Modeling Genetic and Metabolic Networks and their Evolution

Peter Schuster

Institut für Theoretische Chemie der Universität Wien, Austria

40. Winterseminar

Klosters, 28.01.2005

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks

1. What is computational systems biology?
2. Networks and network evolution
3. Forward and inverse problems
4. Reverse engineering – A simple example
5. MiniCellSim – A simulation tool
6. Evolution of genetic and metabolic networks

1. **What is computational systems biology?**
2. Networks and network evolution
3. Forward and inverse problems
4. Reverse engineering – A simple example
5. MiniCellSim – A simulation tool
6. Evolution of genetic and metabolic networks

Structural biology

Sequence $\Rightarrow$ Structure $\Rightarrow$ Function

Computational systems biology

Genome $\Rightarrow$ Proteome $\Rightarrow$ Dynamics of cells and organisms

Structural biology

Sequence $\Rightarrow$ Structure $\Rightarrow$ Function

Computational systems biology

Genome $\Rightarrow$ Proteome $\Rightarrow$ Dynamics of cells and organisms

Goals:

1. Large scale simulation of genetic regulatory and metabolic reaction networks.

Structural biology

Sequence $\Rightarrow$ Structure $\Rightarrow$ Function

Computational systems biology

Genome $\Rightarrow$ Proteome $\Rightarrow$ Dynamics of cells and organisms

Goals:

1. Large scale simulation of genetic regulatory and metabolic reaction networks.

2. Understanding the dynamics of cells and organisms. Are kinetic differential equations applicable to spatially highly heterogeneous media with active transport?

Structural biology

Sequence $\Rightarrow$ Structure $\Rightarrow$ Function

Systems biology

Genome $\Rightarrow$ Proteome $\Rightarrow$ Dynamics of cells and organisms

Goals:

1. Large scale simulation of genetic regulatory and metabolic reaction networks.

2. Understanding the dynamics of cells and organisms. Are kinetic differential equations applicable to spatially highly heterogeneous media with active transport?

3. Design of genetic and metabolic model systems, which allow for **optimization** through **evolution** and which provide explanations for the unique properties of living cells and organisms like **robustness**, **homeostasis**, and **adaptation** to environmental changes.

Structural biology

Sequence $\Rightarrow$ Structure $\Rightarrow$ Function

Systems biology

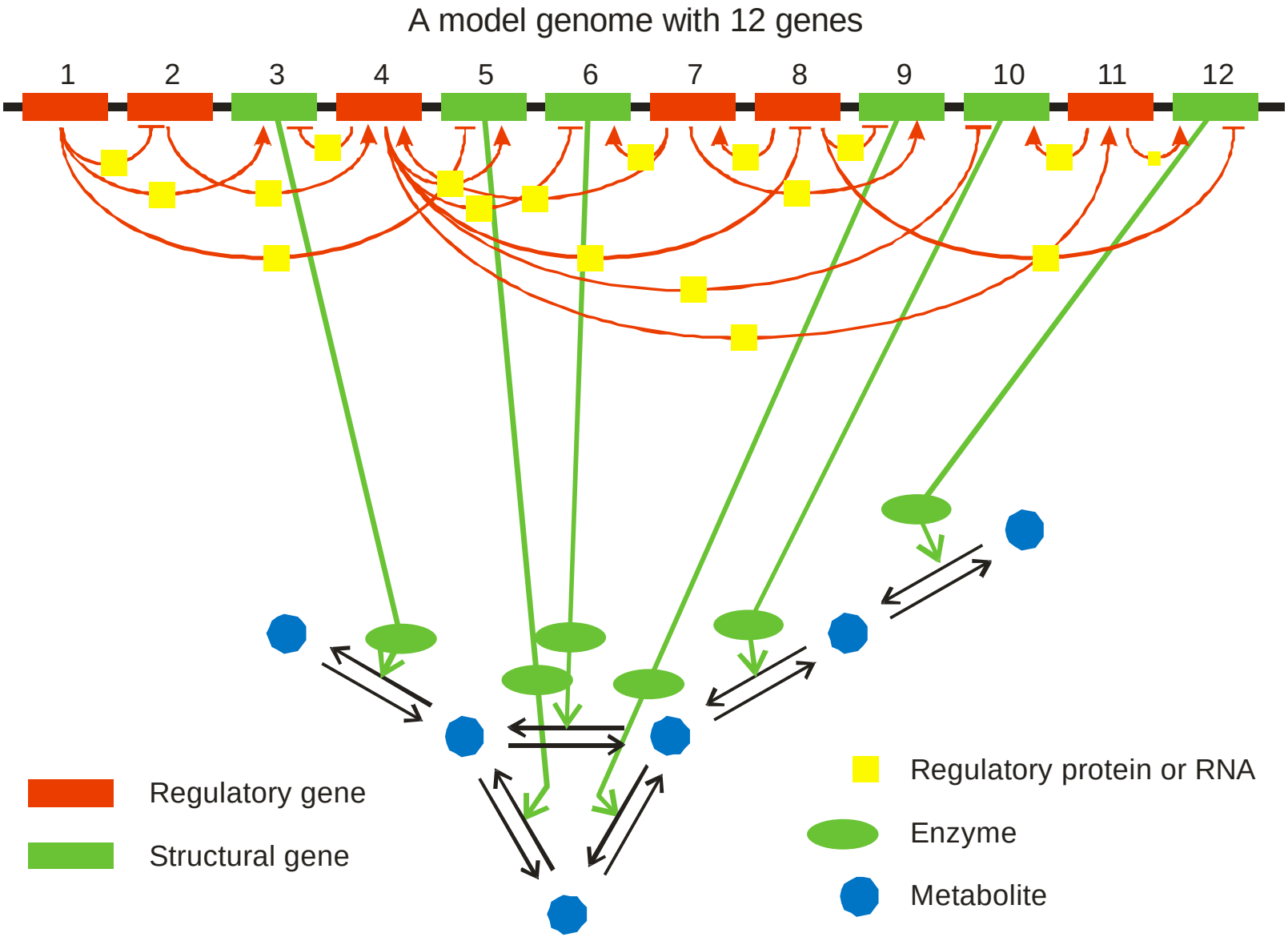
Genome $\Rightarrow$ Proteome $\Rightarrow$ Dynamics of cells and organisms

Goals:

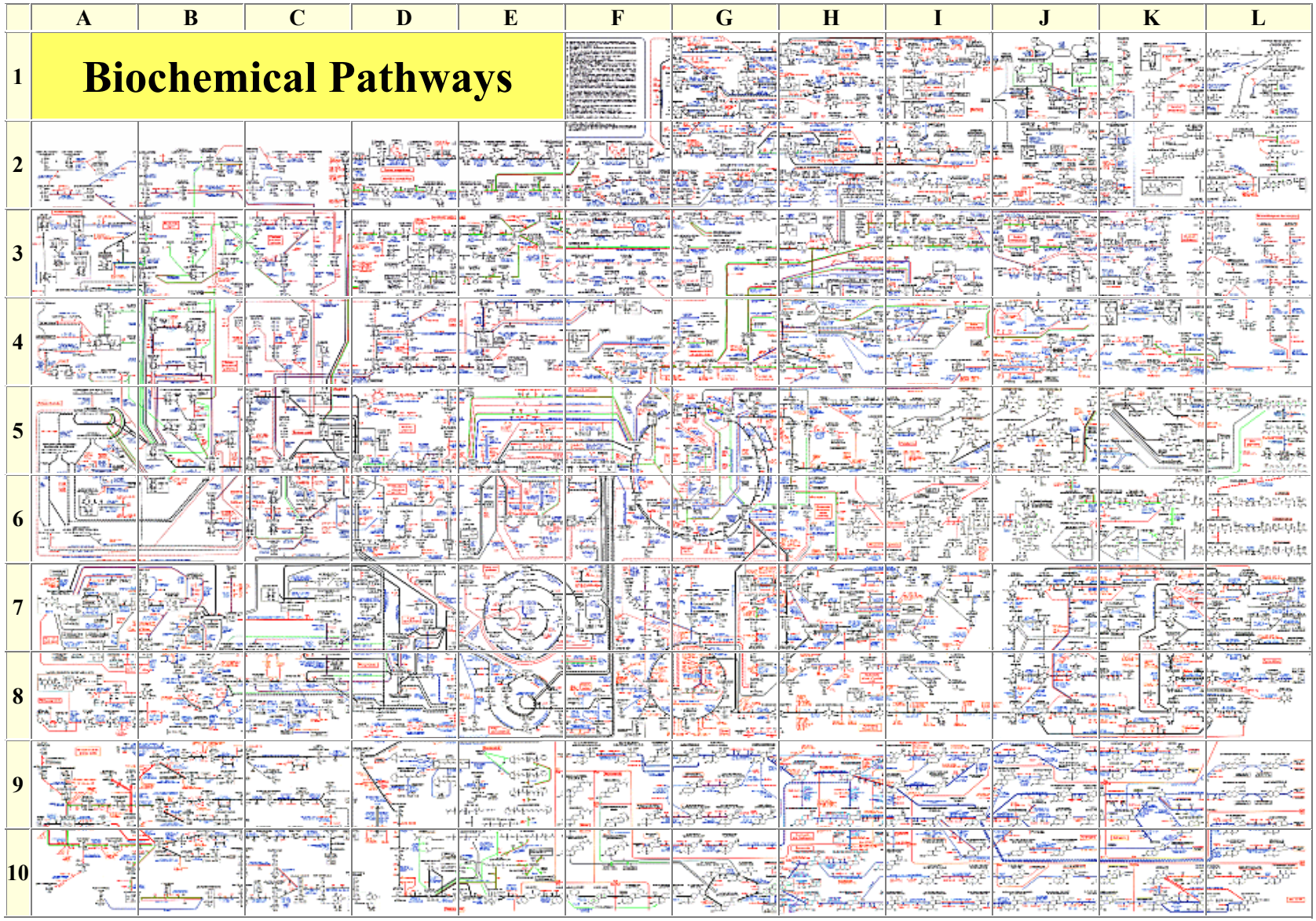
1. Large scale simulation of genetic regulatory and metabolic reaction networks.

2. Understanding of the dynamics of cells and organisms. Are kinetic differential equations applicable to spatially highly heterogeneous media with active transport?

3. Design of genetic and metabolic model systems, which allow for **optimization** through **evolution** and which provide explanations for the unique properties of living cells and organisms like **robustness**, **homeostasis**, and **adaptation** to environmental changes.

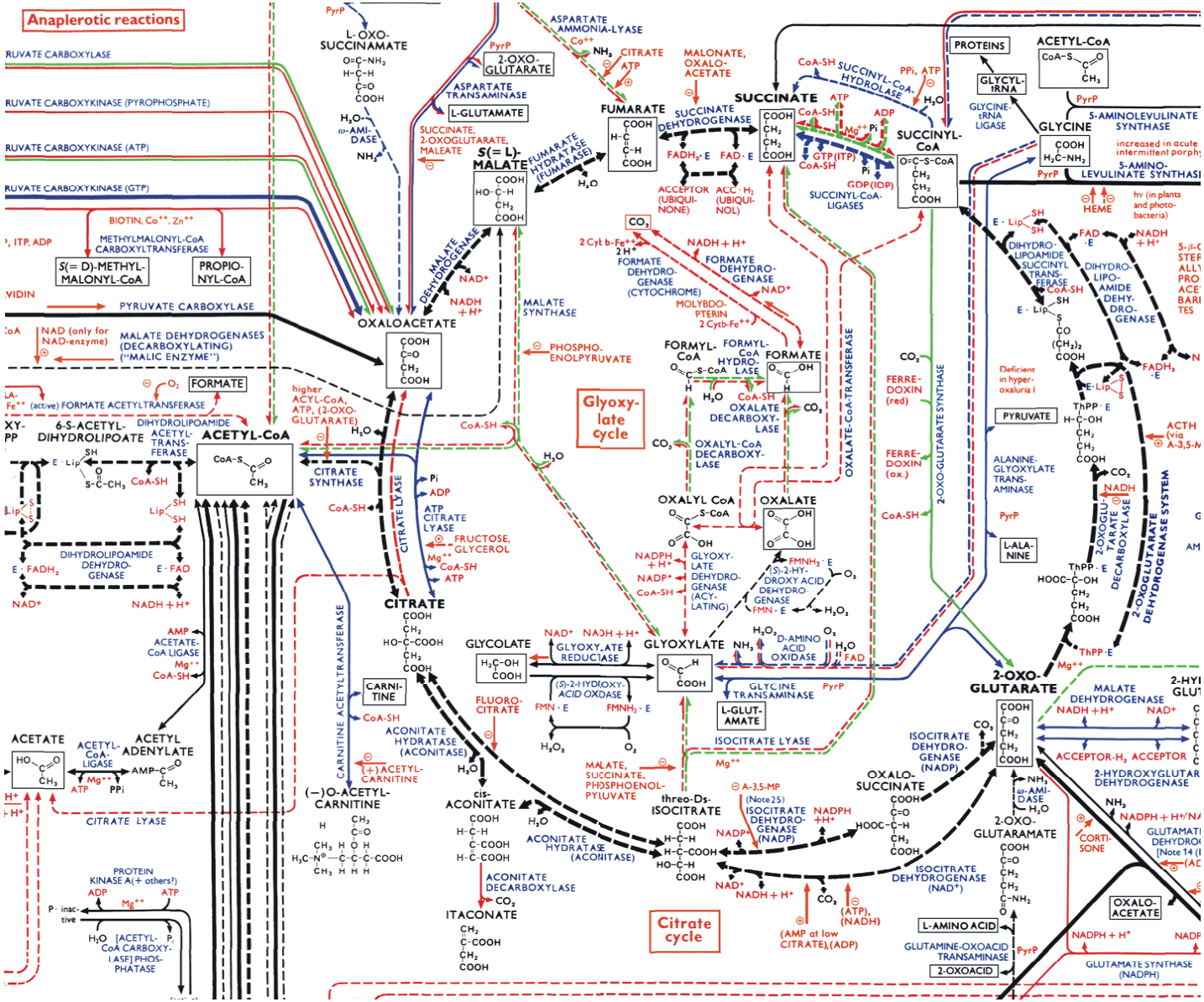


Sketch of a genetic and metabolic network



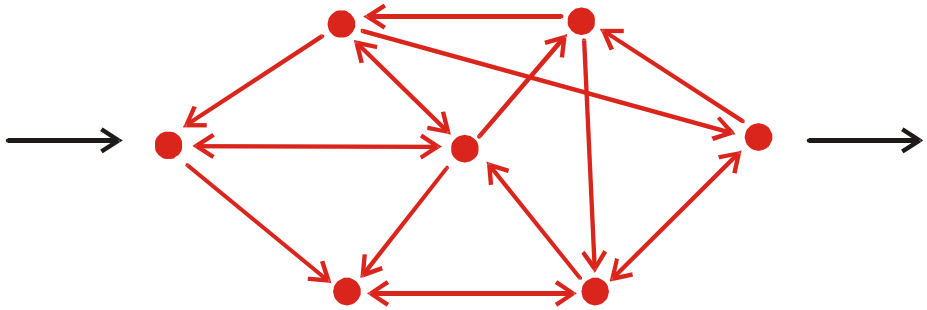
The reaction network of cellular metabolism published by Boehringer-Ingelheim.

The citric acid or Krebs cycle (enlarged from previous slide).



1. What is computational systems biology?
2. **Networks and network evolution**
3. Forward and inverse problems
4. Reverse engineering – A simple example
5. MiniCellSim – A simulation tool
6. Evolution of genetic and metabolic networks

