Van Kampen (1997), *Stochastic Processes in Physics and Chemistry*

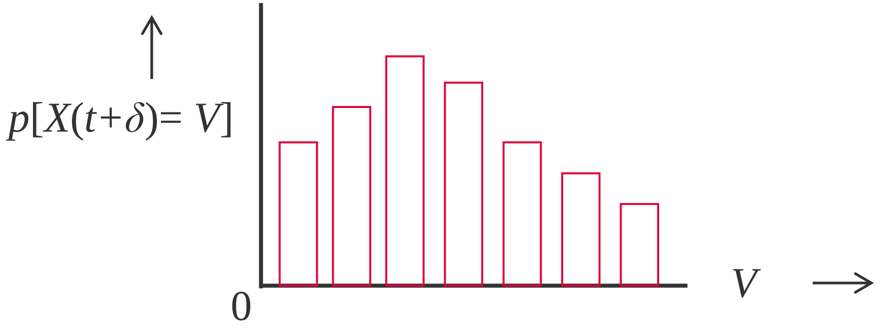


Stochastic master equation

- For $\Delta t \rightarrow 0$ we obtain **stochastic master equation**

$$\partial p[\boldsymbol{X}(t)=\boldsymbol{V}] \,/\, \partial t = \sum_{j=1}^{m} p[\boldsymbol{X}(t)=\boldsymbol{V}-\boldsymbol{\nu}_{\mathrm{j}}]\ \beta_j - p[\boldsymbol{X}(t)=\boldsymbol{V}]\ \alpha_j$$

Van Kampen (1997), *Stochastic Processes in Physics and Chemistry*

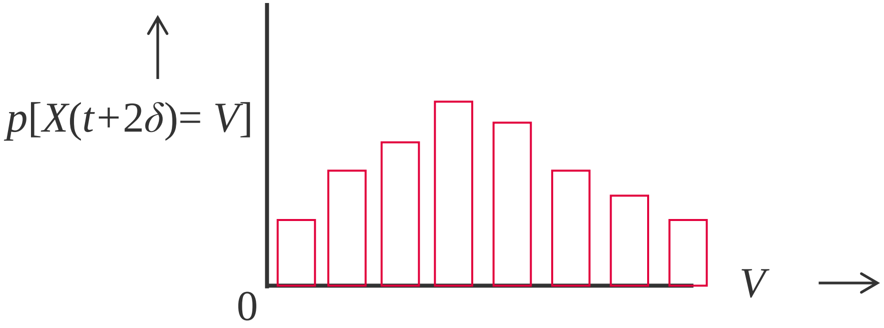


Stochastic master equation

❖ For $\Delta t \to 0$ we obtain **stochastic master equation**

$$\partial p[\boldsymbol{X}(t)=\boldsymbol{V}] \,/\, \partial t = \sum_{j=1}^{m} p[\boldsymbol{X}(t)=\boldsymbol{V}-\boldsymbol{\nu}_{\mathrm{j}}]\ \beta_j - p[\boldsymbol{X}(t)=\boldsymbol{V}]\ \alpha_j$$

Van Kampen (1997), *Stochastic Processes in Physics and Chemistry*

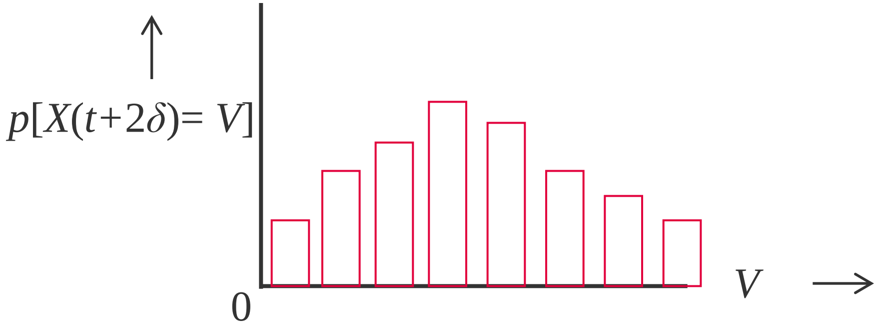


Stochastic master equation

- For $\Delta t \rightarrow 0$ we obtain **stochastic master equation**

$$\partial p[\boldsymbol{X}(t)=\boldsymbol{V}] \,/\, \partial t = \sum_{j=1}^{m} p[\boldsymbol{X}(t)=\boldsymbol{V}-\boldsymbol{\nu}_{\mathrm{j}}]\ \beta_j - p[\boldsymbol{X}(t)=\boldsymbol{V}]\ \alpha_j$$