Linear chain



Network

Processing of information in cascades and networks

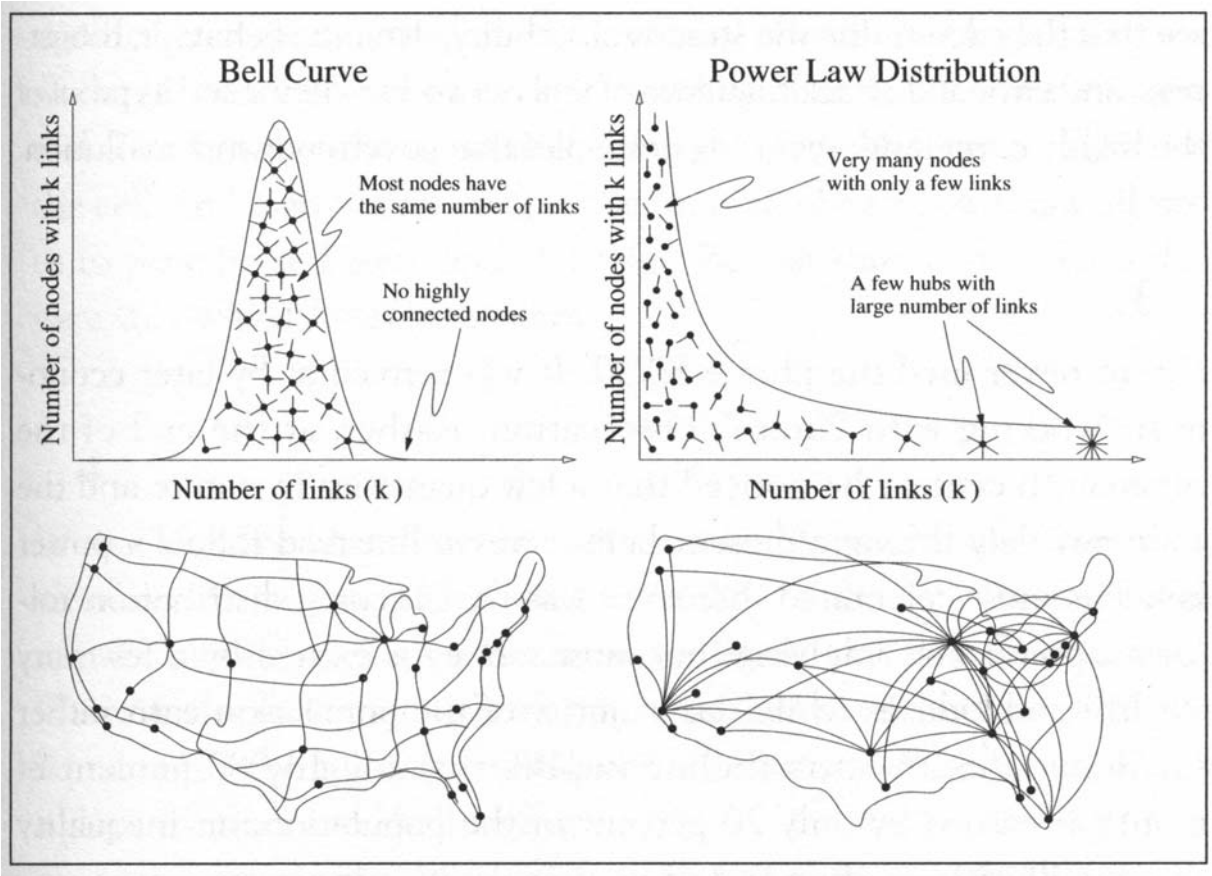
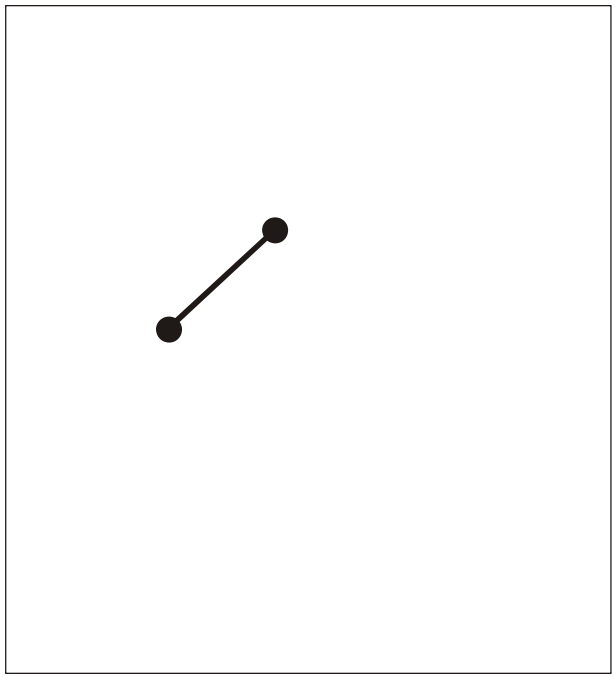
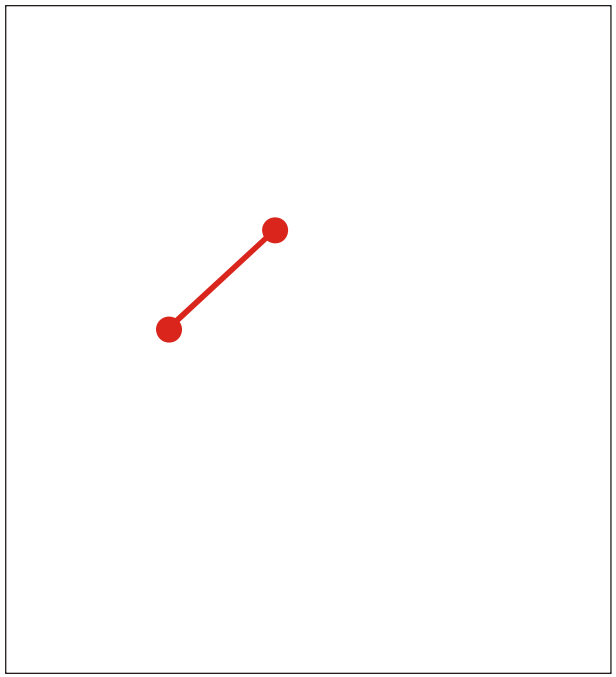


Figure 6.1 Random and Scale-Free Networks. *The degree distribution of a random network follows a bell curve, telling us that most nodes have the same number of links, and nodes with a very large number of links don't exist (top left). Thus a random network is similar to a national highway network, in which the nodes are the cities, and the links are the major highways connecting them. Indeed, most cities are served by roughly the same number of highways (bottom left). In contrast, the power law degree distribution of a scale-free network predicts that most nodes have only a few links, held together by a few highly connected hubs (top right). Visually this is very similar to the air traffic system, in which a large number of small airports are connected to each other via a few major hubs (bottom right).*

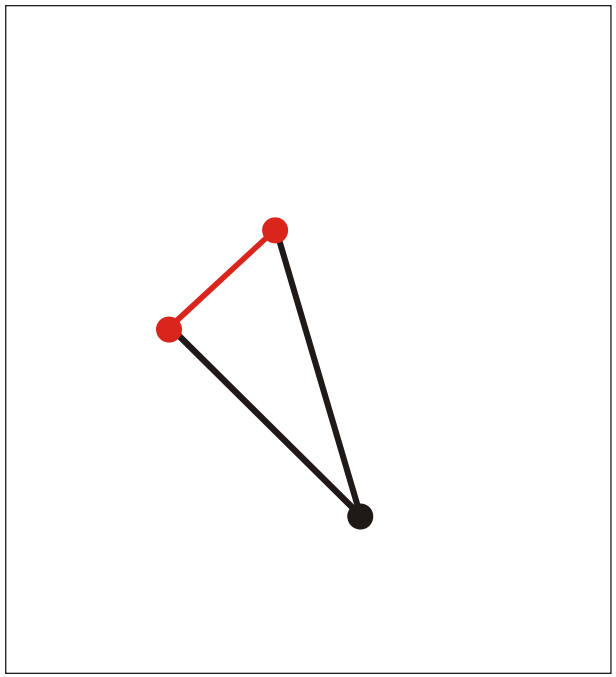
Albert-László Barabási,
Linked – The New Science of Networks
Perseus Publ., Cambridge, MA, 2002



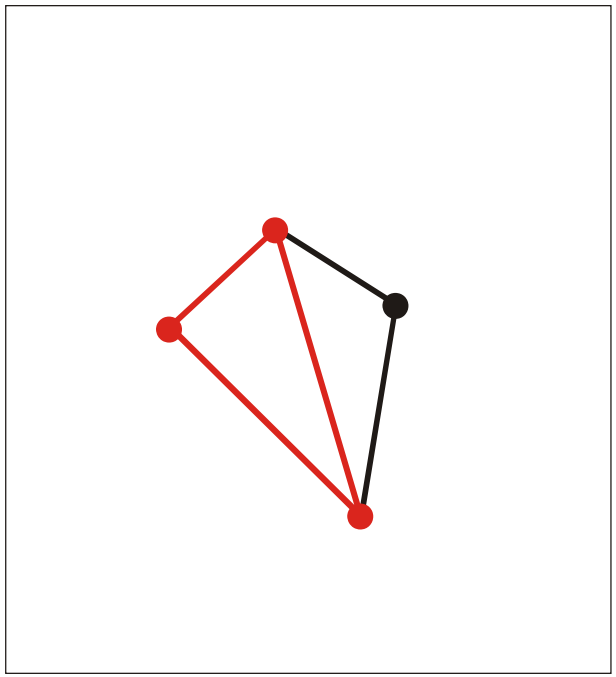
Formation of a scale-free network through evolutionary point by point expansion: Step 000



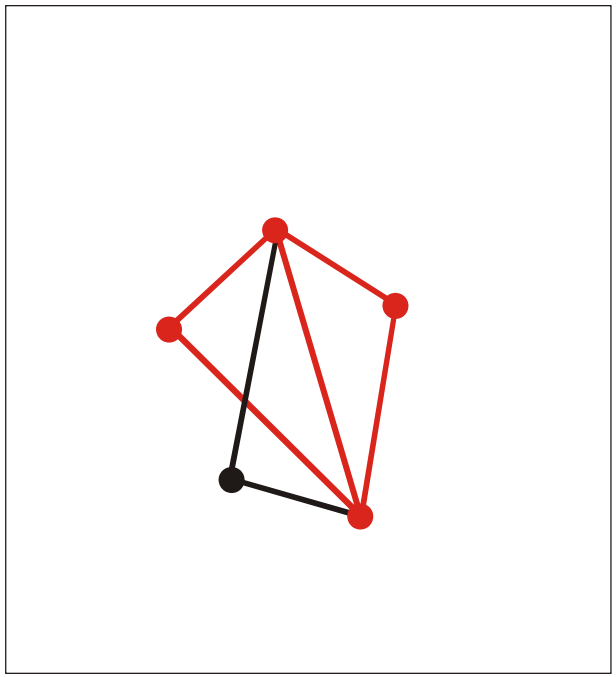
Formation of a scale-free network through evolutionary point by point expansion: Step 001



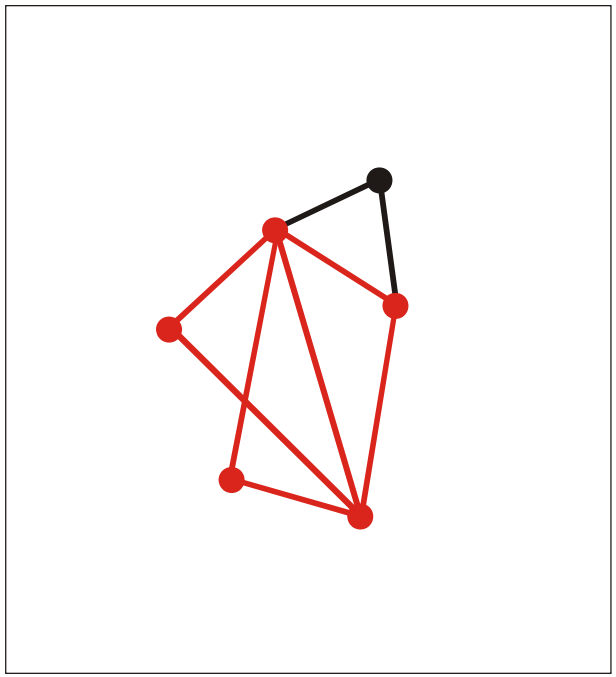
Formation of a scale-free network through evolutionary point by point expansion: Step 002



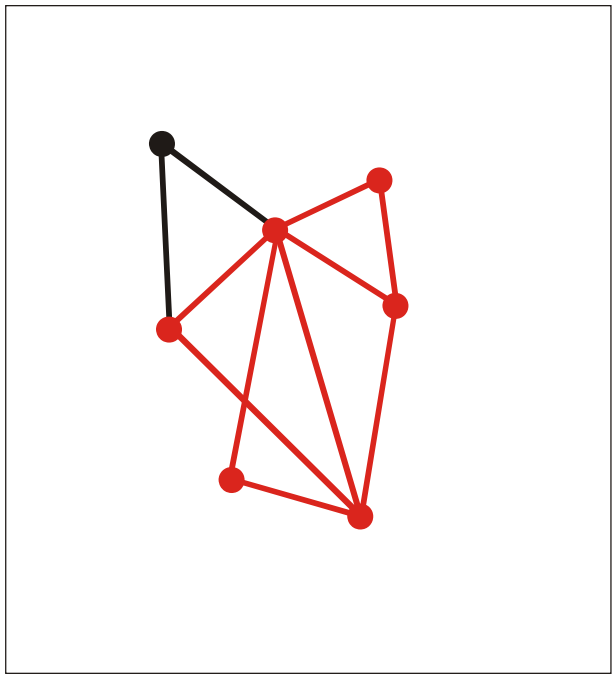
Formation of a scale-free network through evolutionary point by point expansion: Step 003



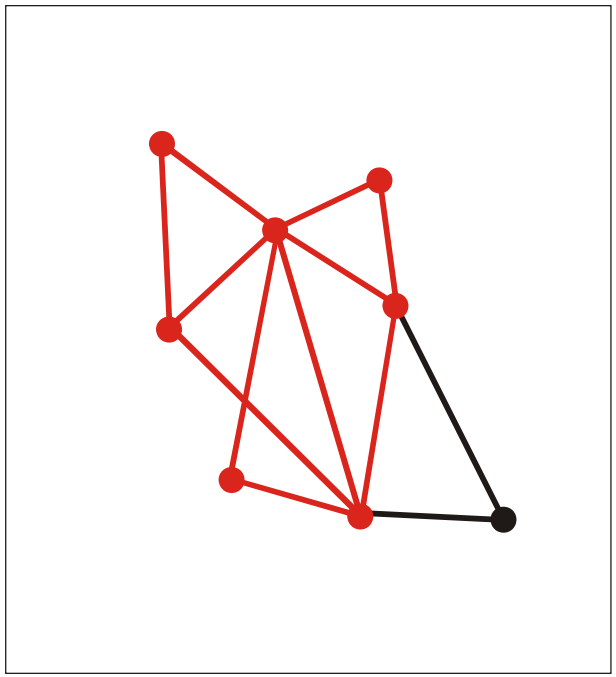
Formation of a scale-free network through evolutionary point by point expansion: Step 004



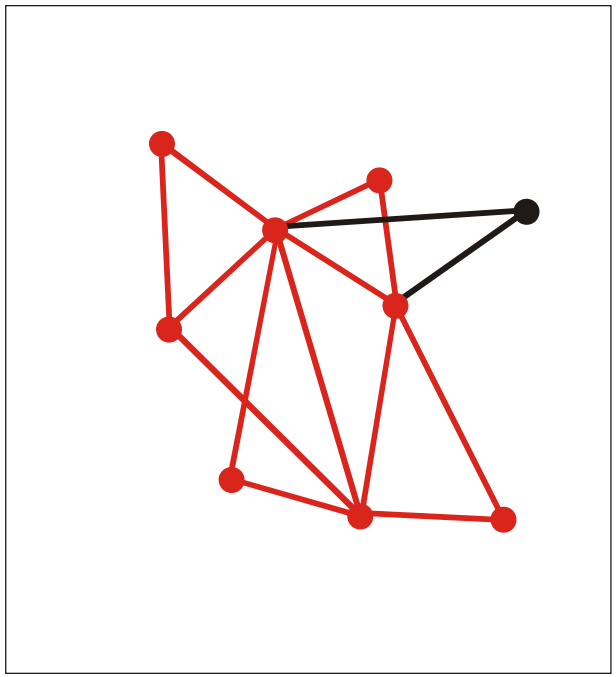
Formation of a scale-free network through evolutionary point by point expansion: Step 005



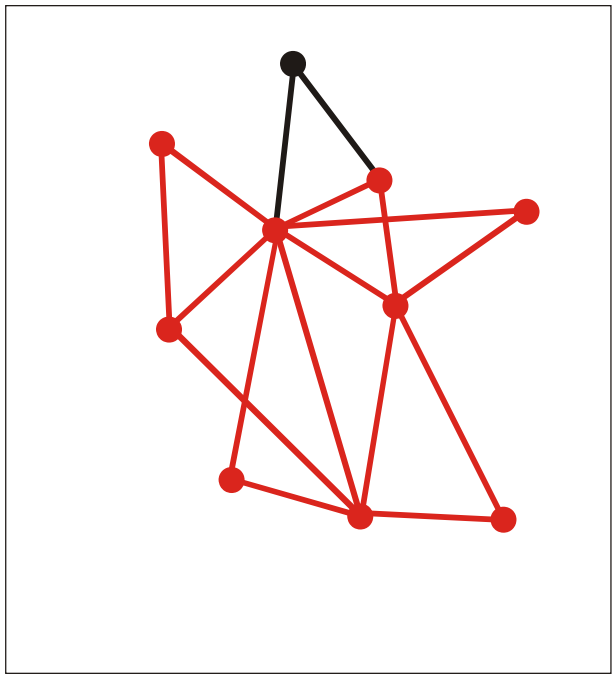
Formation of a scale-free network through evolutionary point by point expansion: Step 006



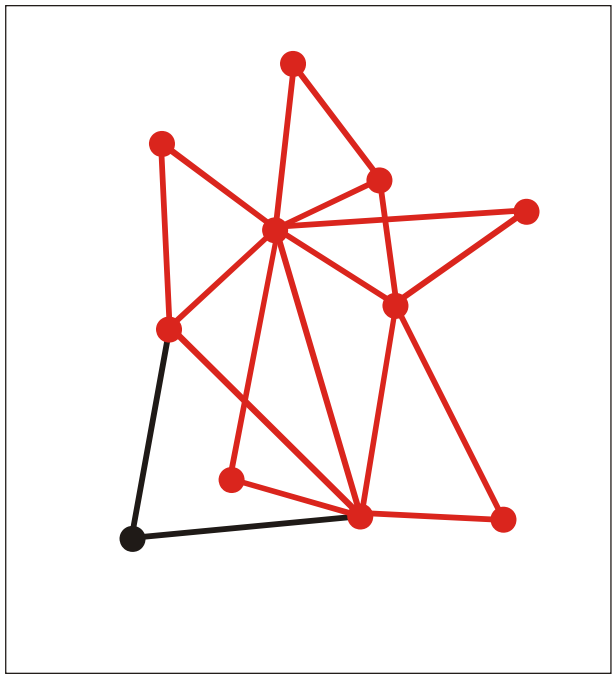
Formation of a scale-free network through evolutionary point by point expansion: Step 007



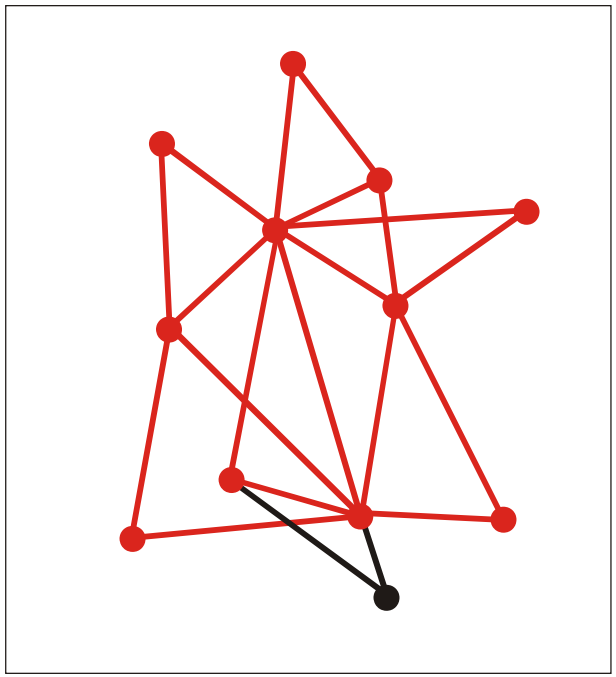
Formation of a scale-free network through evolutionary point by point expansion: Step 008



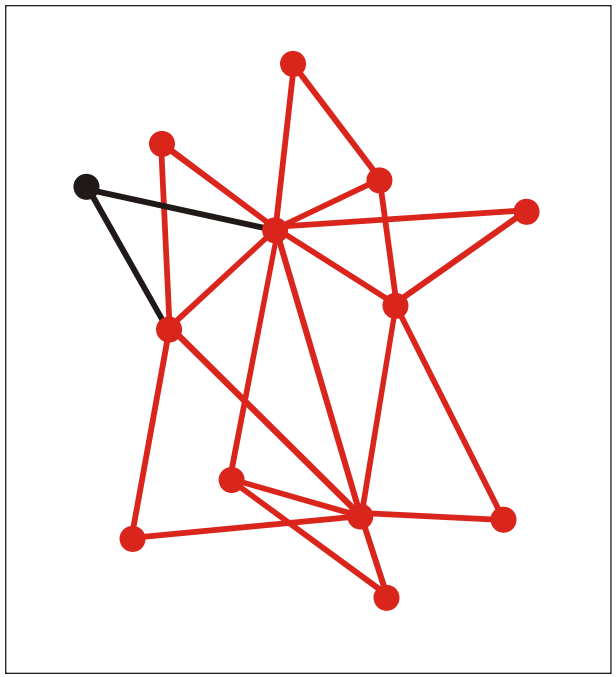
Formation of a scale-free network through evolutionary point by point expansion: Step 009



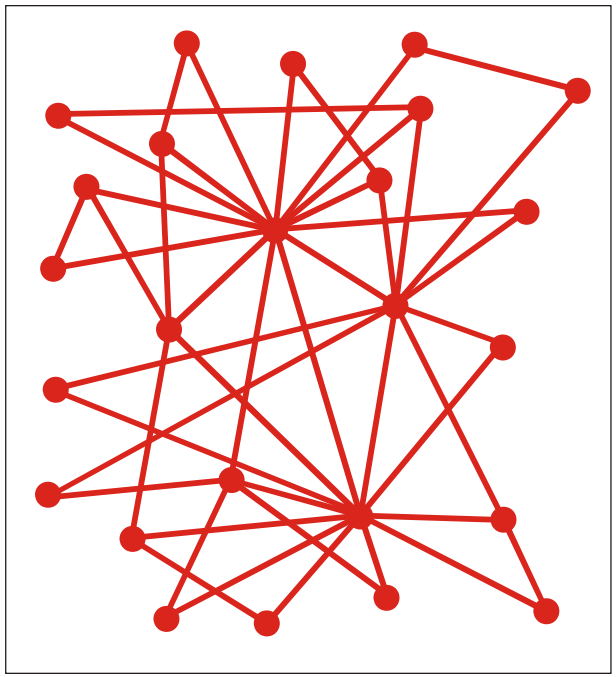
Formation of a scale-free network through evolutionary point by point expansion: Step 010



Formation of a scale-free network through evolutionary point by point expansion: Step 011

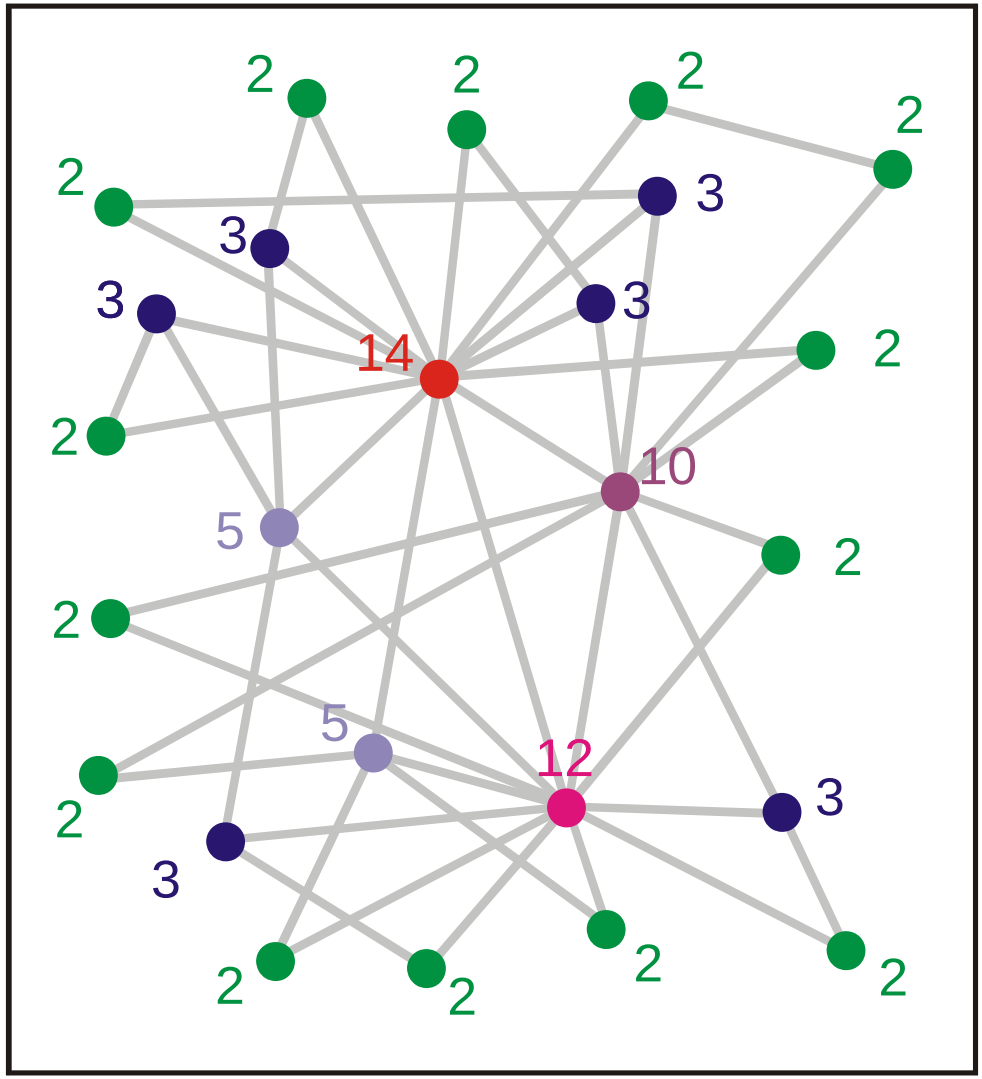


Formation of a scale-free network through evolutionary point by point expansion: Step 012



Formation of a scale-free network through evolutionary point by point expansion: Step 024

links	# nodes
2	14
3	6
5	2
10	1
12	1
14	1



Analysis of nodes and links in a step by step evolved network

1. What is systems biology?
2. Networks and network evolution
3. **Forward and inverse problems**
4. Reverse engineering – A simple example
5. MiniCellSim – A simulation tool
6. Evolution of genetic and metabolic networks

GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA sequence

RNA folding:

Structural biology, spectroscopy of biomolecules, understanding **molecular function**

Iterative determination of a sequence for the given secondary structure

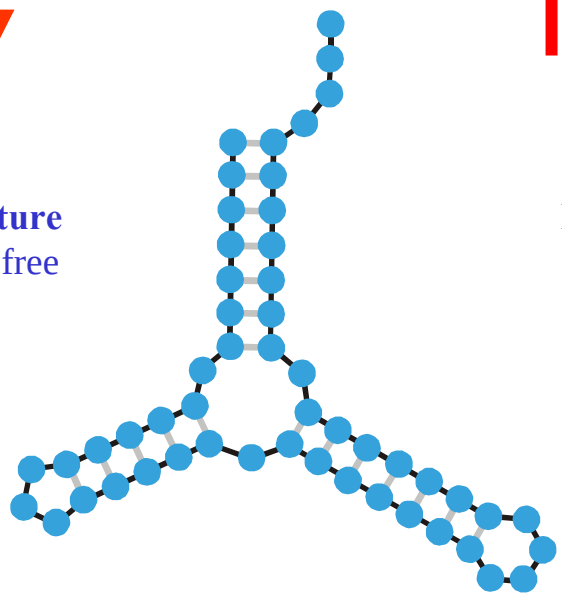
Inverse Folding Algorithm

RNA sequence that forms the structure as minimum free energy structure

Inverse folding of RNA:

Biotechnology, **design of biomolecules** with predefined structures and functions

RNA structure of minimal free energy



RNA structure

Sequence, structure, and design through inverse folding

Kinetic differential equations

$$\frac{dx}{dt} = f(x;k);\ x=(x_1,\ldots,x_n);\ k=(k_1,\ldots,k_m)$$

Reaction diffusion equations

$$\frac{\partial x}{\partial t} = D\,\nabla^2 x + f(x;k)$$

Parameter set

$k_j(\mathrm{T},\mathrm{p},\mathrm{pH},\mathrm{I},\ldots);\ \mathrm{j}=1,2,\ldots,\mathrm{m}$

General conditions: T, p, pH, I, ...

Initial conditions: $x(0)$

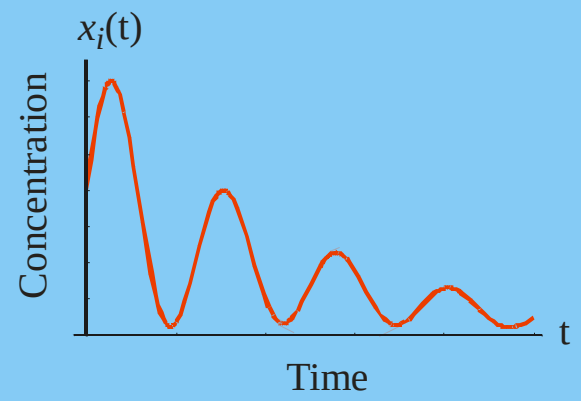
Boundary conditions:

boundary ... S, normal unit vector ... $\hat{u}$

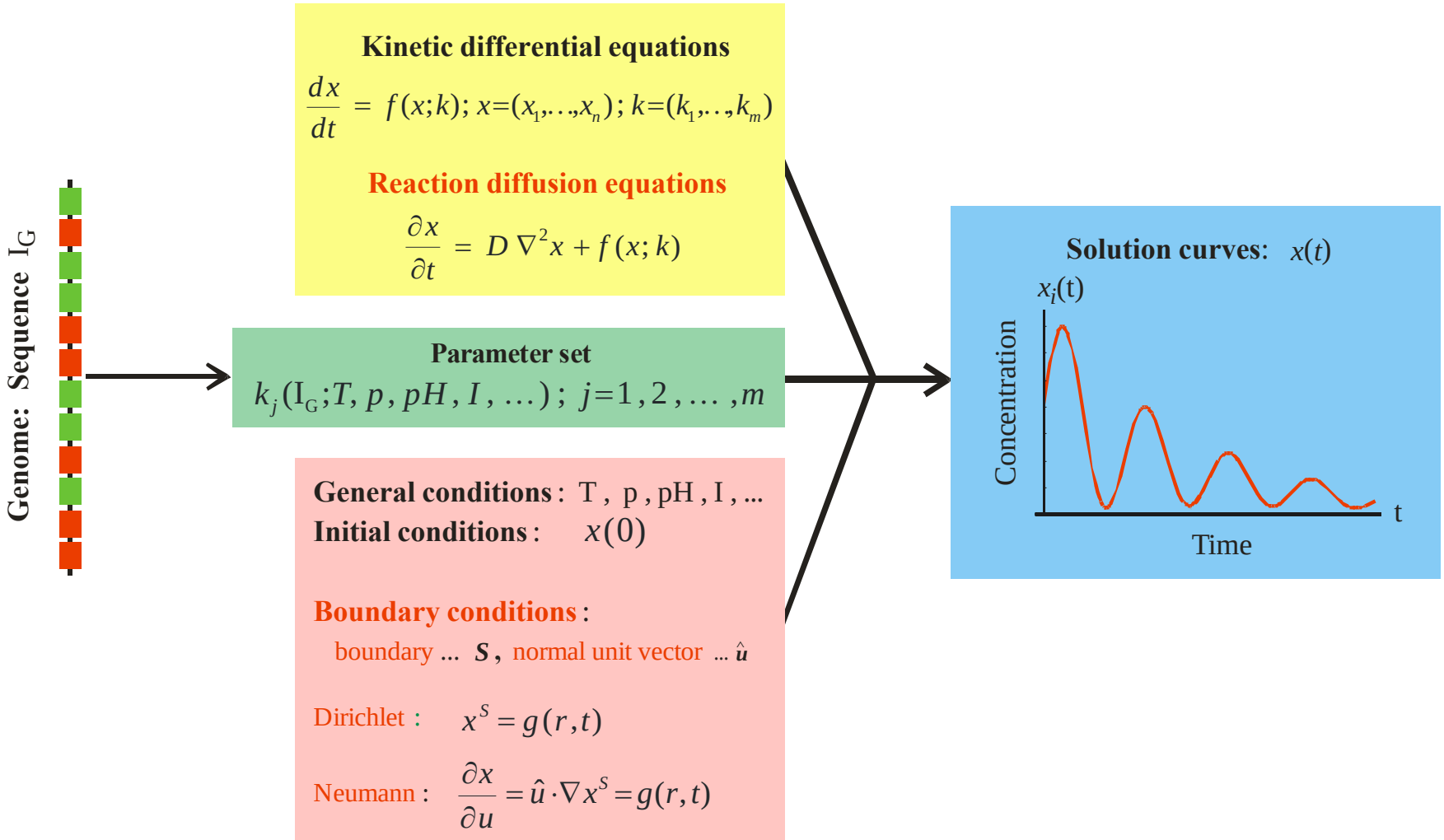
Dirichlet : $x^S = g(r,t)$

Neumann : $\frac{\partial x}{\partial u} = \hat{u}\cdot\nabla x^S = g(r,t)$

Solution curves: $x(t)$



The forward problem of chemical reaction kinetics (Level I)



The forward problem of biochemical reaction kinetics (Level I)

Parameter set

$$k_j(I_G; T, p, pH, I, \ldots);\ j = 1, 2, \ldots, m$$

Kinetic differential equations

$$\frac{dx}{dt} = f(x;k);\ x=(x_1,\ldots,x_n);\ k=(k_1,\ldots,k_m)$$

Reaction diffusion equations

$$\frac{\partial x}{\partial t} = D\,\nabla^2 x + f(x;k)$$

General conditions : T , p , pH , I , ...

Initial conditions : $x(0)$

Boundary conditions :

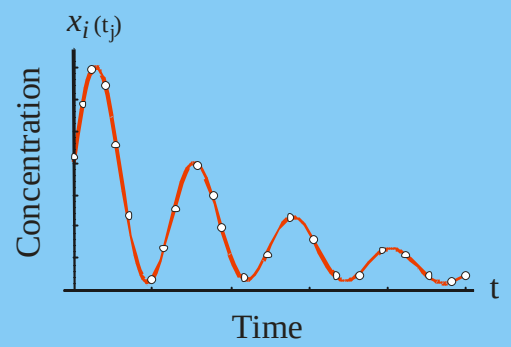
boundary ... S, normal unit vector... $\hat{u}$

Dirichlet : $x^S = g(r,t)$

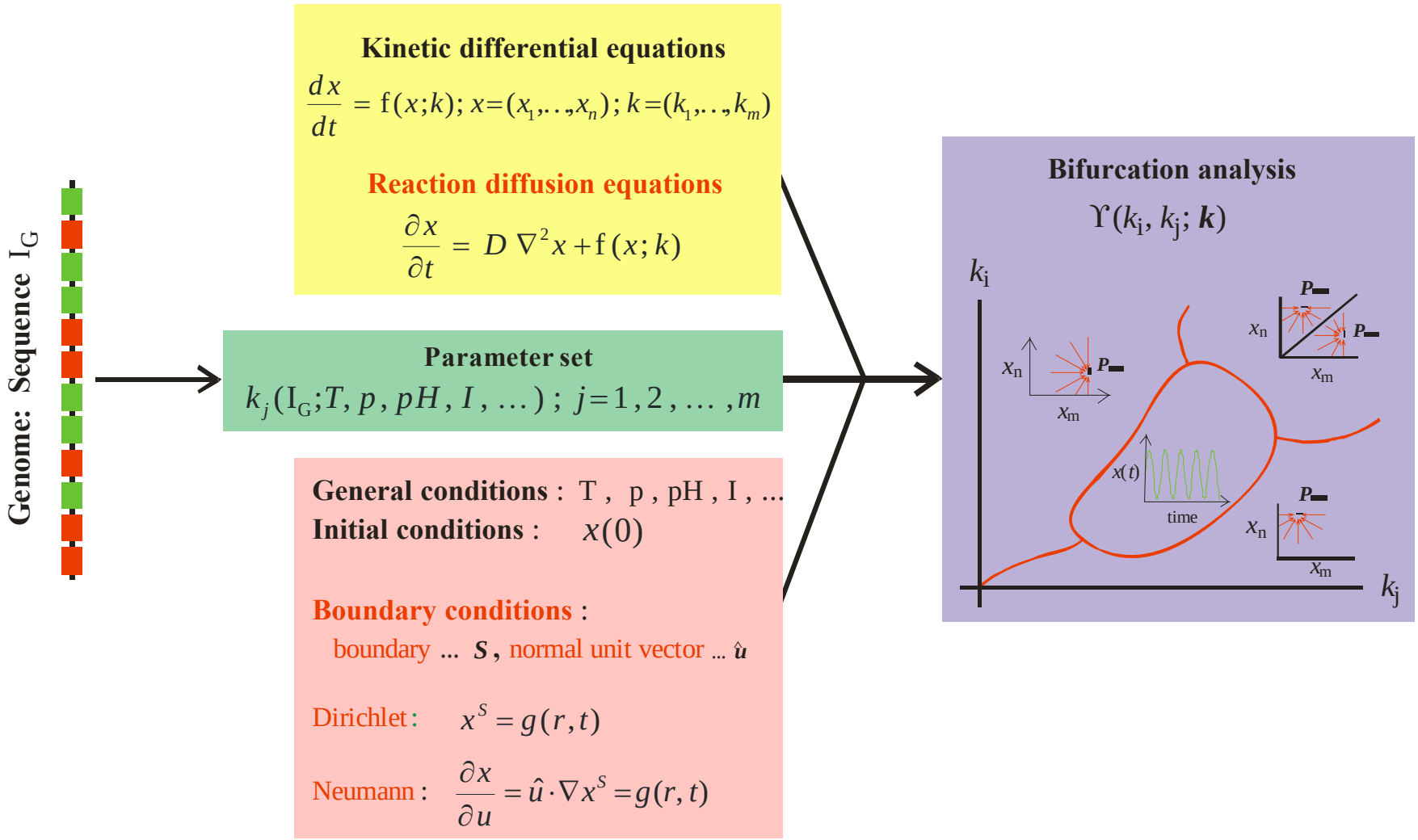
Neumann : $\frac{\partial x}{\partial u} = \hat{u} \cdot \nabla x^S = g(r,t)$

Data from measurements

$x(t_j);\ j = 1, 2, \ldots, N$



The inverse problem of biochemical reaction kinetics (Level I)



The forward problem of bifurcation analysis (Level II)

Genome: Sequence I_G

Parameter set

$k_j(I_G; T, p, pH, I, \ldots);\ j = 1, 2, \ldots, m$

Kinetic differential equations

$$\frac{dx}{dt} = f(x;k);\ x=(x_1,\ldots,x_n);\ k=(k_1,\ldots,k_m)$$

Reaction diffusion equations

$$\frac{\partial x}{\partial t} = D\,\nabla^2 x + f(x;k)$$

General conditions: T , p , pH , I , ...