Van Kampen (1997), *Stochastic Processes in Physics and Chemistry*



- Probabilities α_j, β_j are defined in terms of kinetic constants of reactions

- Analytical solution of master equations is not possible in general

Stochastic simulation

- **Stochastic simulation** predicts sequences of reactions that change state of system, starting from initial state $\boldsymbol{X}(0) = \boldsymbol{V}_{\mathbf{0}}$

  Stochastic simulation samples joint probability density function

  $p[\tau, j|\boldsymbol{X}(t) = \boldsymbol{V}]$

  $\tau =$ time interval until occurrence of next reaction

  $j =$ index of next reaction

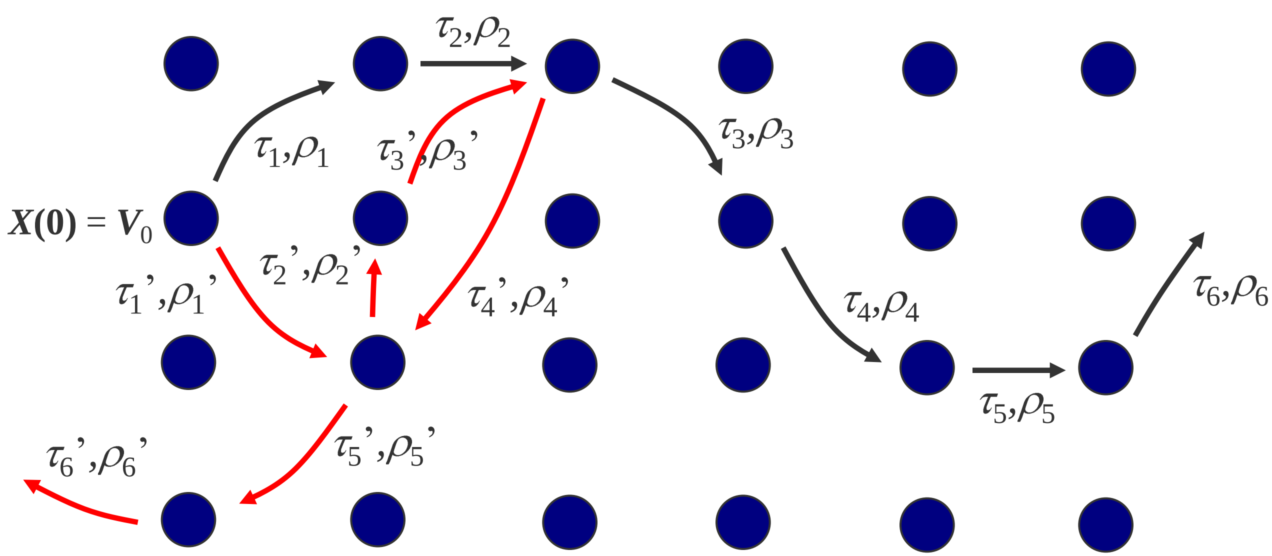
  Probability density function defined in terms of α_j, β_k (reaction constants)

- Repeating stochastic simulations yields approximation of $p(\boldsymbol{X}(t)=\boldsymbol{V})$, and thus solution of stochastic master equation

Gillespie (2002), *J. Phys. Chem.*, 81(25): 2340-2361

Stochastic simulation

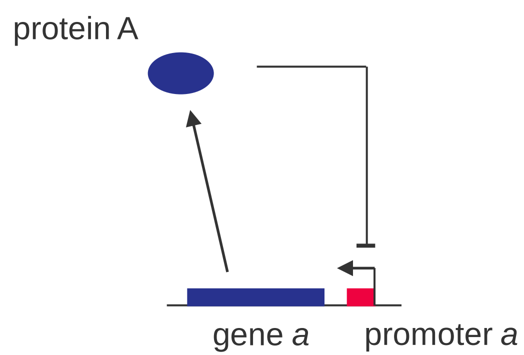
- Stochastic simulation generates sequences of reactions and time intervals between reactions, starting from initial state $\boldsymbol{X}(0)$