Initial conditions: $x(0)$

Boundary conditions:

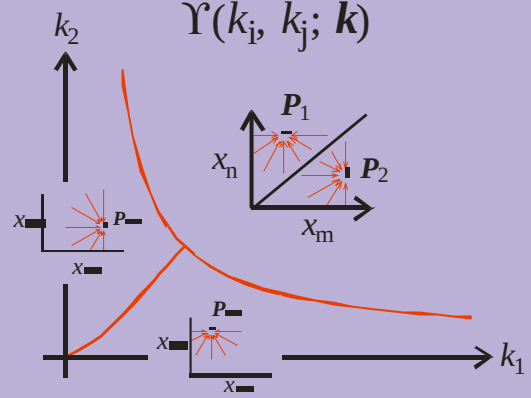
boundary ... S , normal unit vector... $\hat{u}$

Dirichlet : $x^S = g(r,t)$

Neumann : $\frac{\partial x}{\partial u} = \hat{u} \cdot \nabla x^S = g(r,t)$

Bifurcation pattern

$\Upsilon(k_i, k_j; \boldsymbol{k})$



The inverse problem of bifurcation analysis (Level II)

1. What is computational systems biology?
2. Networks and network evolution
3. Forward and inverse problems
4. **Reverse engineering – A simple example**
5. MiniCellSim – A simulation tool
6. Evolution of genetic and metabolic networks

$$* \xrightarrow{r} \mathbf{A}$$

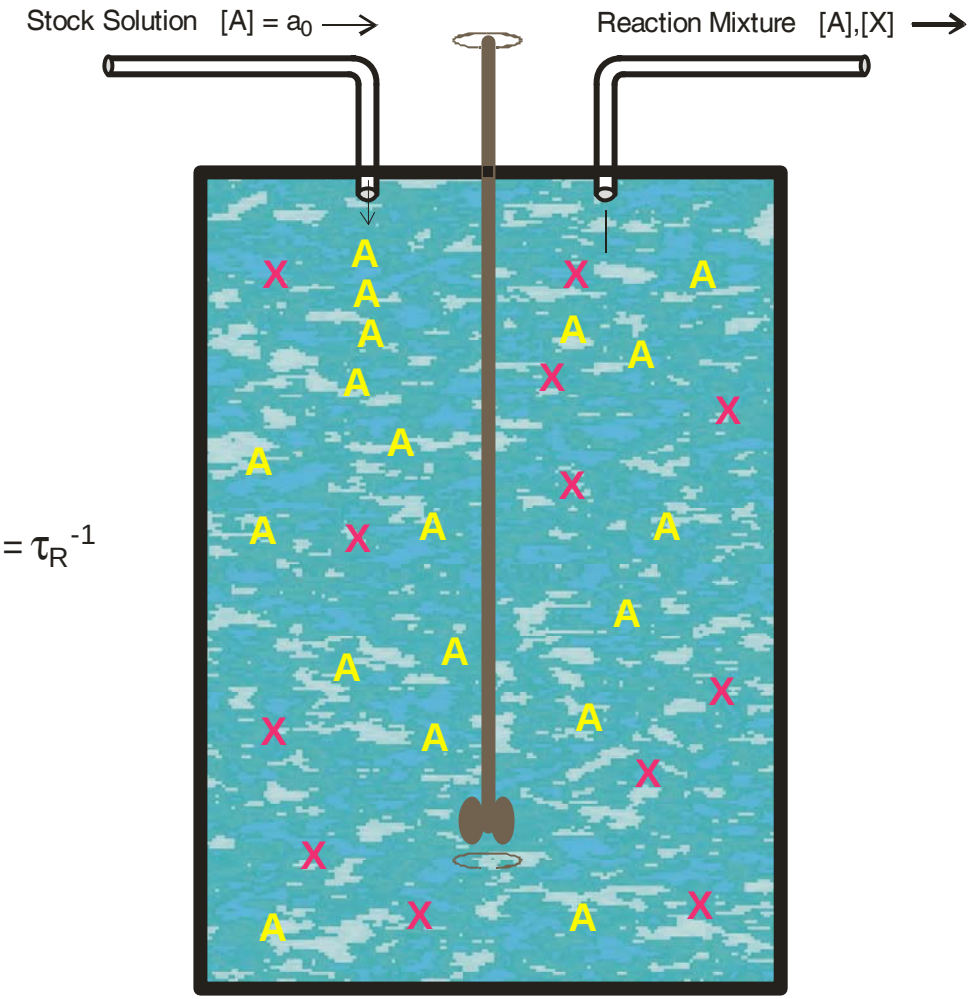
$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{X}$$

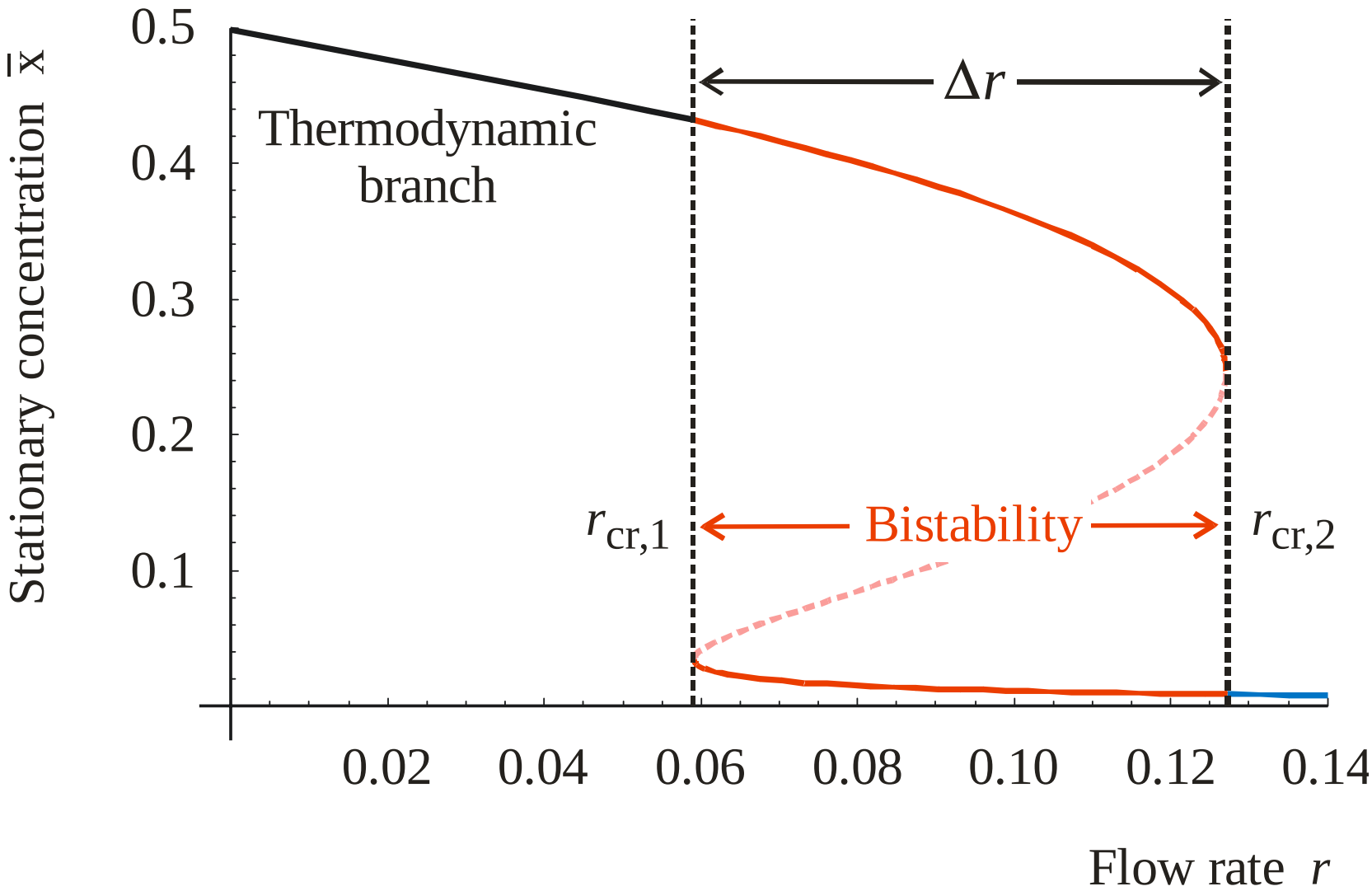
$$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} 3\mathbf{X}$$

$$\mathbf{A} \xrightarrow{r} 0$$

$$\mathbf{X} \xrightarrow{r} 0$$

Flow rate $r = \tau_R^{-1}$




Stationary concentration $\bar{x}$
0.5
0.4
0.3
0.2
0.1
Thermodynamic branch
Δr
$r_{cr,1}$
Bistability
$r_{cr,2}$
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Flow rate r

$$* \xrightarrow{r} \mathbf{A}$$

$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathbf{X}$$

$$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftarrows}} 3\mathbf{X}$$

$$\mathbf{A} \xrightarrow{r} 0$$

$$\mathbf{X} \xrightarrow{r} 0$$

Kinetic differential equations:

$$\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t} = r\,(a_0 - a) - (k_1 + k_3 x^2)\,a + (k_2 + k_4 x^2)\,x$$

$$\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t} = -\,r\,x + (k_1 + k_3 x^2)\,a - (k_2 + k_4 x^2)\,x$$

$$* \xrightarrow{r} \mathbf{A}$$

$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathbf{X}$$

$$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftarrows}} 3\mathbf{X}$$

$$\mathbf{A} \xrightarrow{r} 0$$

$$\mathbf{X} \xrightarrow{r} 0$$

Kinetic differential equations:

$$\frac{\mathrm{d[A]}}{\mathrm{dt}} = \frac{\mathrm{d}a}{\mathrm{dt}} = r\,(a_0 - a) - (k_1 + k_3 x^2)\,a + (k_2 + k_4 x^2)\,x$$

$$\frac{\mathrm{d[X]}}{\mathrm{dt}} = \frac{\mathrm{d}x}{\mathrm{dt}} = -\,r\,x + (k_1 + k_3 x^2)\,a - (k_2 + k_4 x^2)\,x$$

Steady states:

$$\bar{x}^3(k_3 + k_4) - \bar{x}^2 k_3 a_0 + \bar{x}(k_1 + k_2 + r) - k_1 a_0 = 0$$

$$* \xrightarrow{r} \mathbf{A}$$

$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathbf{X}$$

$$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftarrows}} 3\mathbf{X}$$

$$\mathbf{A} \xrightarrow{r} 0$$

$$\mathbf{X} \xrightarrow{r} 0$$

Kinetic differential equations:

$$\frac{\mathrm{d[A]}}{\mathrm{dt}} = \frac{\mathrm{d}a}{\mathrm{dt}} = r\,(a_0 - a) - (k_1 + k_3 x^2)\,a + (k_2 + k_4 x^2)\,x$$

$$\frac{\mathrm{d[X]}}{\mathrm{dt}} = \frac{\mathrm{d}x}{\mathrm{dt}} = -\,r\,x + (k_1 + k_3 x^2)\,a - (k_2 + k_4 x^2)\,x$$

Steady states:

$$\bar{x}^3(k_3+k_4) - \bar{x}^2 k_3 a_0 + \bar{x}(k_1+k_2+r) - k_1 a_0 = 0$$

$$k_1 = k_2 = \alpha,\ k_3 = k_4 = 1: \quad 2\bar{x}^3 - \bar{x}^2 a_0 + \bar{x}(r+2\alpha) - \alpha a_0 = 0$$

$$* \xrightarrow{r} \mathbf{A}$$

Kinetic differential equations:

$$\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t} = r\,(a_0 - a) - (k_1 + k_3 x^2)\,a + (k_2 + k_4 x^2)\,x$$

$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{X}$$

$$\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t} = -\,r\,x + (k_1 + k_3 x^2)\,a - (k_2 + k_4 x^2)\,x$$

$$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} 3\mathbf{X}$$

Steady states:

$$\bar{x}^3(k_3+k_4) - \bar{x}^2 k_3 a_0 + \bar{x}(k_1+k_2+r) - k_1 a_0 = 0$$

$$\mathbf{A} \xrightarrow{r} 0$$

$$k_1 = k_2 = \alpha,\ k_3 = k_4 = 1: \quad 2\bar{x}^3 - \bar{x}^2 a_0 + \bar{x}(r+2\alpha) - \alpha a_0 = 0$$

Polynomial discriminant of the cubic equation:

$$\mathbf{X} \xrightarrow{r} 0$$

$$216\,\mathrm{D} = r^3 + r^2\left(6\alpha - \frac{{a_0}^2}{8}\right) + r\,(12\alpha^2 - 5\alpha\,{a_0}^2) + 8\alpha^3 + 4\alpha^2 {a_0}^2 + \frac{\alpha\,{a_0}^4}{2} = 0$$

$\ast \xrightarrow{r} \mathbf{A}$

Kinetic differential equations:

$$\frac{d[A]}{dt} = \frac{da}{dt} = r\,(a_0 - a) - (k_1 + k_3 x^2)\,a + (k_2 + k_4 x^2)\,x$$

$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{X}$

$$\frac{d[X]}{dt} = \frac{dx}{dt} = -r\,x + (k_1 + k_3 x^2)\,a - (k_2 + k_4 x^2)\,x$$

$\mathbf{A} + 2\,\mathbf{X} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} 3\mathbf{X}$

Steady states:

$$\bar{x}^3(k_3+k_4) - \bar{x}^2 k_3 a_0 + \bar{x}(k_1+k_2+r) - k_1 a_0 = 0$$

$\mathbf{A} \xrightarrow{r} 0$

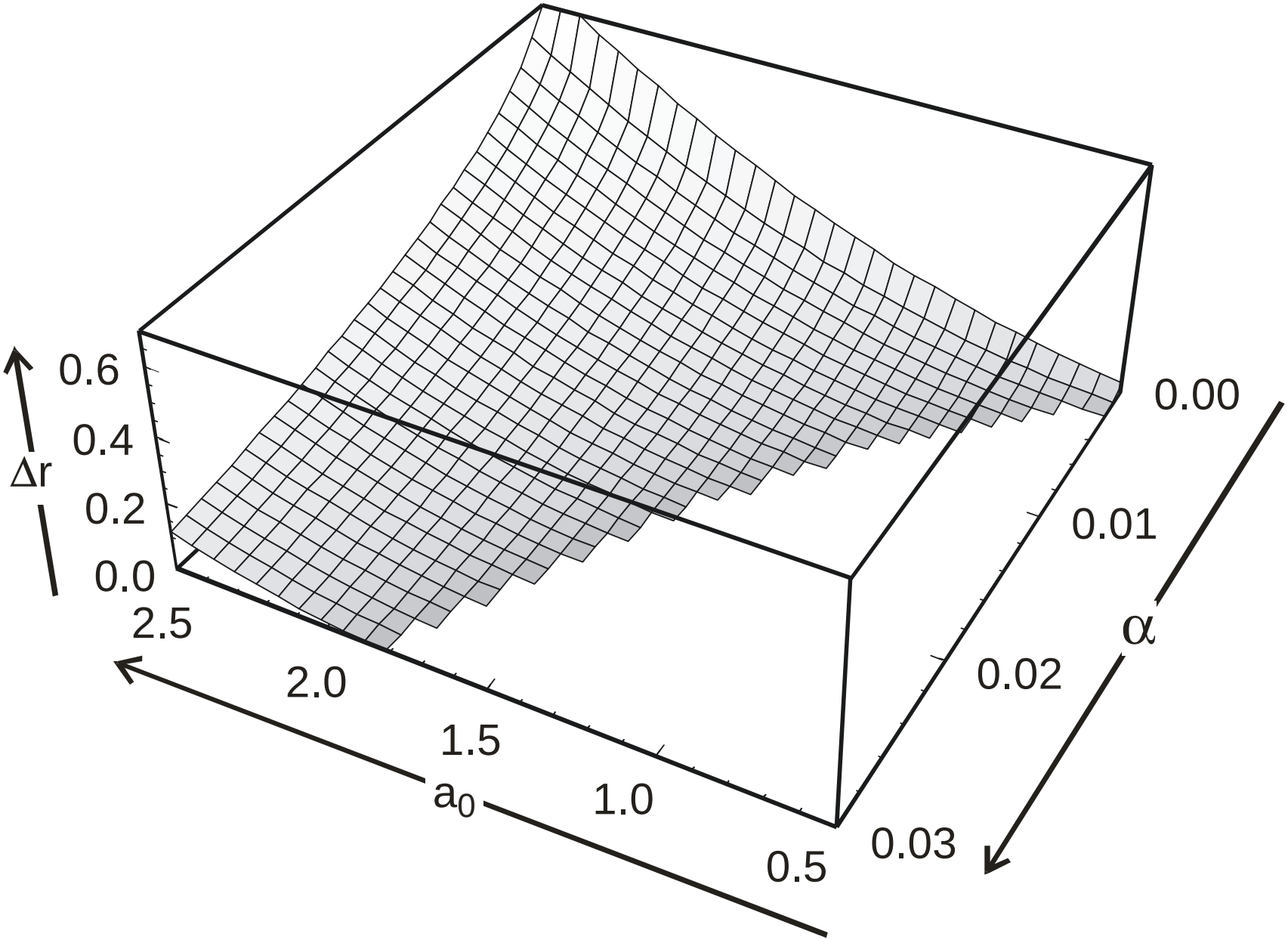
$$k_1 = k_2 = \alpha,\ k_3 = k_4 = 1: \quad 2\bar{x}^3 - \bar{x}^2 a_0 + \bar{x}(r+2\alpha) - \alpha a_0 = 0$$

Polynomial discriminant of the cubic equation:

$\mathbf{X} \xrightarrow{r} 0$

$$216\,D = r^3 + r^2\left(6\alpha - \frac{{a_0}^2}{8}\right) + r\,(12\alpha^2 - 5\alpha\,{a_0}^2) + 8\alpha^3 + 4\alpha^2 {a_0}^2 + \frac{\alpha\,{a_0}^4}{2} = 0$$

D < 0: 3 roots r_1, r_2, and r_3, 2 are positive $\Rightarrow$ $\Delta\mathbf{r} = r_1 - r_2$



Range of hysteresis as a function of the parameters a_0 and α

1. What is computational systems biology?
2. Networks and network evolution
3. Forward and inverse problems
4. Reverse engineering – A simple example
5. **MiniCellSim – A simulation tool**
6. Evolution of genetic and metabolic networks

$$A + B \rightarrow X$$
$$2\,X \rightarrow Y$$
$$Y + X \rightarrow D$$

Stoichiometric equations

↓

SBML – systems biology markup language

↓

Kinetic differential equations

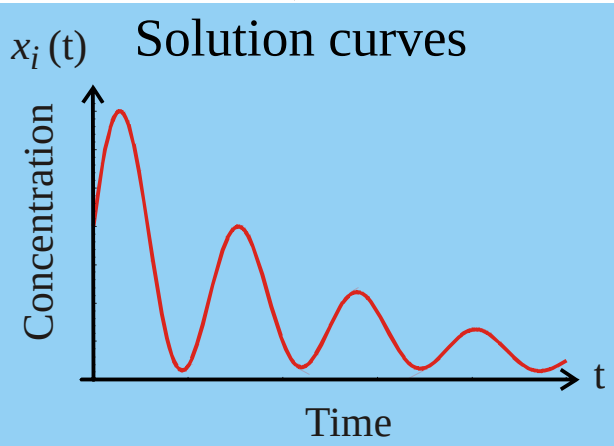
$$\frac{da}{dt} = \frac{db}{dt} = -k_1 a b$$
$$\frac{dx}{dt} = k_1 a b - k_2 x^2 - k_3 x y$$
$$\frac{dy}{dt} = k_2 x^2 - k_3 x y$$
$$\frac{dd}{dt} = k_3 x y$$

↓

ODE Integration by means of *CVODE*

↓

Solution curves



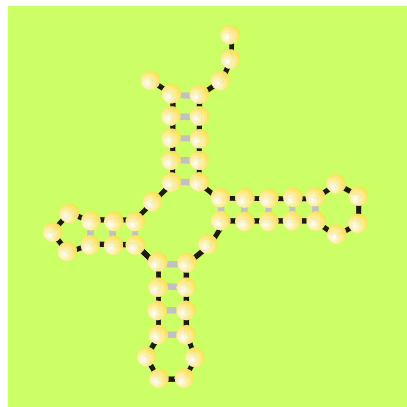
Sequences

↓

Vienna RNA Package

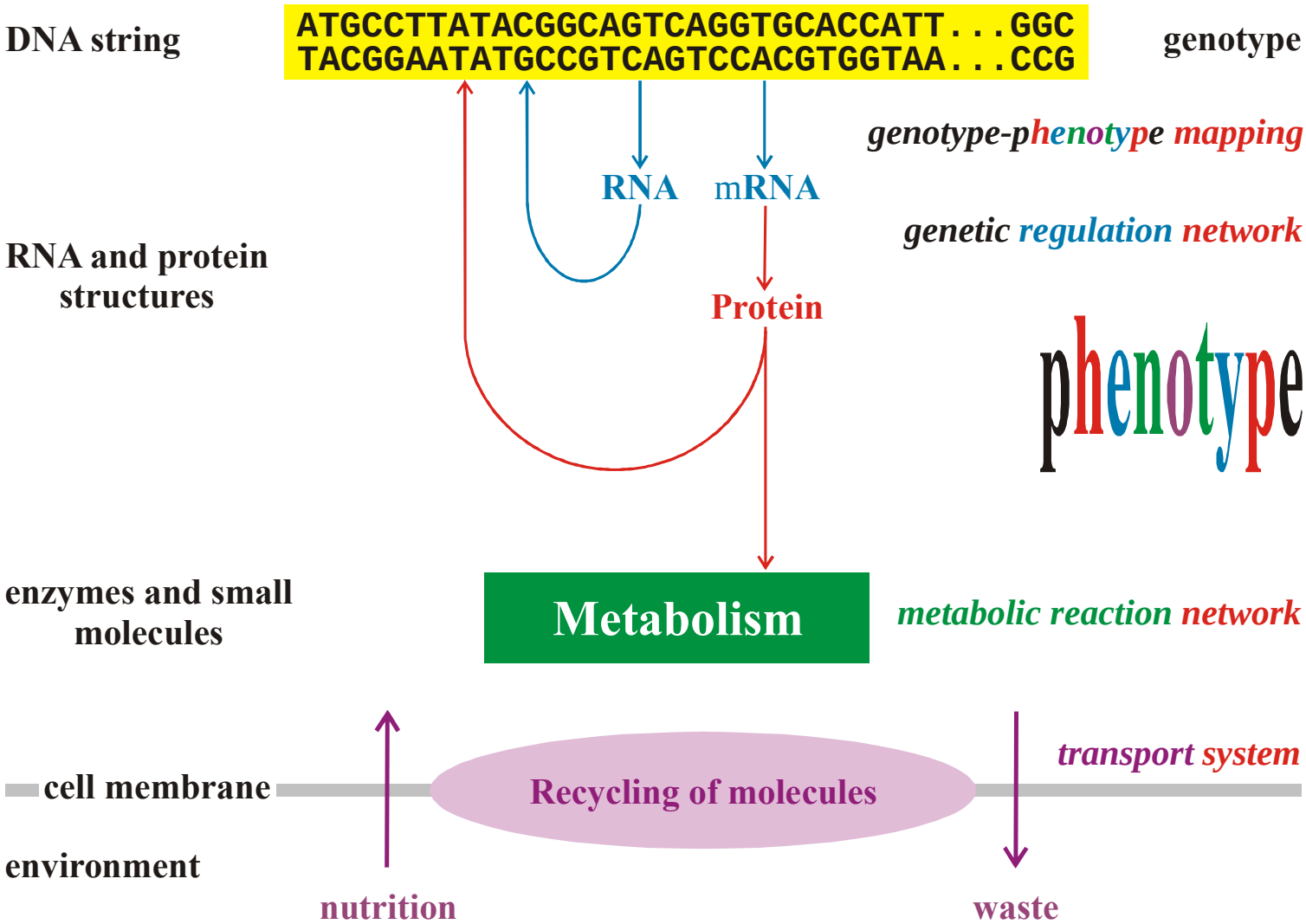
↓

Structures and kinetic parameters

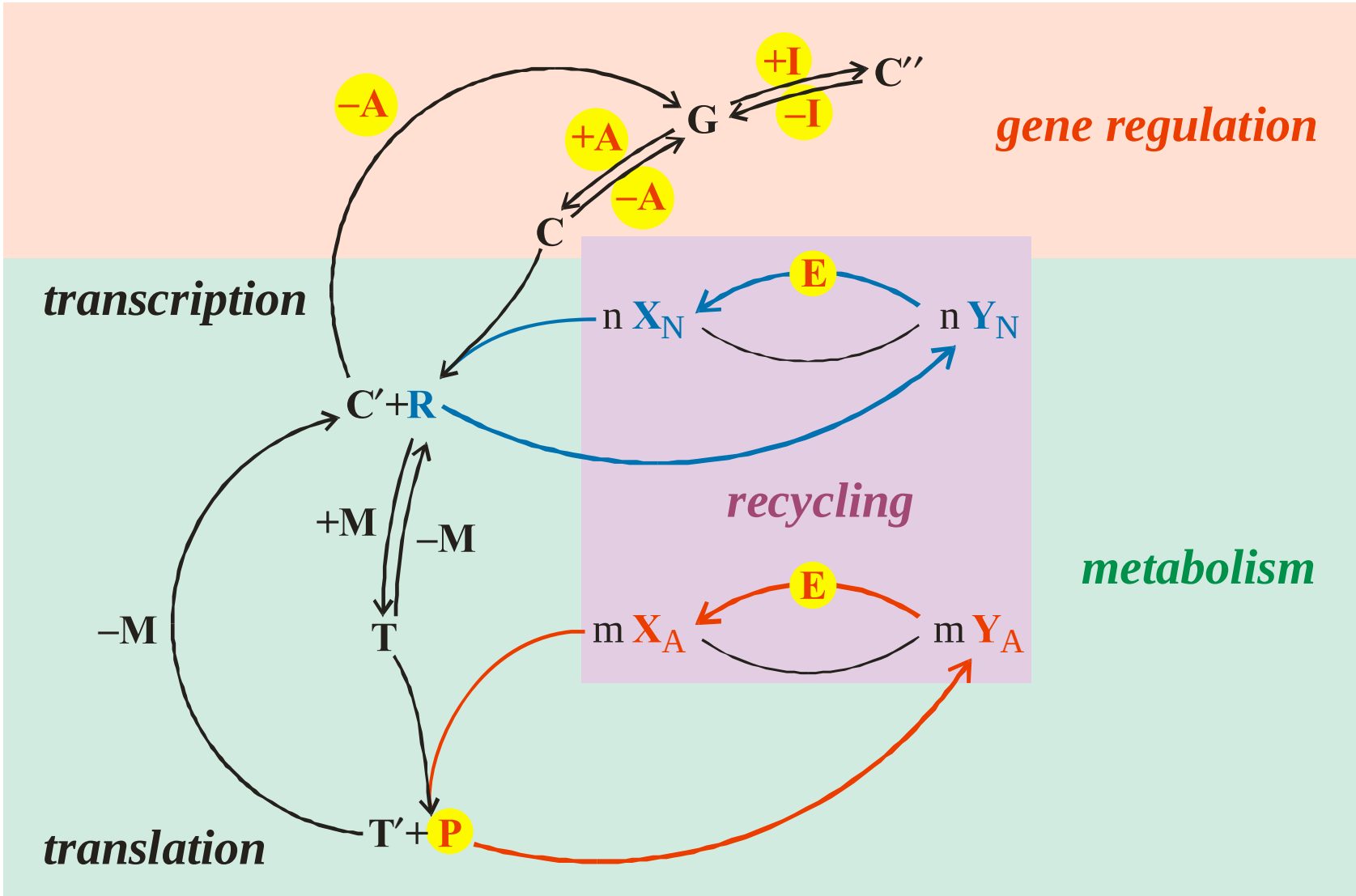


The elements of the simulation tool MiniCellSim

SBML: *Bioinformatics* **19**:524-531, 2003; *CVODE*: *Computers in Physics* **10**:138-143, 1996



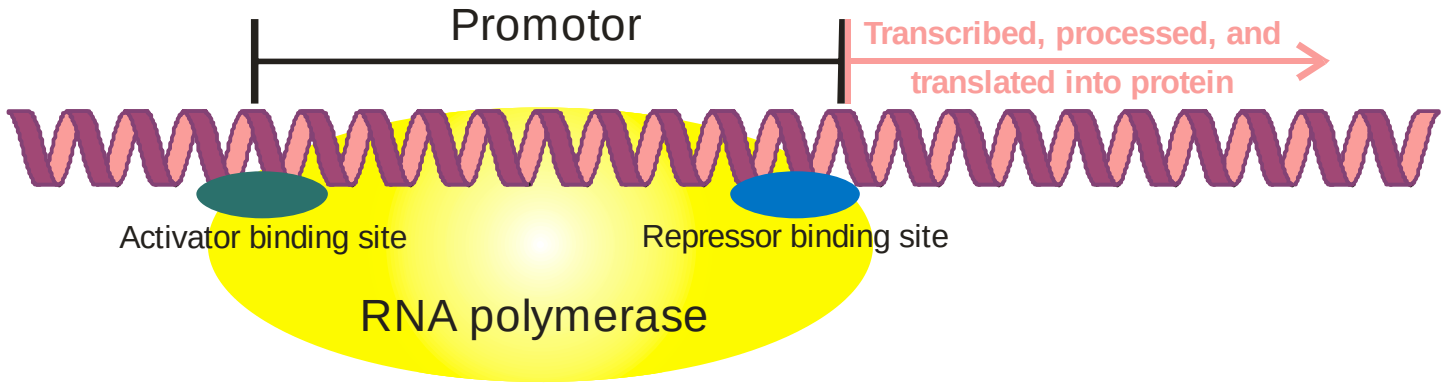
The regulatory logic of MiniCellSym



The chemical reaction dynamics of MiniCellSym

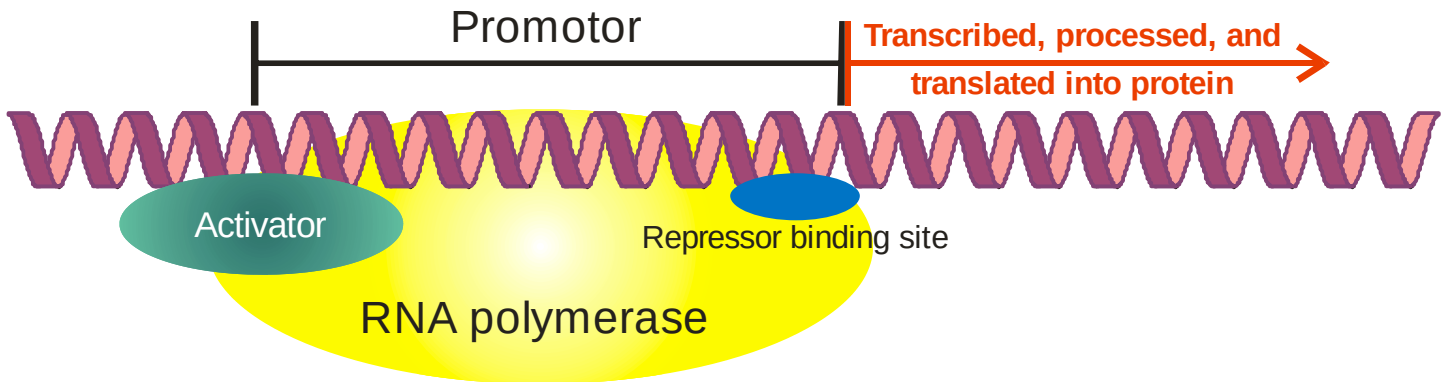
State I:

basal state



State II:

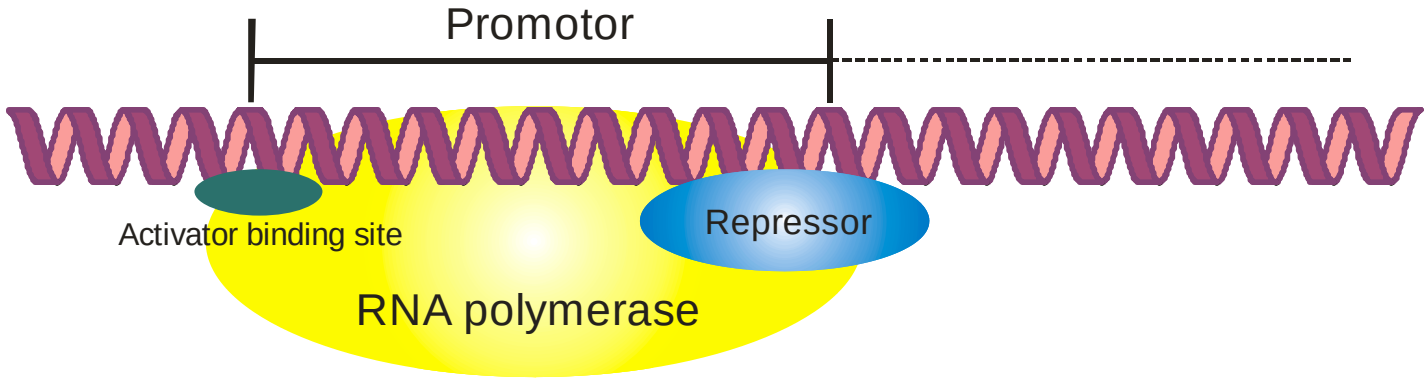
active state



Active states of gene regulation

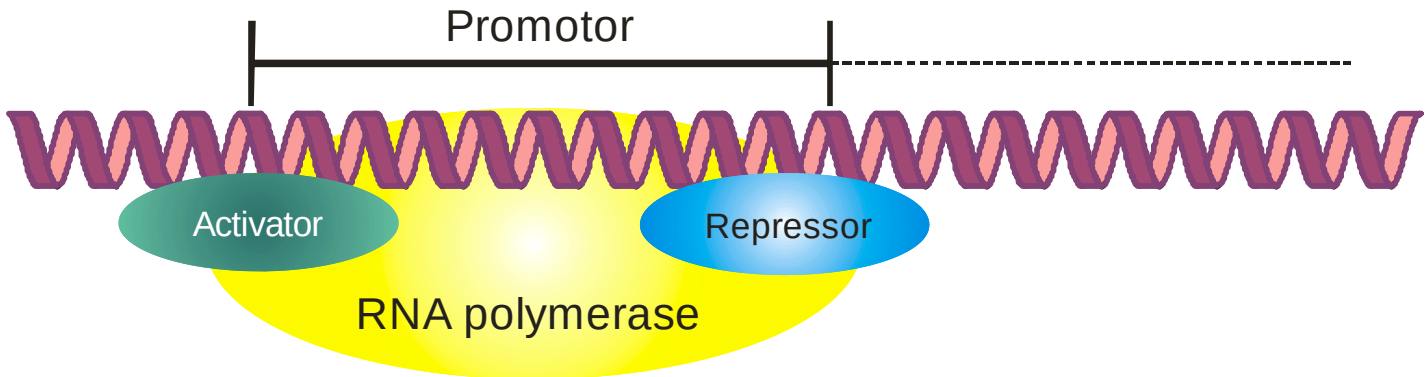
State III:

inactive state

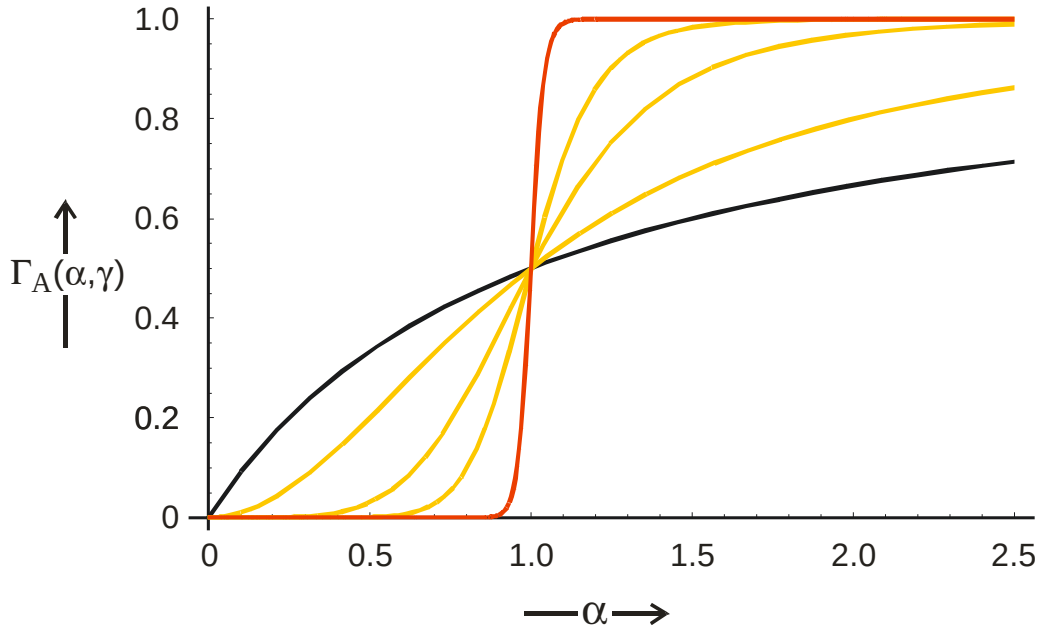


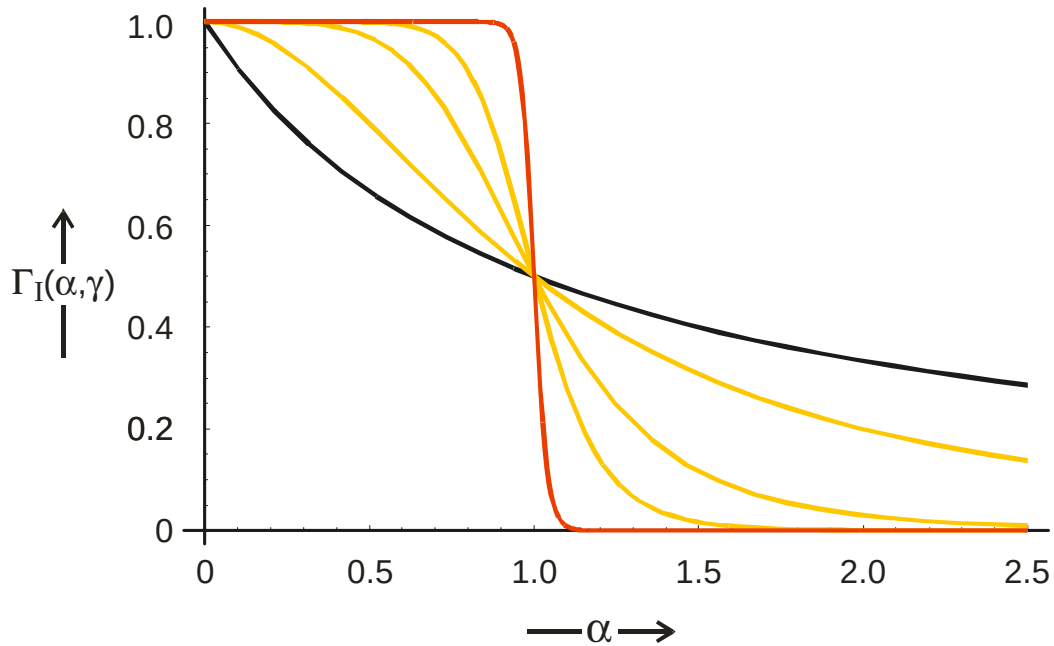
State III:

inactive state

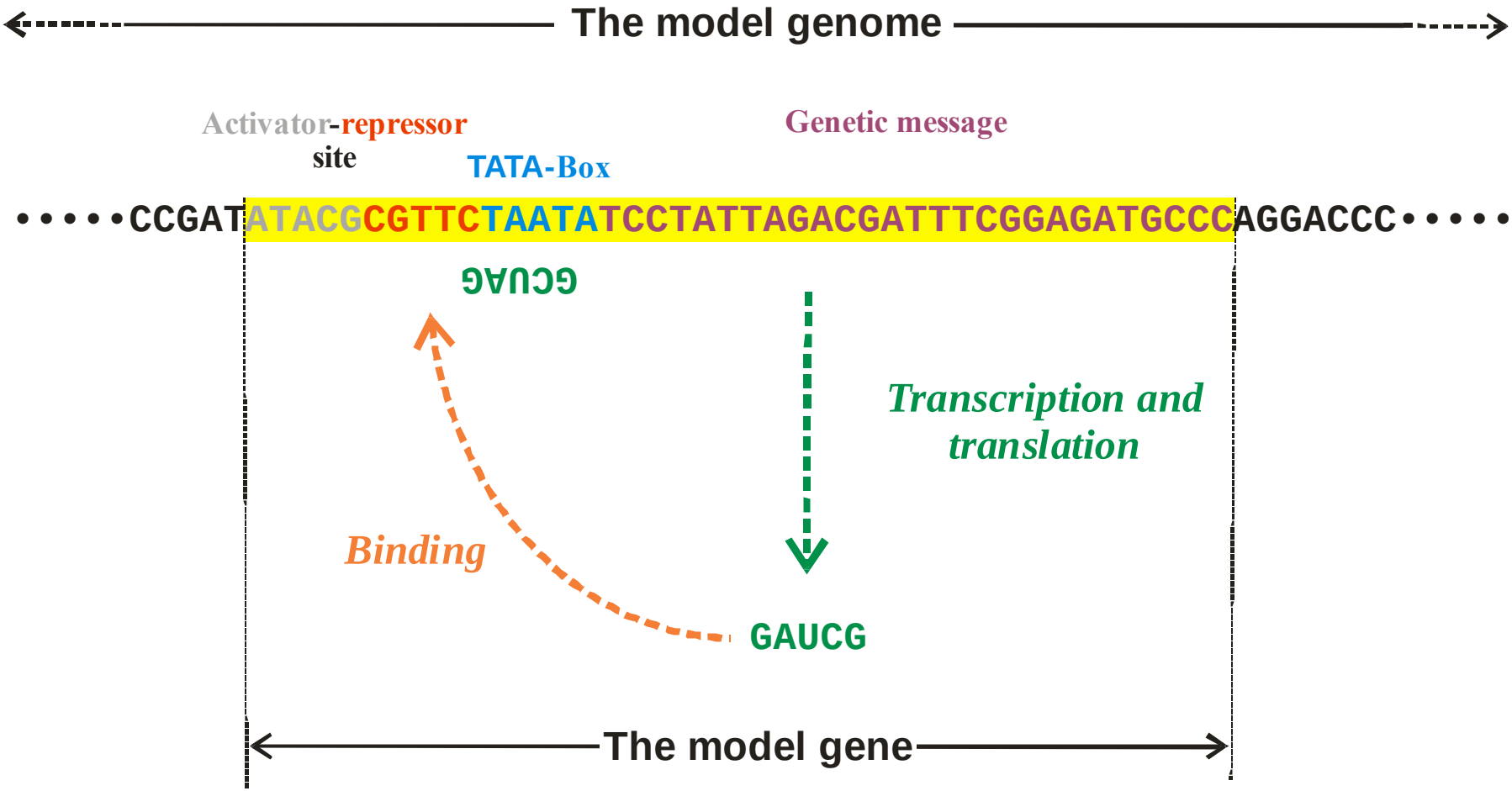


Inactive states of gene regulation

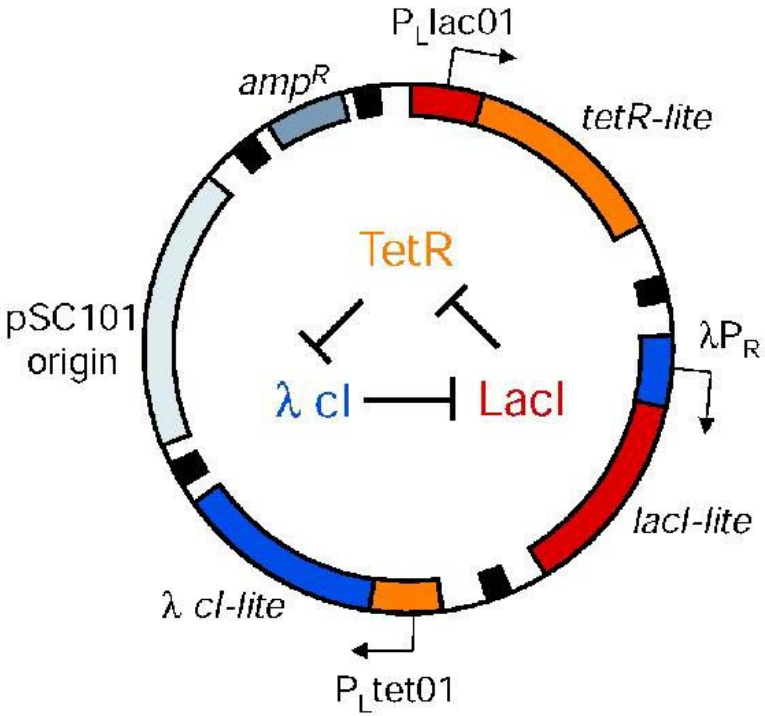


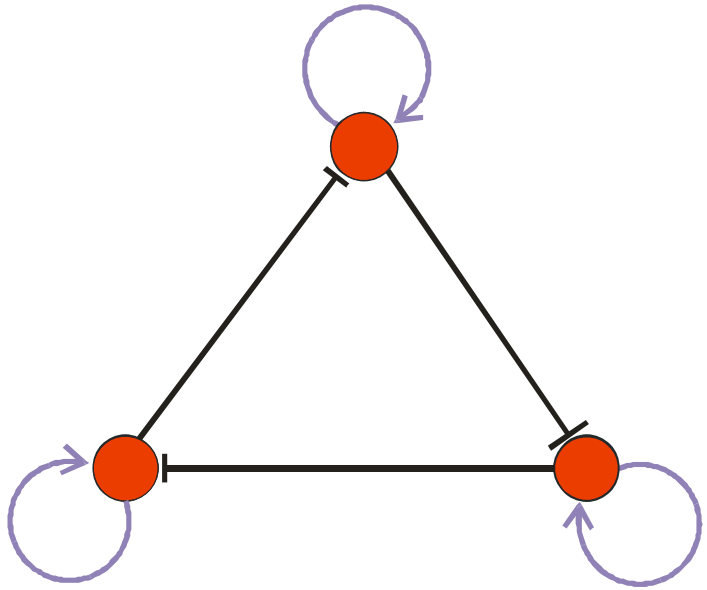


Gene activity for cooperative binding of activator and inhibitor

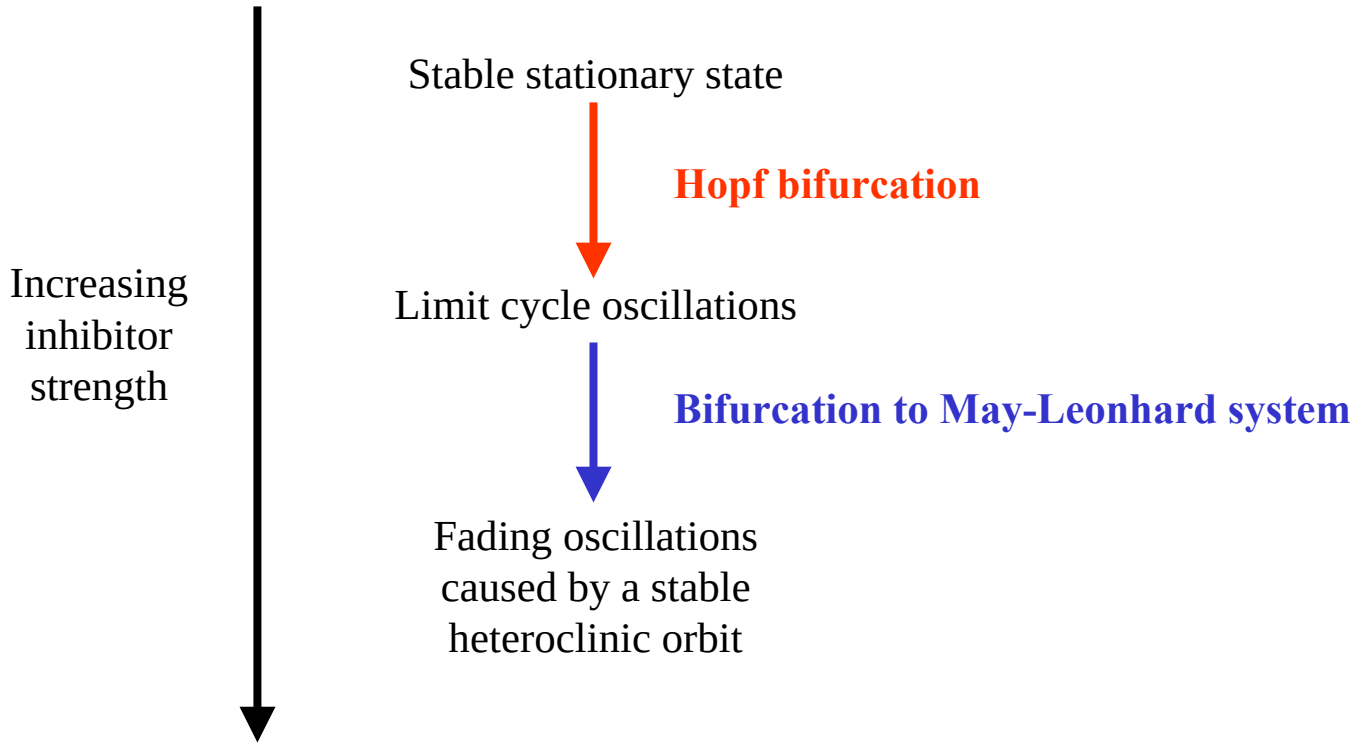


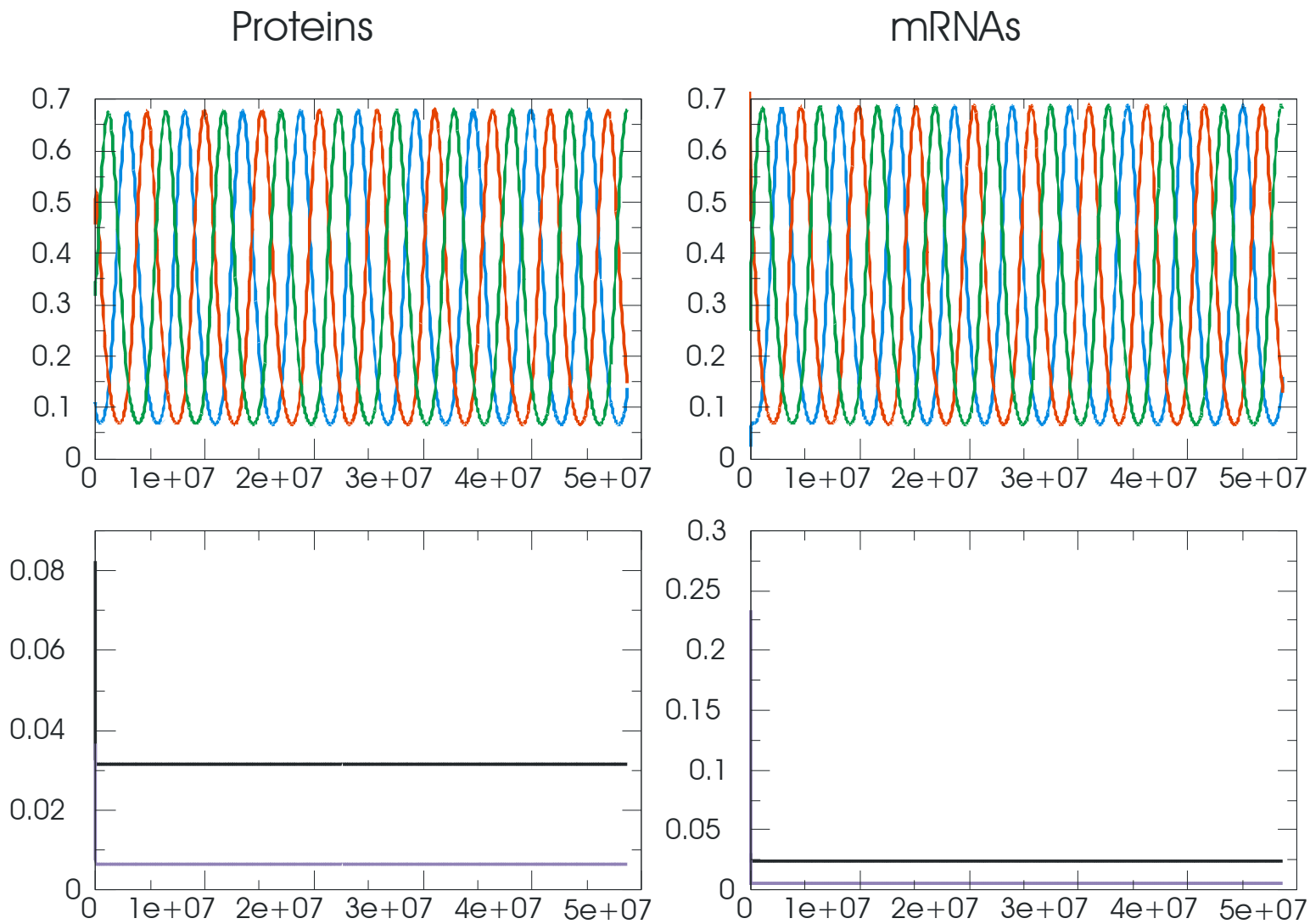
The model gene in MiniCellSim



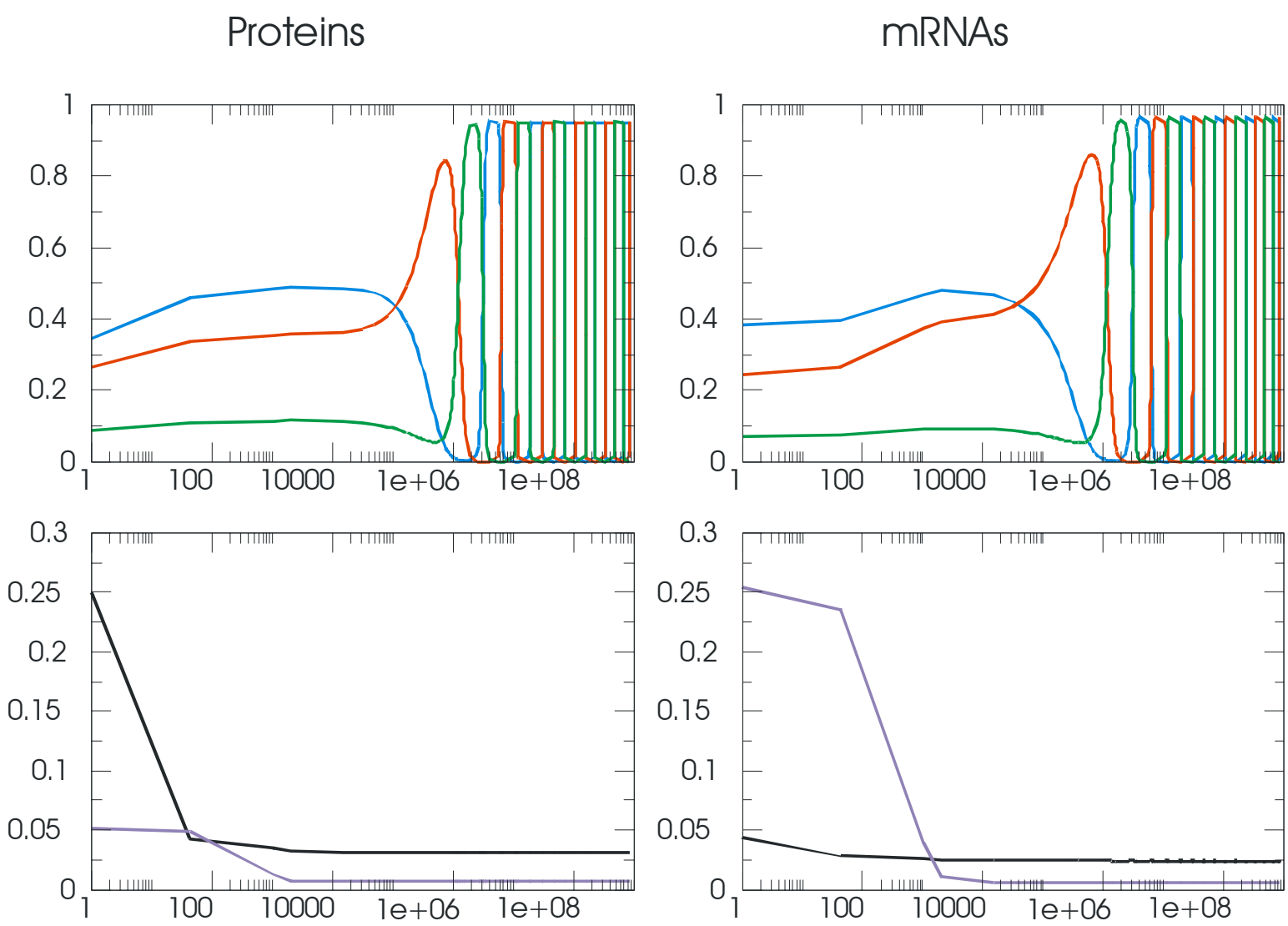


The repressilator: M.B. Ellowitz, S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**:335-338, 2002

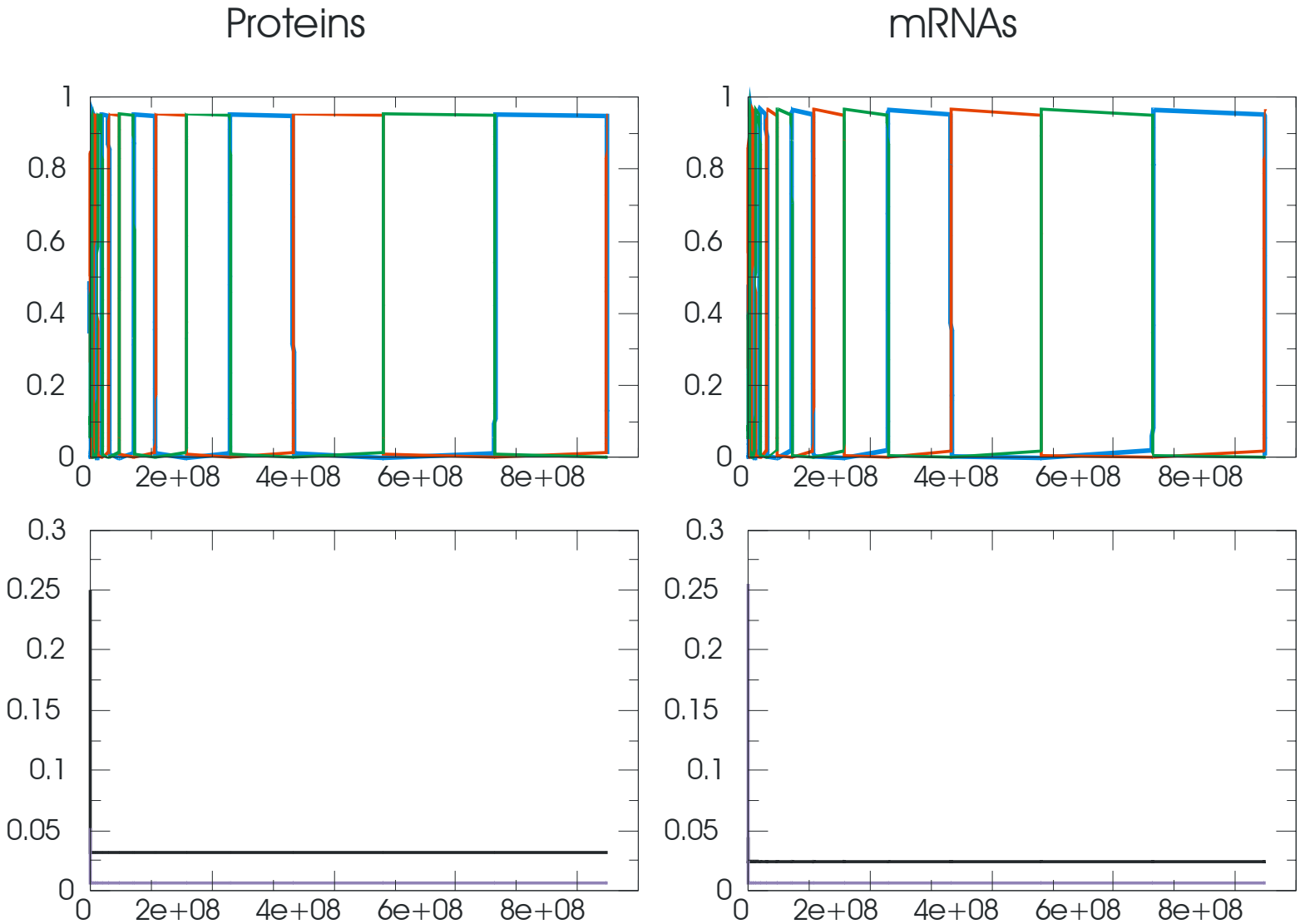

Increasing inhibitor strength
Stable stationary state
Hopf bifurcation
Limit cycle oscillations
Bifurcation to May-Leonhard system
Fading oscillations caused by a stable heteroclinic orbit



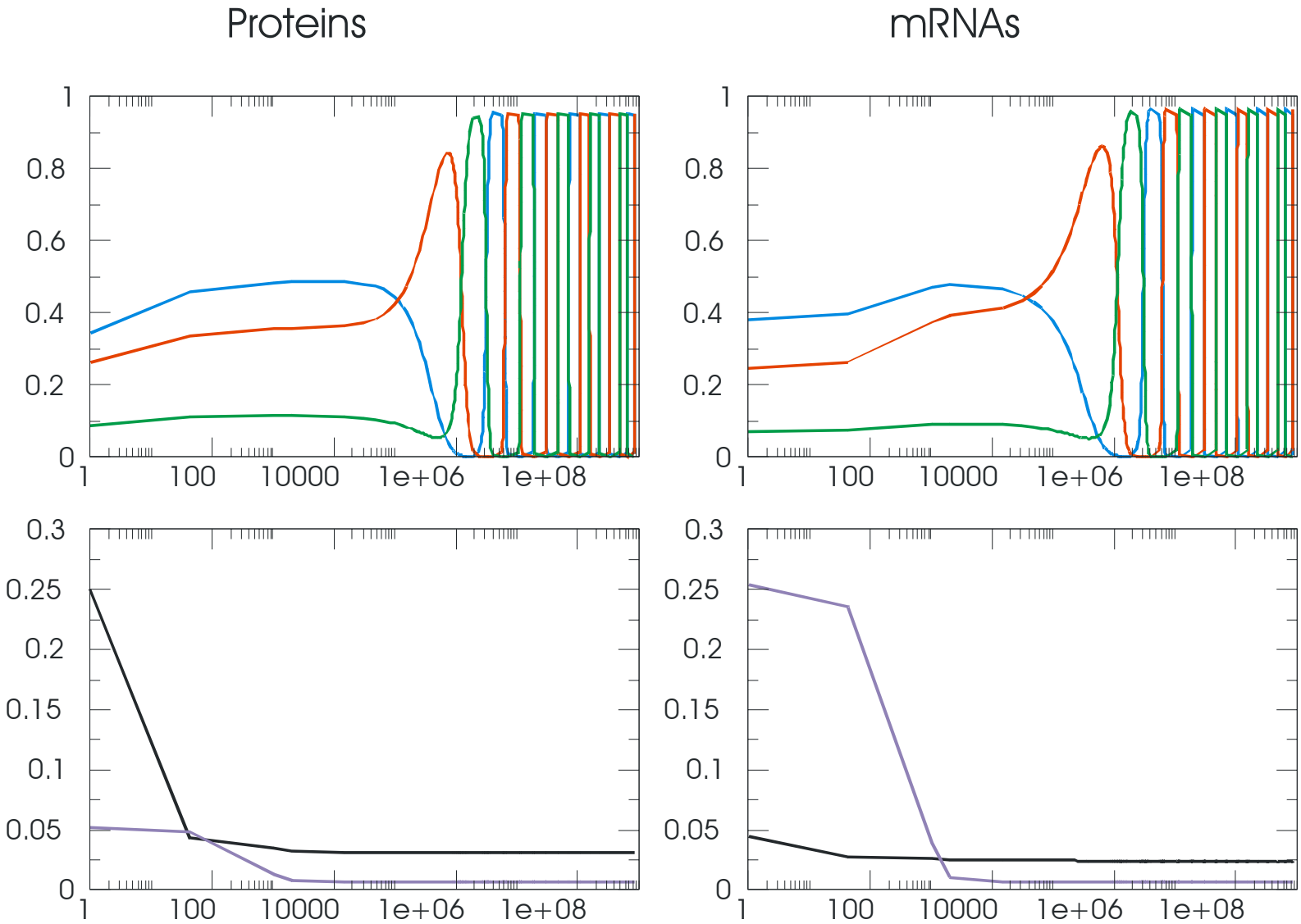
The repressilator limit cycle



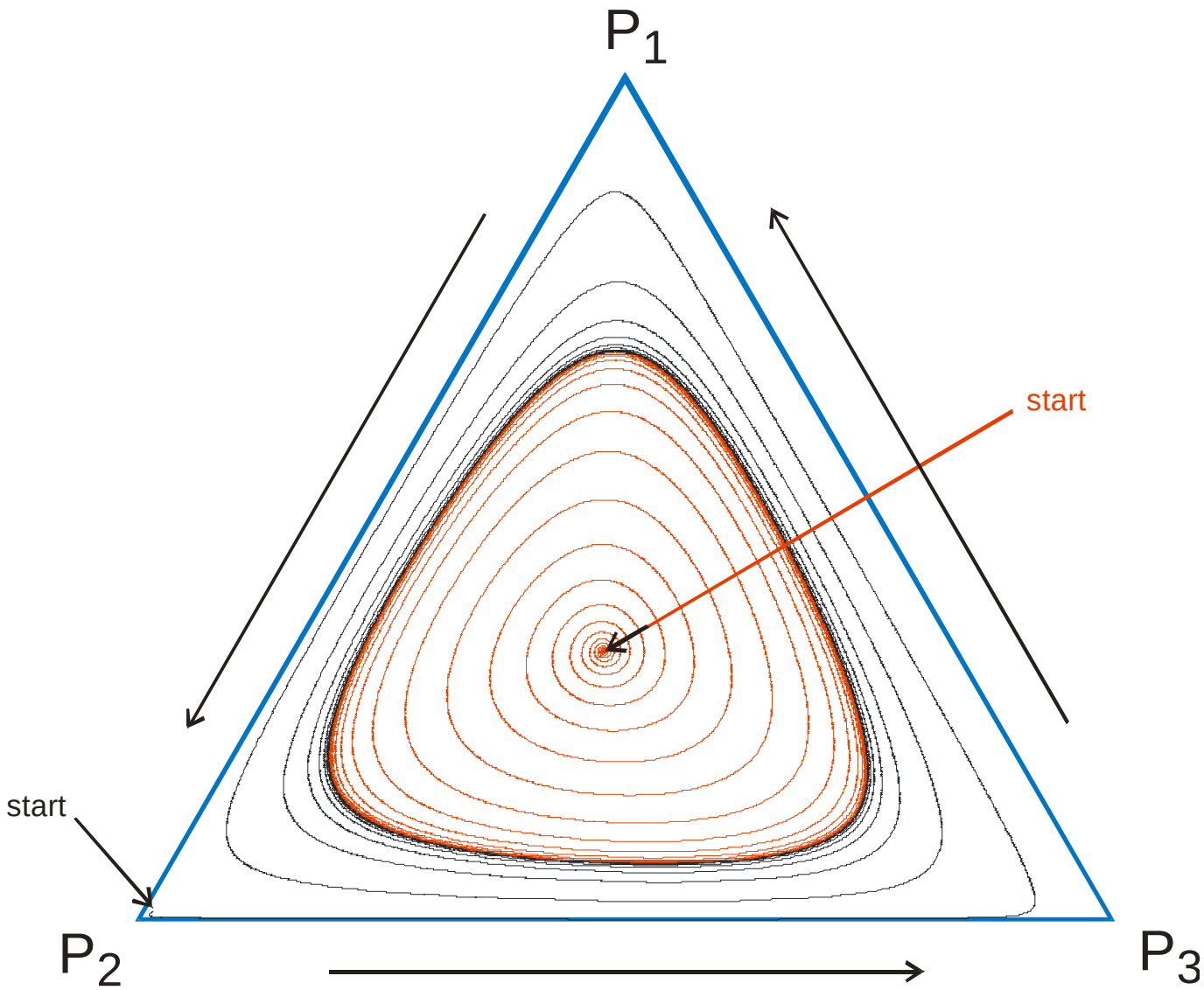
The repressilator limit cycle (logarithmic time scale)



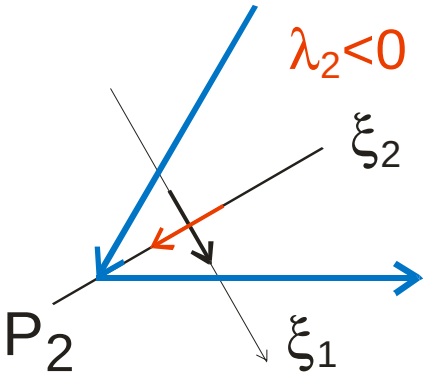
The repressilator heteroclinic orbit



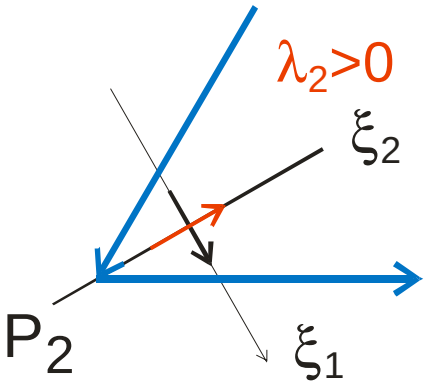
The repressilator heteroclinic orbit (logarithmic time scale)