- Stochastic simulation may lead to different dynamical behaviors starting from identical initial conditions

Auto-inhibition network

- **Auto-inhibition** network consists of a single gene, coding for transcription regulator inhibiting expression of its own gene

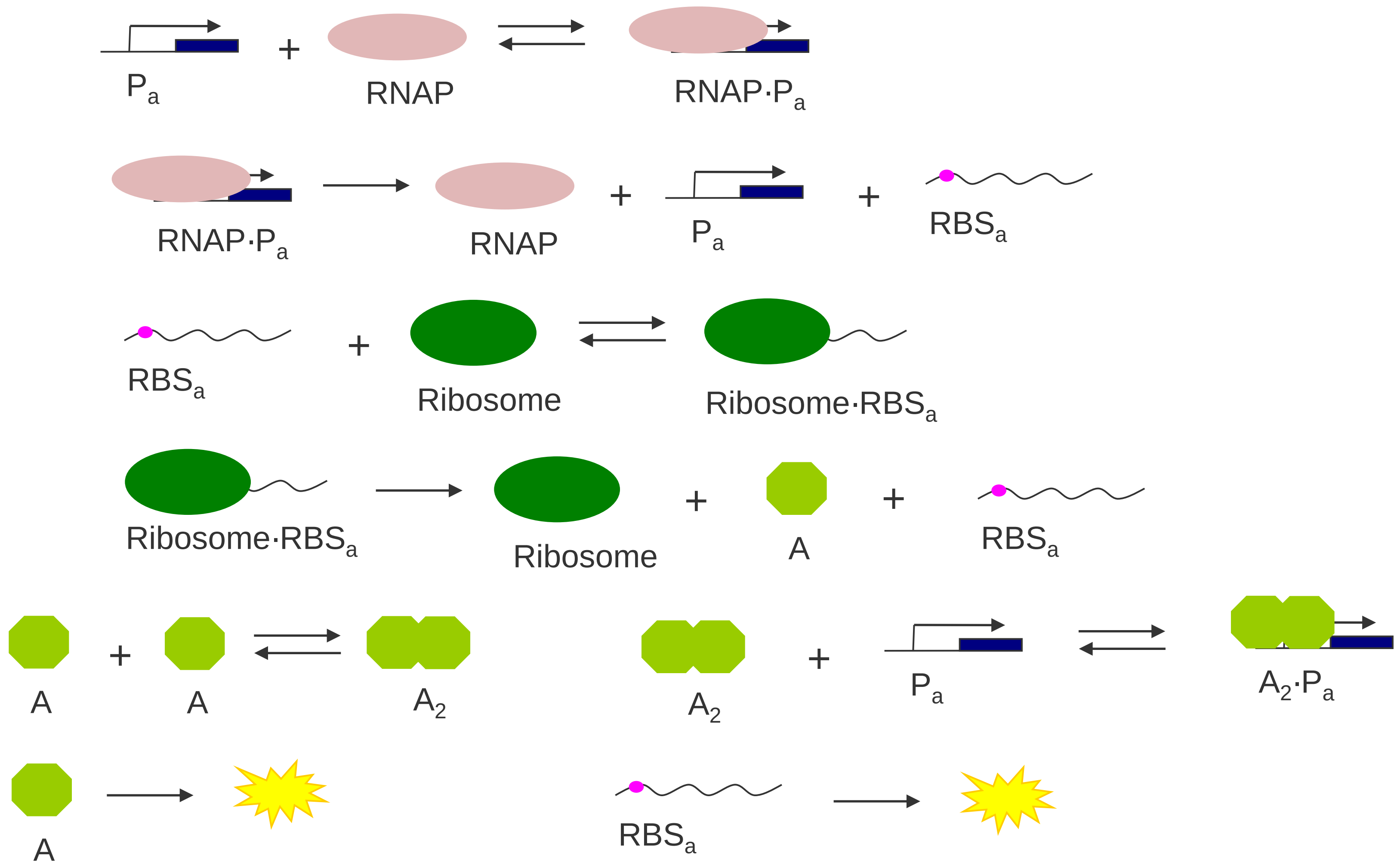


- Auto-inhibition is example of **negative feedback**, and frequently occurs in bacterial regulatory networks

Thieffry *et al.* (1998), *BioEssays*, 20(5):433-440

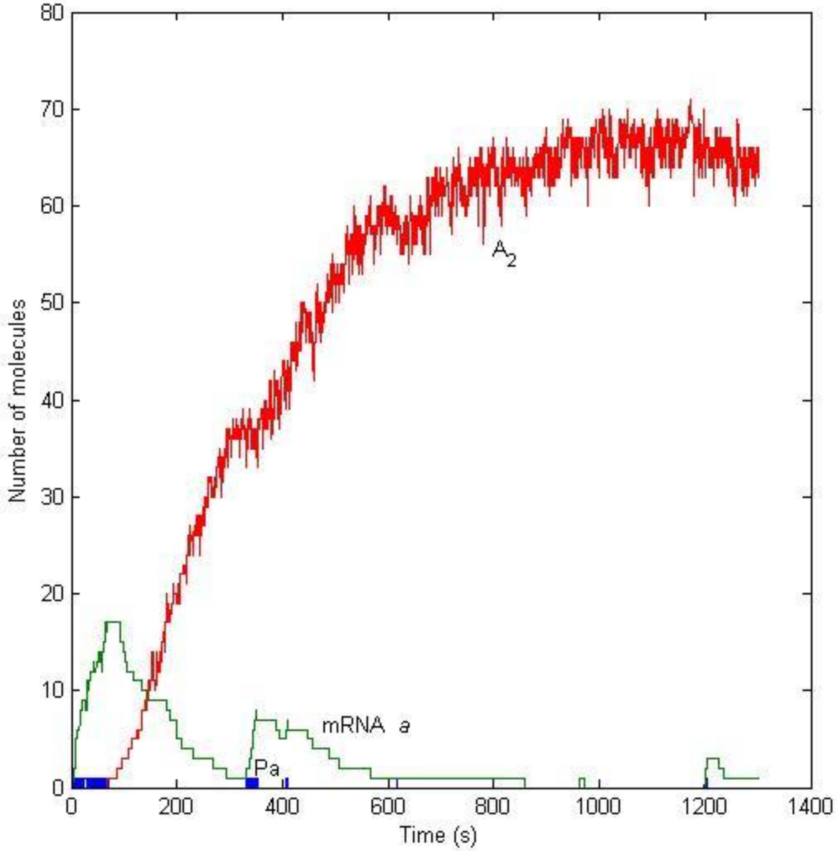
- Development of stochastic model requires list of species, reactions, and kinetic constants

Reactions and species



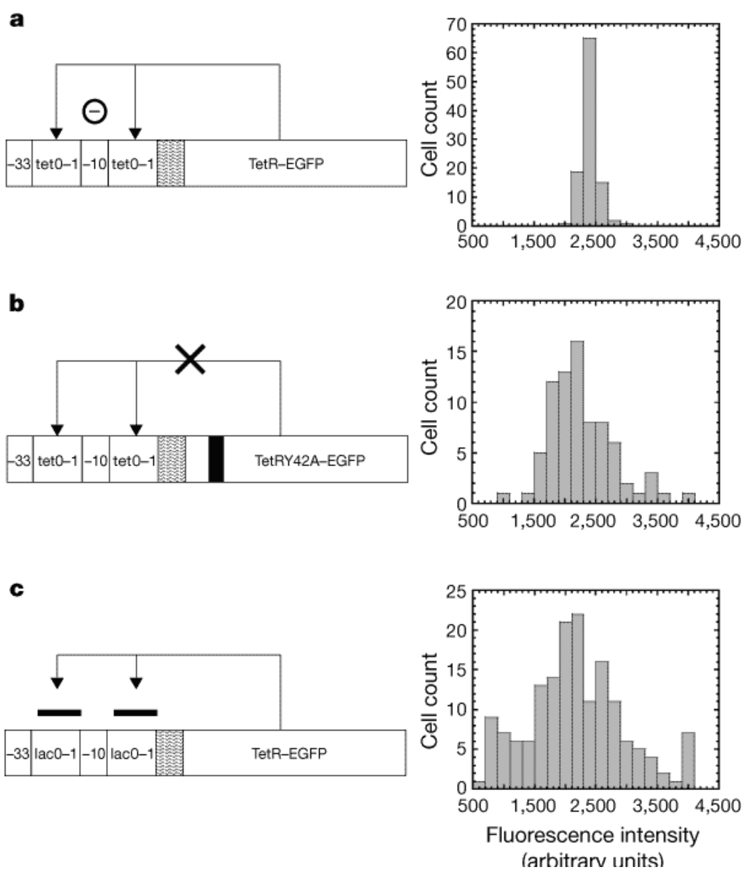
Stochastic simulation of auto-inhibition

❖ Occurrence of fluctuations and bursts in gene expression



Auto-inhibition and noise reduction

- Auto-inhibition reduces fluctuations in gene expression level

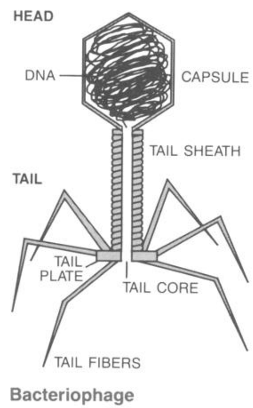


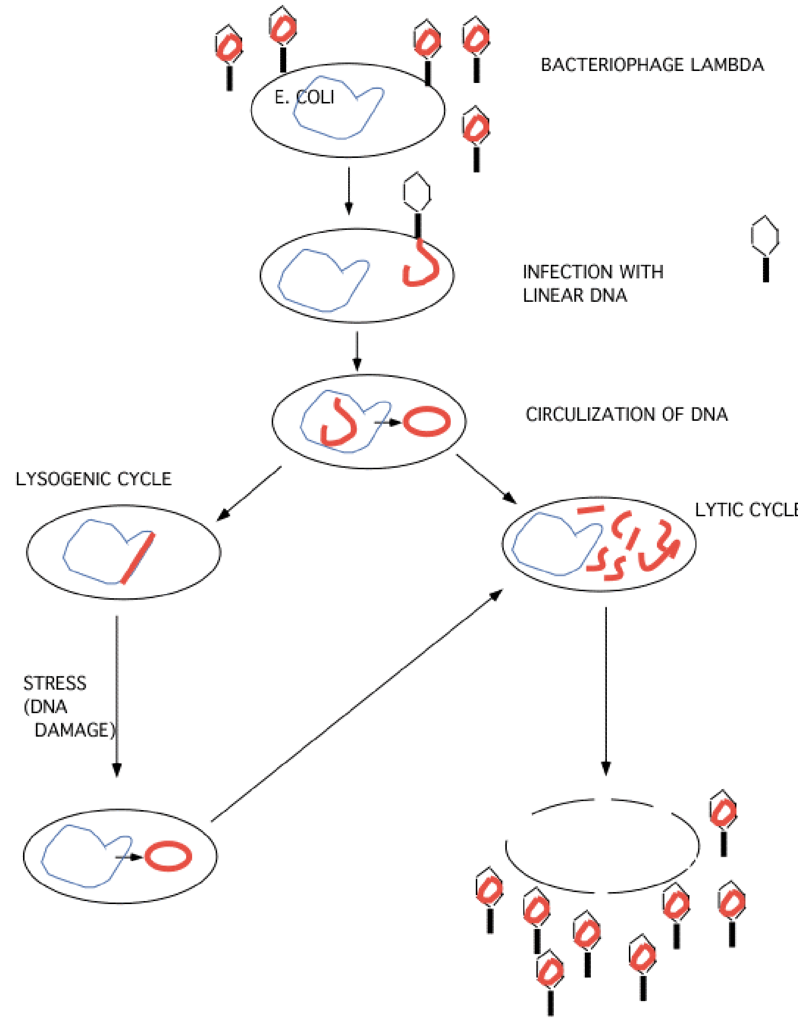
Becskei and Serrano (2000), *Nature*, 405(6785):590-591

Bacteriophage λ infection of *E. coli*

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

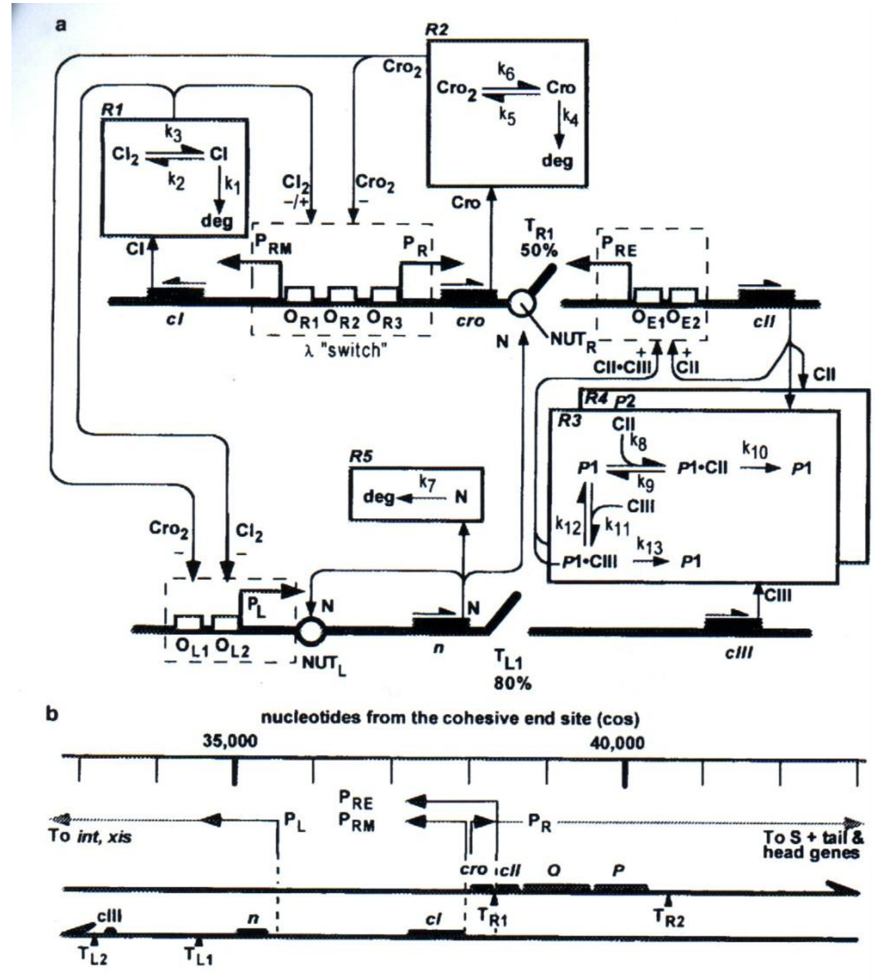
Ptashne (1997), *A Genetic Switch: Phage λ and Higher Organisms*





Stochastic analysis of phage λ infection

- Stochastic model of λ lysis-lysogeny decision network

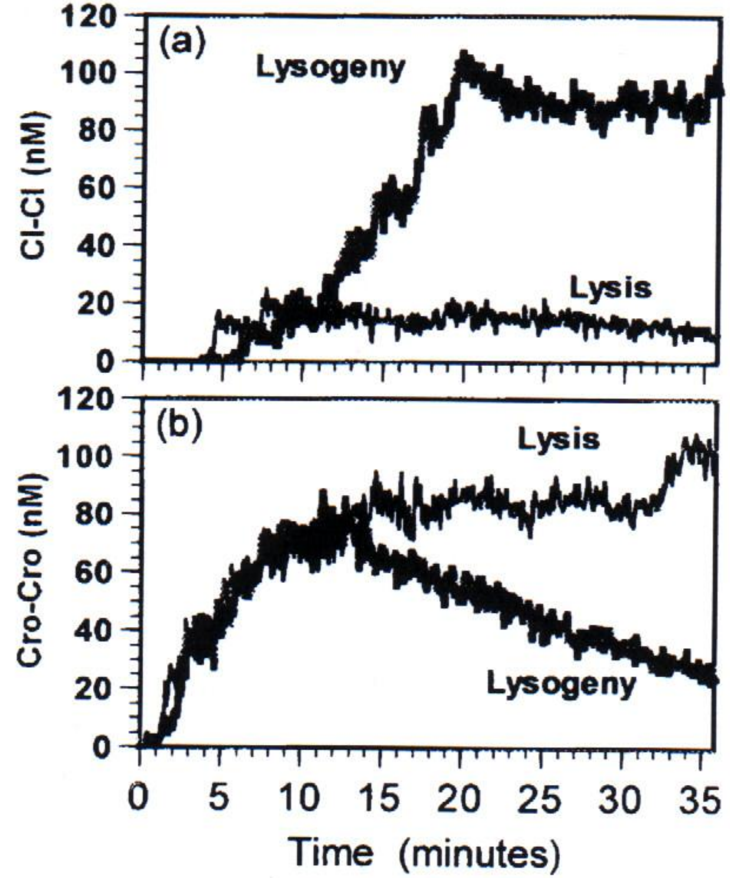


Arkin *et al.* (1998), *Genetics*, 149(4): 1633-1648

Stochastic analysis of phage λ 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), *Genetics*, 149(4): 1633-1648

Comparison with deterministic approach

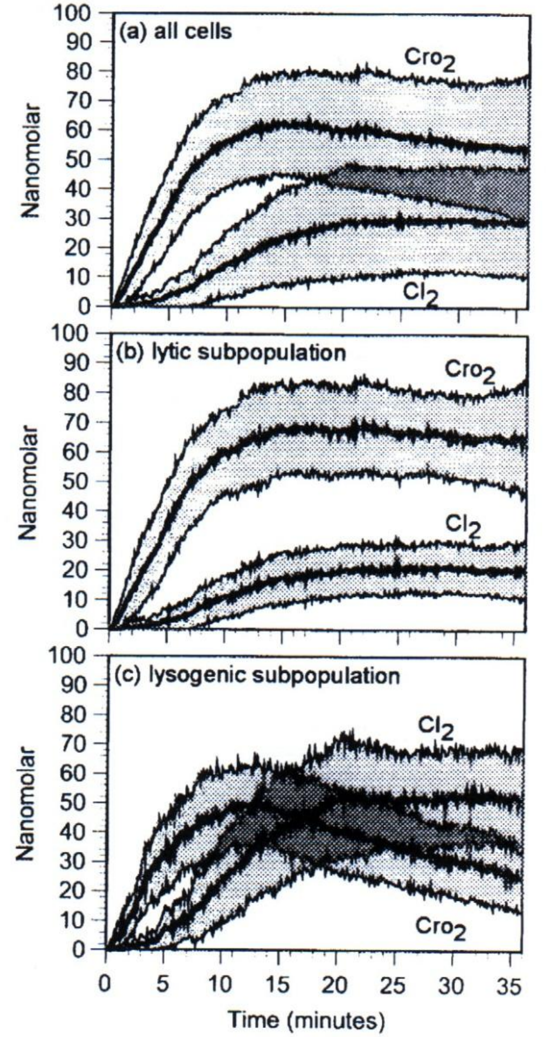
- Deterministic models can be seen as predicting **average behavior** of cell population

  Gillespie. (2000), *J. Chem. Phys.*, 113(1): 297-306

- Analysis of average behavior may obscure that one part of population chooses one pathway rather than another

  Arkin *et al.* (1998), *Genetics*, 149(4): 1633-1648

- However, under some conditions deterministic models yield good approximation



Other stochastic models

- Effect of noise on carbon assimilation in *E. coli*

  Puchalka and Kierzek (2004), *Biophys. J.*, 86(3):1357-1372

- Regulation of expression of virulence factor in pathogenic *E. coli*

  Jarboe *et al.* (2004), *Biotechnol. Bioengin.*, 88(2):189-203

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 kinetic constants, necessary for generating values of parameters τ and ρ, are not or incompletely known

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

  Large networks cannot currently be handled, but a host of extensions and approximations have been developed

Conclusions

- Mathematical methods and **computer tools** for modeling and simulation necessary to understand genetic regulatory processes

- Variety of approaches available, representing genetic regulatory systems on different **levels of abstraction**

- Choice of approach depends on **biological problem** and on **available information**:
  - knowledge on reaction mechanisms
  - quantitative data on model parameters and gene expression levels

- Lots of **applications** on bacteria and higher organisms

Challenges

- Integration of models and experimental data

  New techniques for obtaining real-time measurements in living cells, on level of populations and single cells

- Upscaling to large networks of dozens or even hundreds of genes, proteins, metabolites, …
  - Formal verification tools
  - Model reduction

- Perturbation and redesign of regulatory networks

  Synthetic biology

- From model systems to organisms of medical and biotechnical interest