The repressilator limit cycle

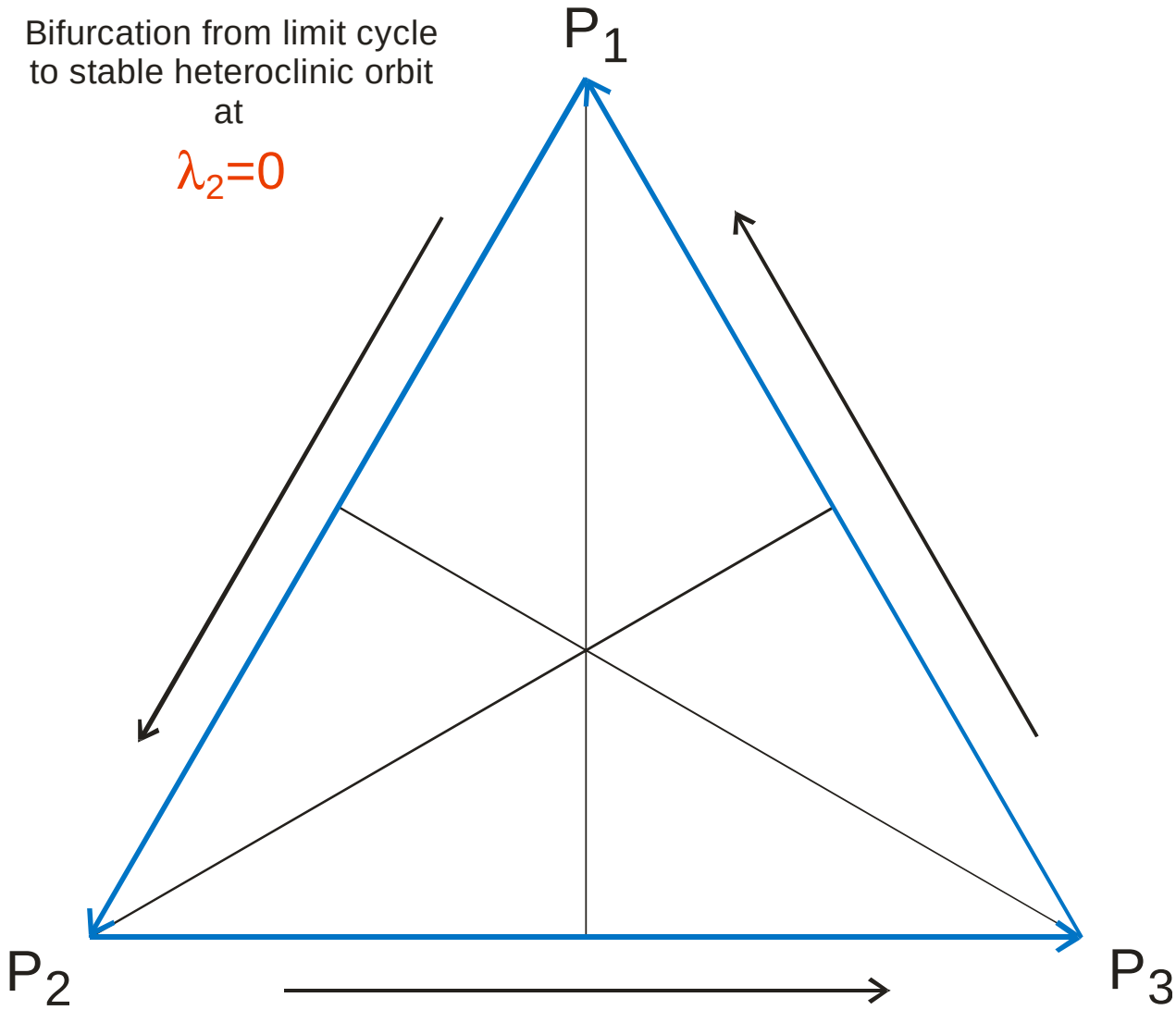


Stable heteroclinic orbit



Unstable heteroclinic orbit

Bifurcation from limit cycle to stable heteroclinic orbit at $\lambda_2=0$



The repressilator heteroclinic orbit

1. What is computational systems biology?
2. Networks and network evolution
3. Forward and inverse problems
4. Reverse engineering – A simple example
5. MiniCellSim – A simulation tool
6. **Evolution of genetic and metabolic networks**

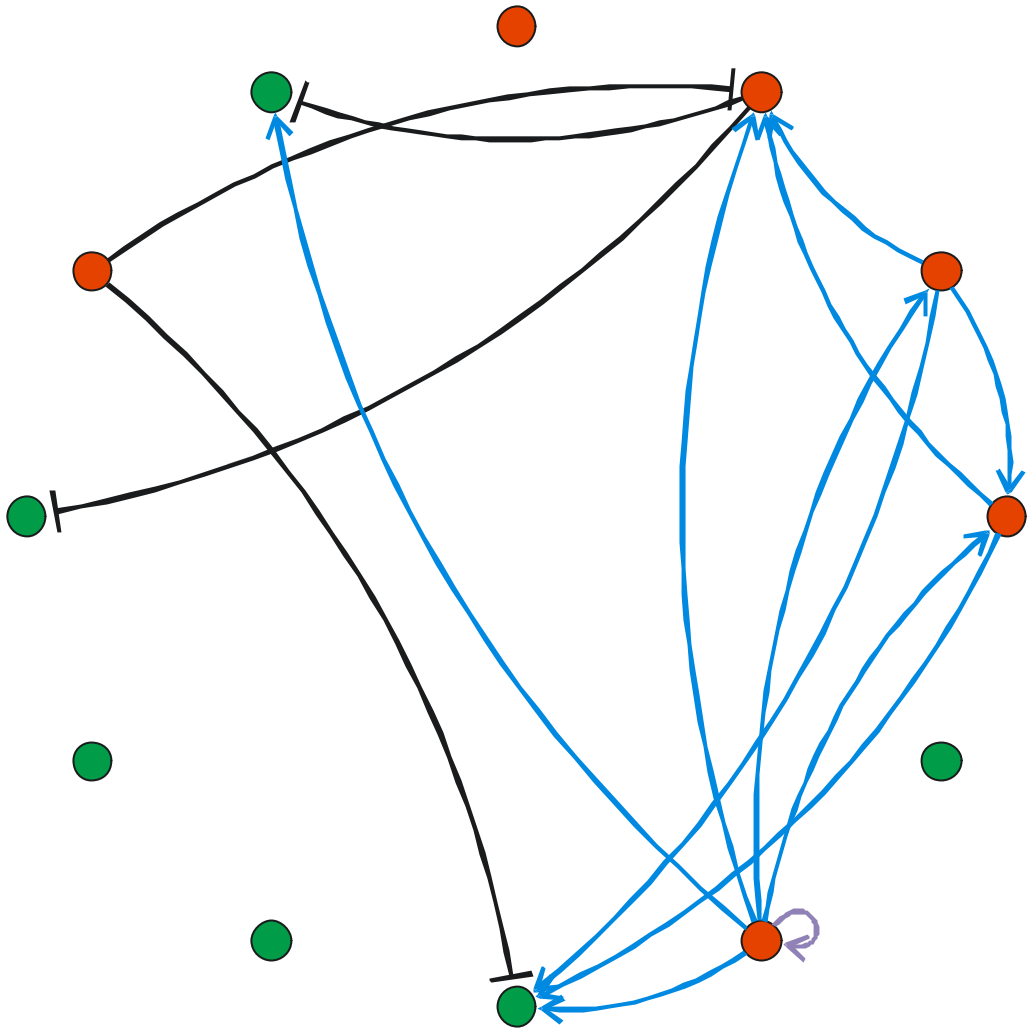
Evolutionary time: 0000

Number of genes: **12**

06 structural + 06 regulatory

Number of interactions: **15**

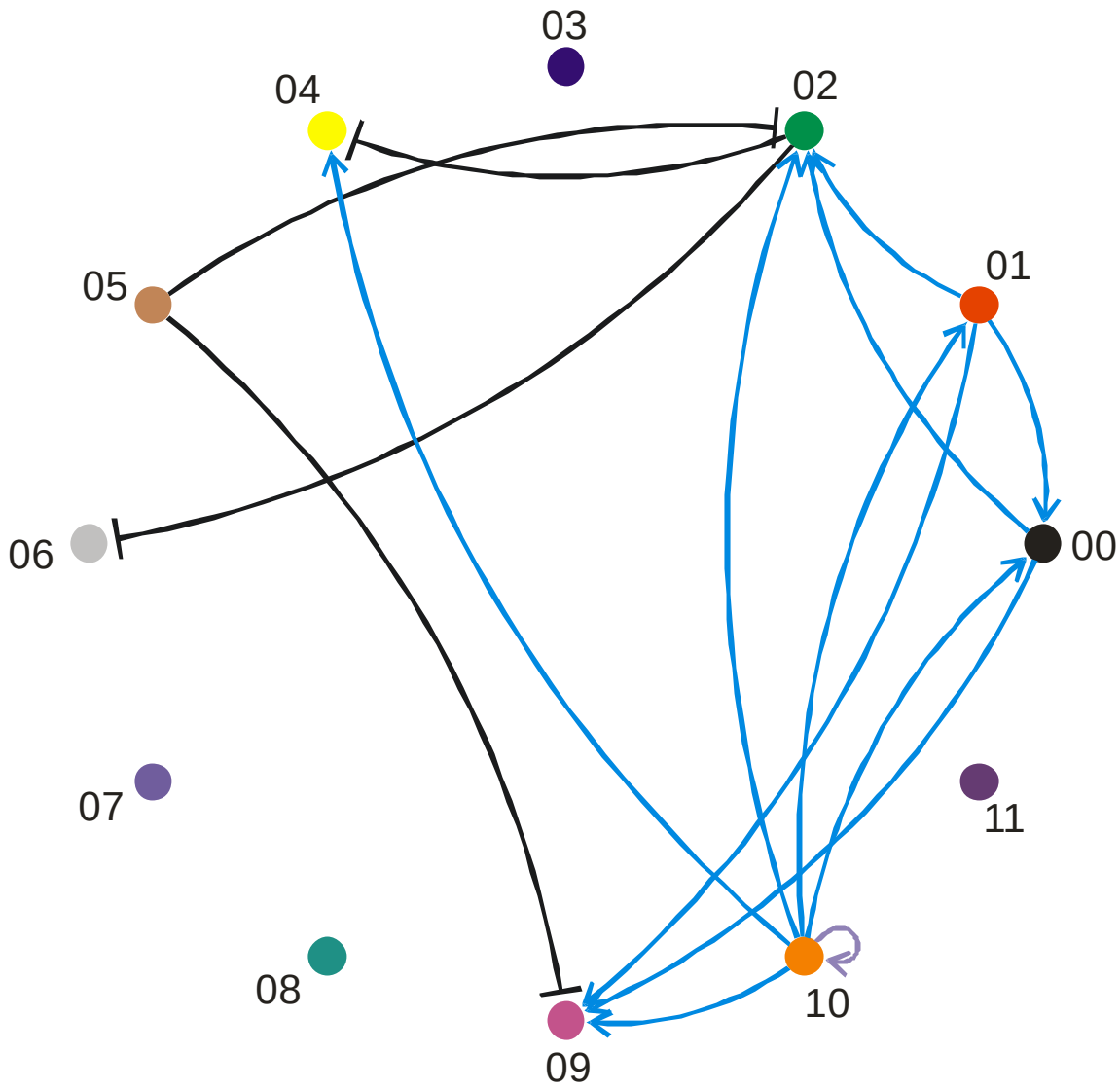
04 inhibitory + 10 activating +
+ 1 self-activating

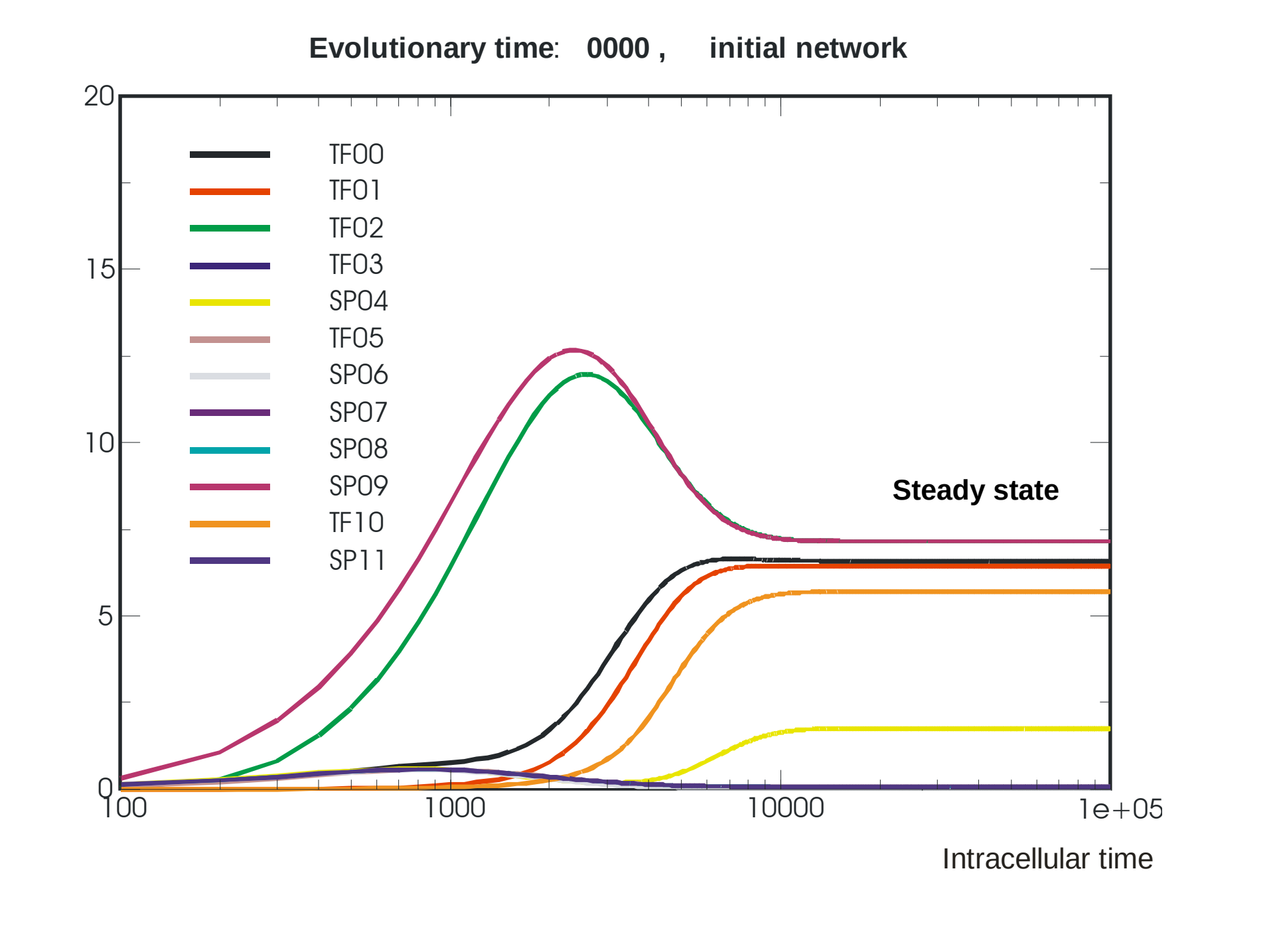


Network formed by a random sequence of 200 nucleotides

Numbering and color code of genes

TF00
TF01
TF02
TF03
SP04
TF05
SP06
SP07
SP08
SP09
TF10
SP11



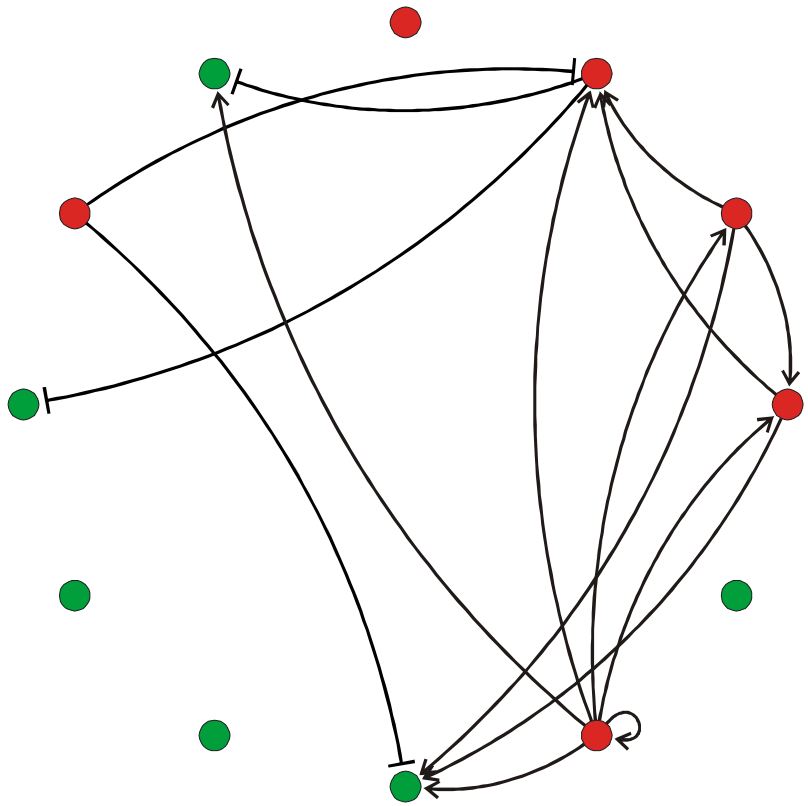

Evolutionary time: 0000 , initial network
20
15
10
5
0
100
1000
10000
1e+05
Intracellular time
TF00
TF01
TF02
TF03
SP04
TF05
SP06
SP07
SP08
SP09
TF10
SP11
Steady state

Evolution of a genetic and metabolic network:

Initial genome: Random sequence of chain length **n = 200**,
AUGC alphabet

Simulation with a mutation rate: **p = 0.01**

Evolutionary time unit >> time unit of regulatory kinetics



Observed events:

(i) Loss of a gene through corruption of the initiation signal "**TA**" (analogue of the **TATA** box)

(ii) Creation of a gene

(iii) Change in the connections through mutation driven changes in the binding affinities of translation factors to the regulatory sites

(iv) Genes may change their class (tf ⇔sp)

Windows Media Player.lnk

Conclusion and outlook on inverse problems

☑	1.	RNA minimum free energy folding and inverse folding for the design of secondary structures.
✓ in progress	2.	Kinetic folding of RNA and design of molecules with multiple states and predefined folding kinetics.
✓ in progress	3.	Computation of the dynamics of cellular genetic and metabolic networks for known rate constants and its inverse problem (Level I).
in progress ✓	4.	Genetic and metabolic dynamics in parameter space and reverse engineering of model systems with predefined full dynamical behavior (Level II) seems doable. Mathematical tools can be applied successfully also to multidimensional dynamical systems.
☑	5.	Random sequences give rise to functional networks in the model.
✓ in progress	6.	Evolution of small genetic and metabolic networks can be simulated properly and with reasonable efforts.
	7.	Upscaling still remains a hard but challenging problem.

Acknowledgement of support

Universität Wien

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)

Projects No. 09942, 10578, 11065, 13093
13887, and 14898

Jubiläumsfonds der Österreichischen Nationalbank

Project No. Nat-7813

European Commission: Project No. EU-980189

Austrian Genome Research Program – GEN-AU

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Coworkers

Universität Wien

Walter Fontana, Harvard Medical School, MA

Christian Reidys, **Christian Forst**, Los Alamos National Laboratory, NM

Peter Stadler, **Bärbel Stadler**, Universität Leipzig, GE

Josef Hofbauer, University College, London, UK

Robert Giegerich, Universität Bielefeld, GE

Heinz Engl, **Stefan Müller**, Johann Radon-Institut für Angewandte und Computergestützte Mathematik der Österreichischen Akademie der Wissenschaften, Linz, AT

Christoph Flamm, **Ivo L.Hofacker**,
Andreas Svrček-Seiler, Universität Wien, AT

Stefanie Widder, **Lukas Endler**, Universität Wien, AT
Ulrike Langhammer, **Ulrike Mückstein**, **Stefan Bernhart**

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks