Systems biology and complexity research

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Interdisciplinary Challenges for Complexity Sciences

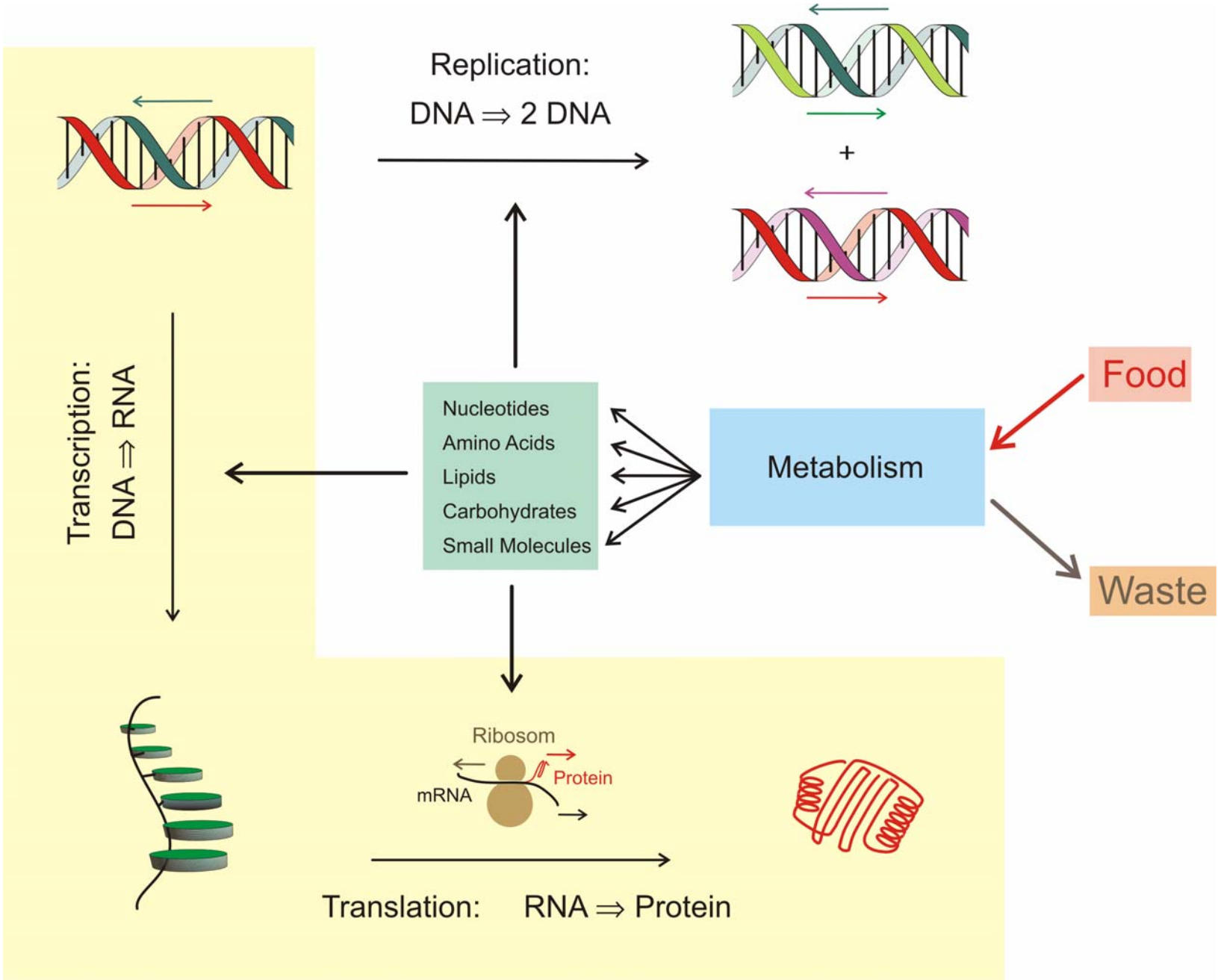
Brussels, 27.– 28.05.2009

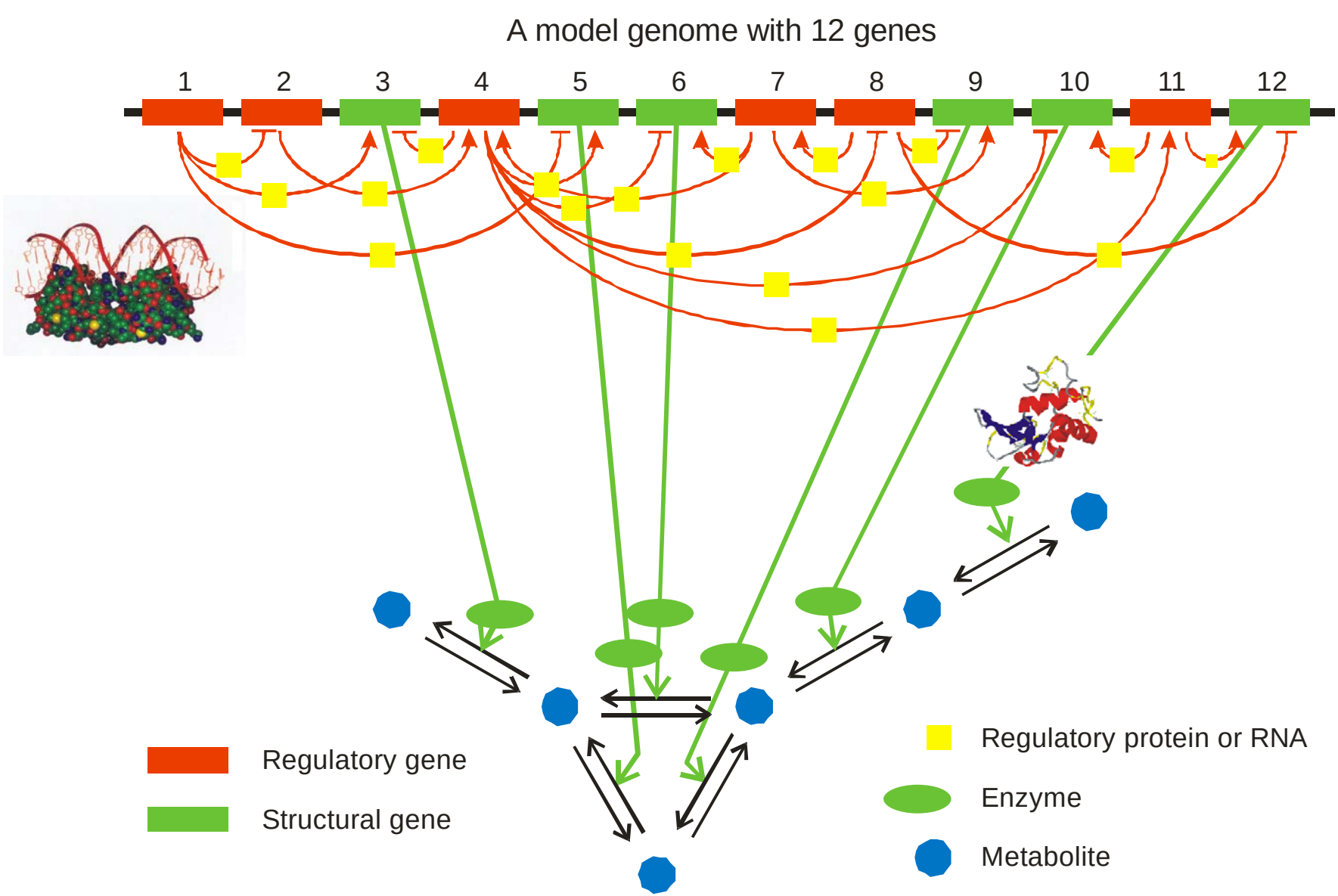
Web-Page for further information:

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

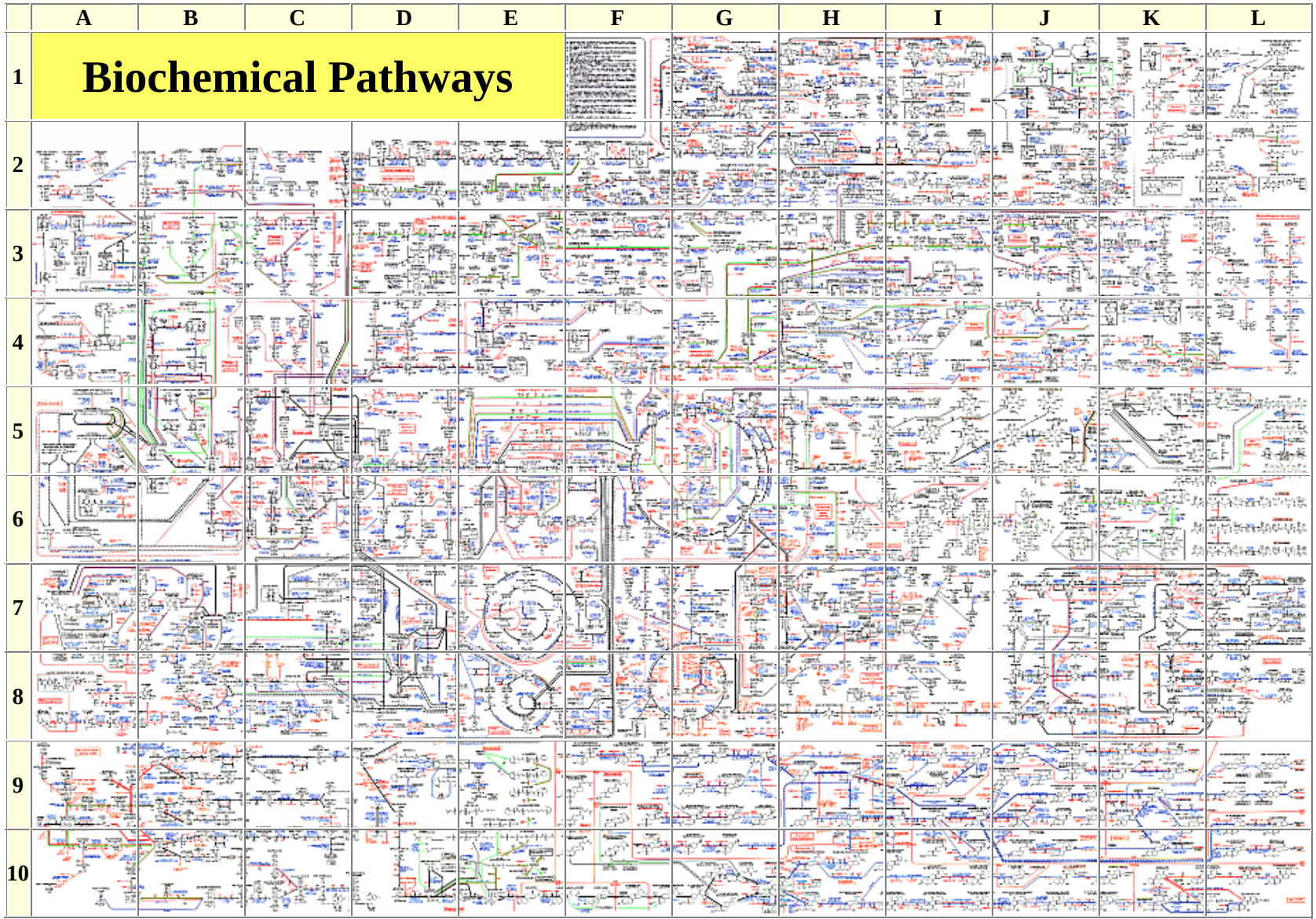
1. Complex networks in cellular regulation
2. Experimental data and modeling in biology
3. Parameter determination and reverse engineering
4. Gene regulation dynamics
5. Inverse bifurcation analysis
6. Current challenges in biology

1. **Complex networks in cellular regulation**
2. Experimental data and modeling in biology
3. Parameter determination and reverse engineering
4. Gene regulation dynamics
5. Inverse bifurcation analysis
6. Current challenges in biology


Replication:
DNA ⇒ 2 DNA
+
Transcription:
DNA ⇒ RNA
Nucleotides
Amino Acids
Lipids
Carbohydrates
Small Molecules
Metabolism
Food
Waste
Ribosom
Protein
mRNA
Translation: RNA ⇒ Protein

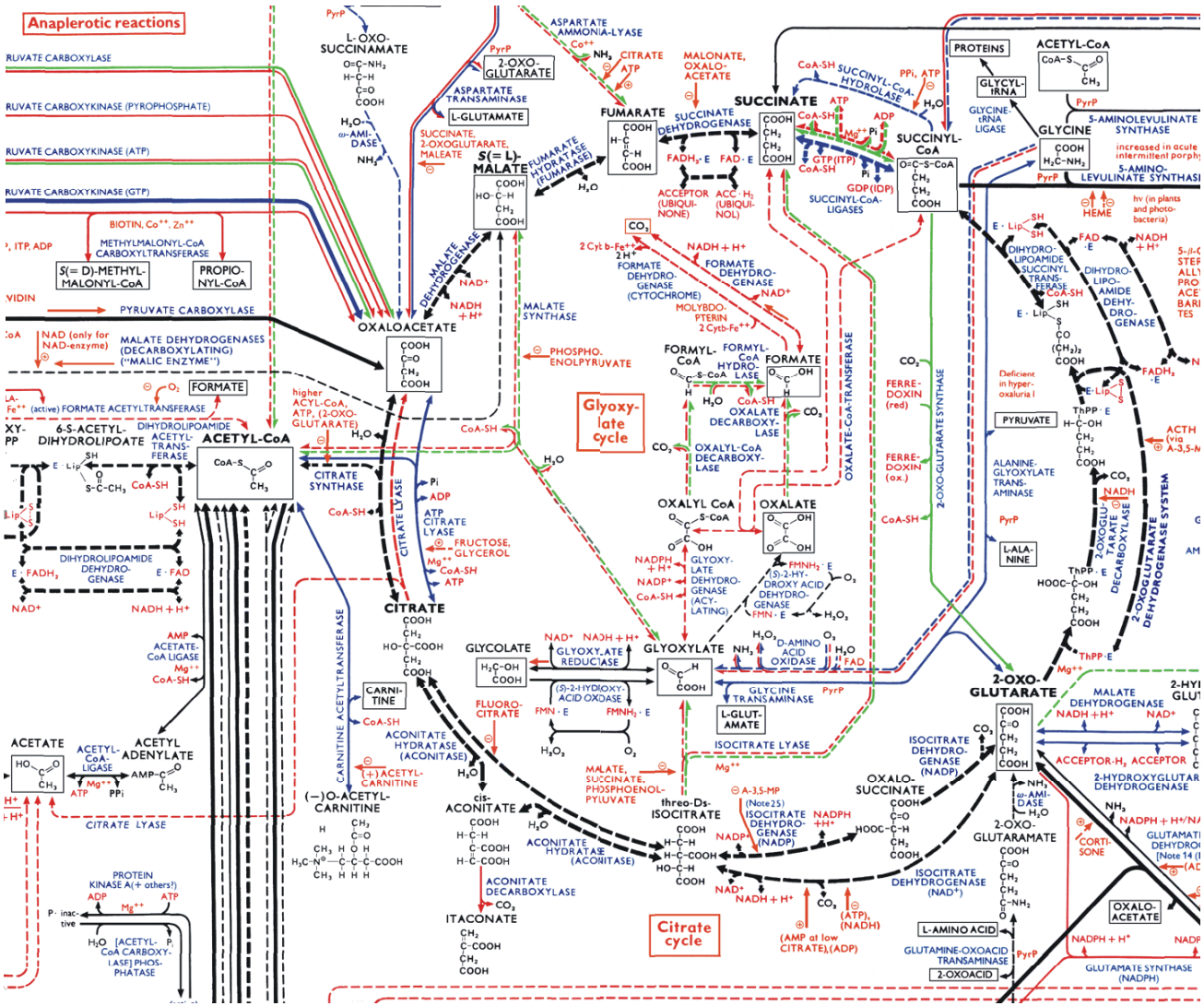


Sketch of a genetic and metabolic network



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

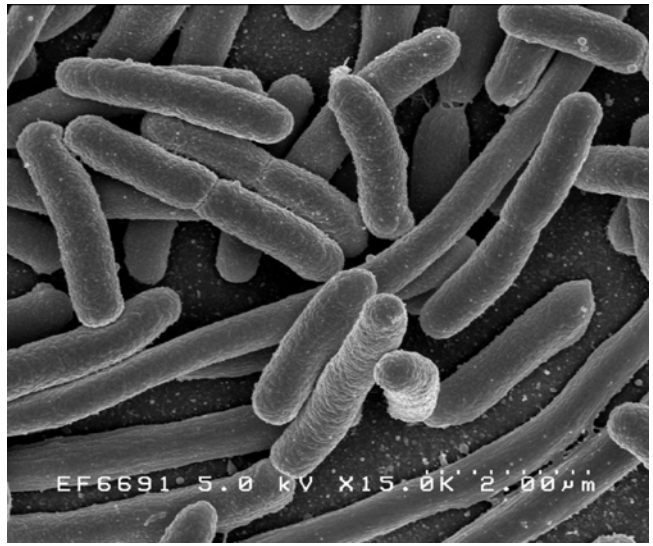
The citric acid or Krebs cycle (enlarged from previous slide).



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

E. coli:	Genome length	4×10^6 nucleotides
	Number of cell types	1
	Number of genes	4 460

Four books, 300 pages each



Man:	Genome length	3×10^9 nucleotides
	Number of cell types	200
	Number of genes	≈ 30 000

A library of 3000 volumes,
300 pages each

Complexity in biology

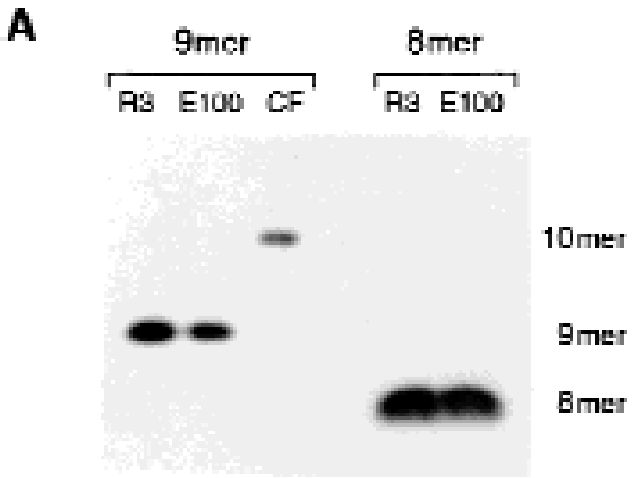
1. Complex networks in cellular regulation
2. **Experimental data and modeling in biology**
3. Parameter determination and reverse engineering
4. Gene regulation dynamics
5. Inverse bifurcation analysis
6. Current challenges in biology

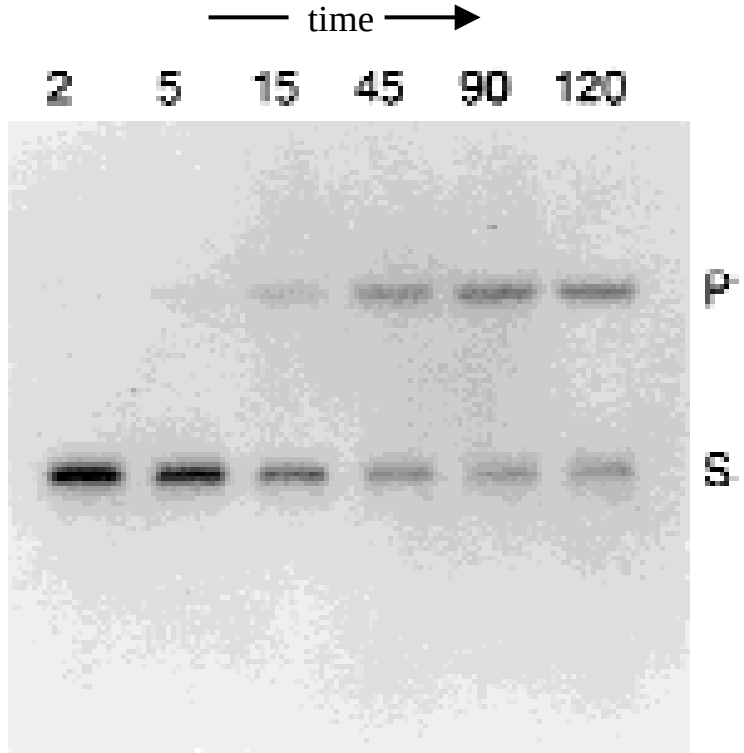
From qualitative data to quantitative modeling

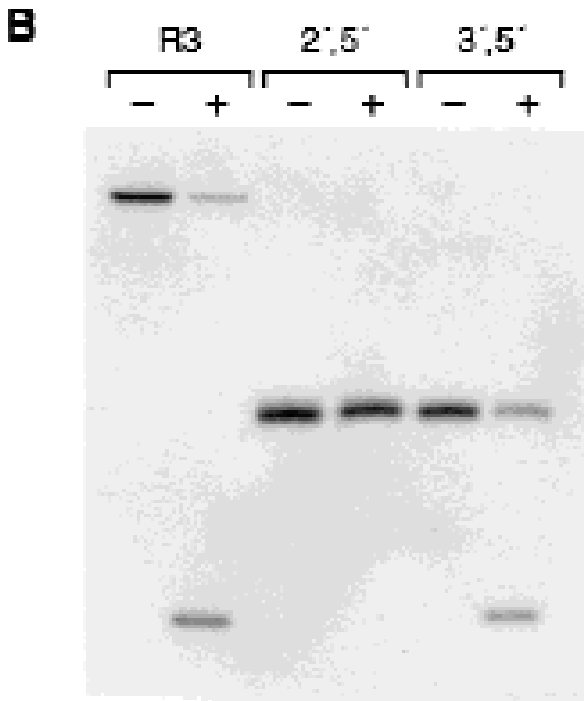
Genomics, transcriptomics, proteomics

Metabolomics, functional genomics

Computational systems biology

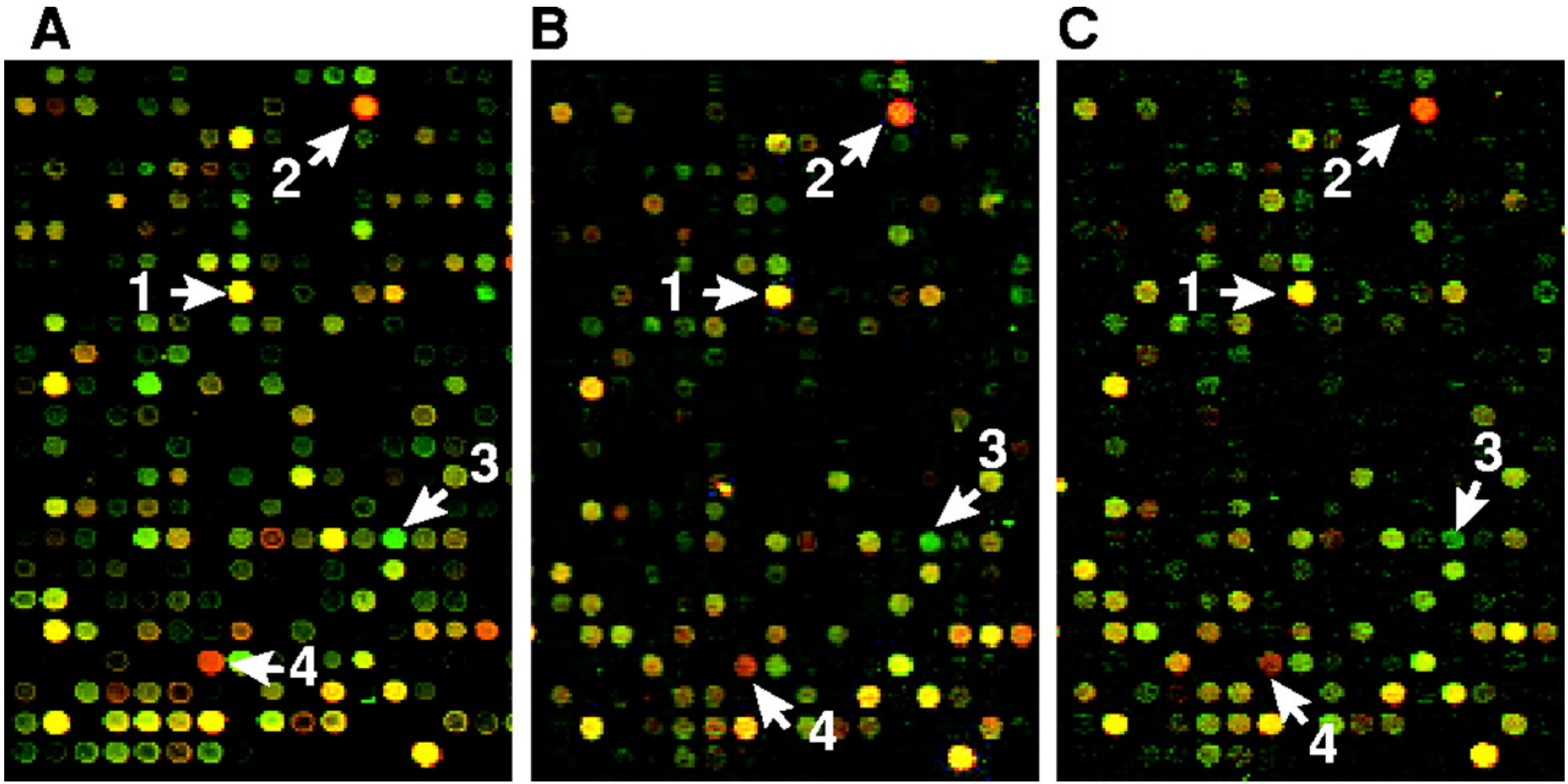






Analysis by gel electrophoresis

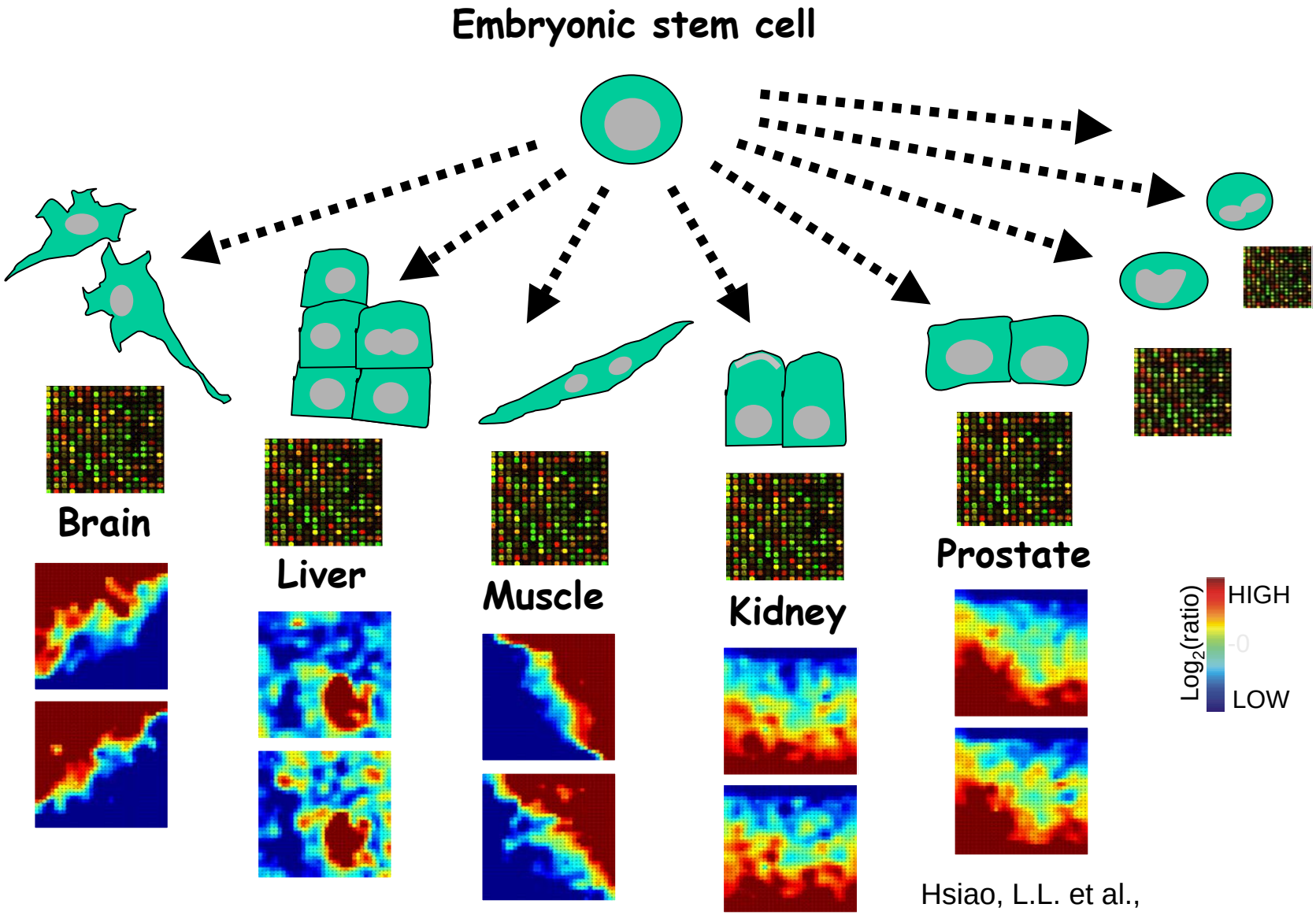
Jeff Rogers, Gerald F. Joyce.
RNA **7**:395-404, 2001



The same section of the microarray is shown in three independent hybridizations. Marked spots refer to: (1) protein disulfide isomerase related protein P5, (2) IL-8 precursor, (3) EST AA057170, and (4) vascular endothelial growth factor.

Gene expression DNA microarray representing **8613** human genes used to study transcription in the response of human fibroblasts to serum.

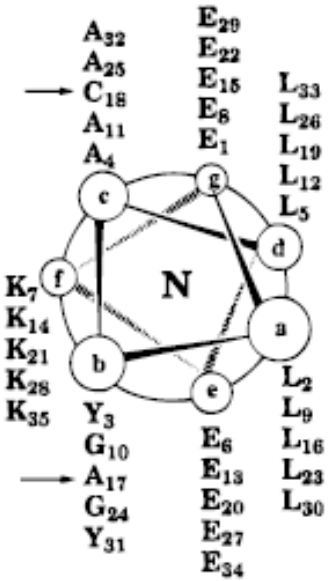
V.R.Iyer *et al.*, *Science* **283**: 83-87, 1999



Hsiao, L.L. et al.,
Physiol.Genomics 2001
Affymetrix, ~ 7000 genes

SOM-based "GEDI maps"
(Eichler, G.S. et al., *Bioinformatics* 2003)

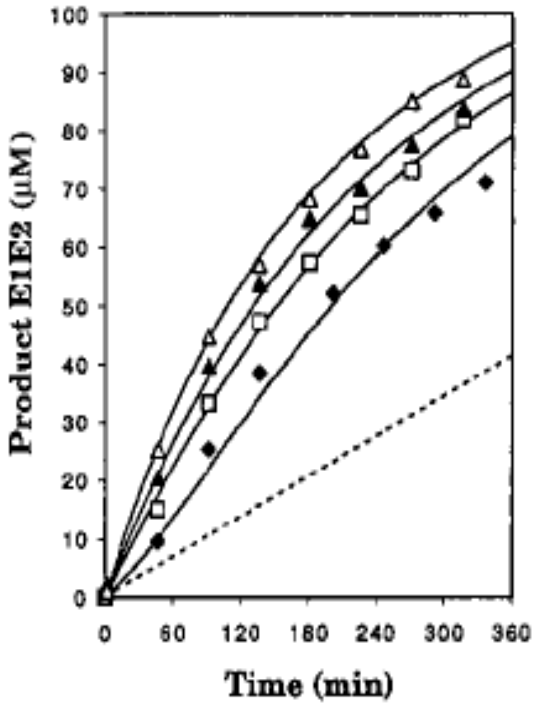
Drawings by Stuart A. Kauffman, 2009

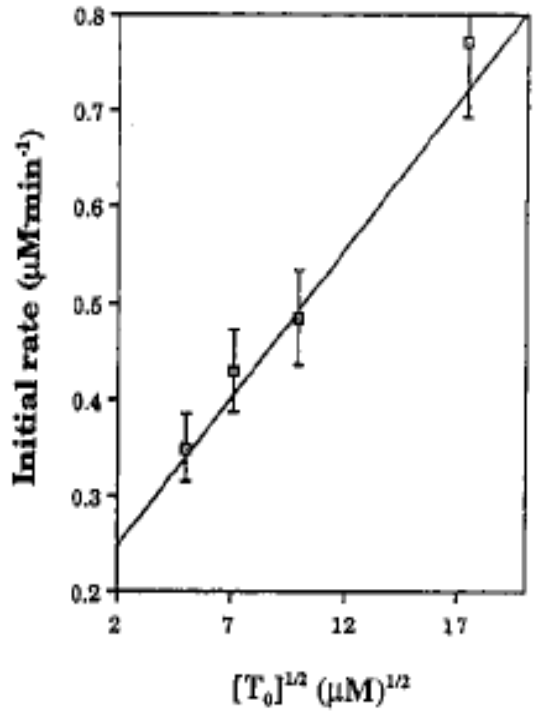


E1E2: Ac-ELYALEK-ELGALEK-ELACLEK-ELGALEK-ELYALEK-$CONH_2$

E1: Ac-ELYALEK-ELGALEK-ELA-COSR

E2: H_2N-CLEK-ELGALEK-ELYALEK-$CONH_2$





A pH-modulated, self-replicating peptide

Shao Yao, Indraneel Ghosh, Reena Zutshi, Jean Chmielewski.
J.Am Chem.Soc. **119**:10559-10560, 1997

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

$$\frac{da}{dt} = \frac{db}{dt} = -k_1 a b$$

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

$$\frac{dy}{dt} = k_2 x^2 - k_3 x y$$

$$\frac{dd}{dt} = k_3 x y$$

Stoichiometric equations

↓

SBML – systems biology markup language

↓

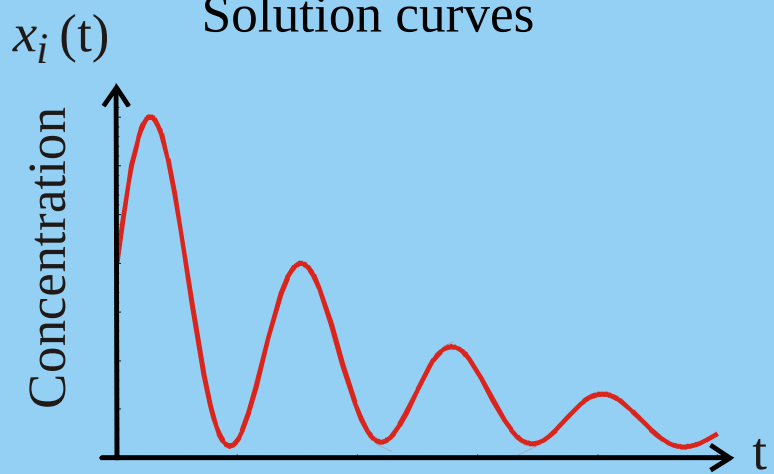
Kinetic differential equations

↓

ODE Integration

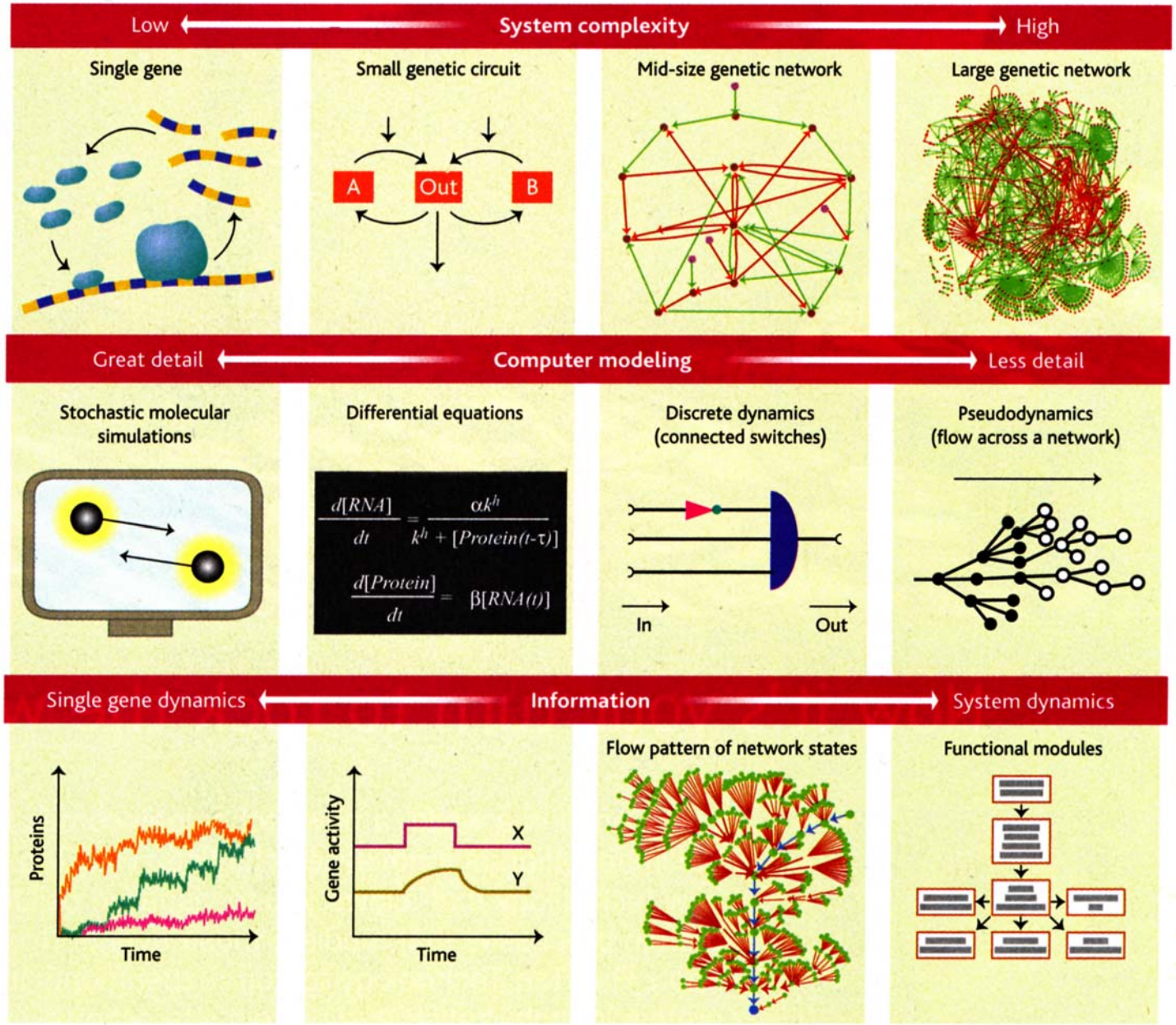
↓

Solution curves



The elements of the simulation tool MiniCellSim

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



Stefan Bornholdt. Less is more in modeling large genetic networks. ***Science*** **310**, 449-450 (2005)

1. Complex networks in cellular regulation
2. Experimental data and modeling in biology
3. **Parameter determination and reverse engineering**
4. Gene regulation dynamics
5. Inverse bifurcation analysis
6. Current challenges in biology

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: $x(0)$

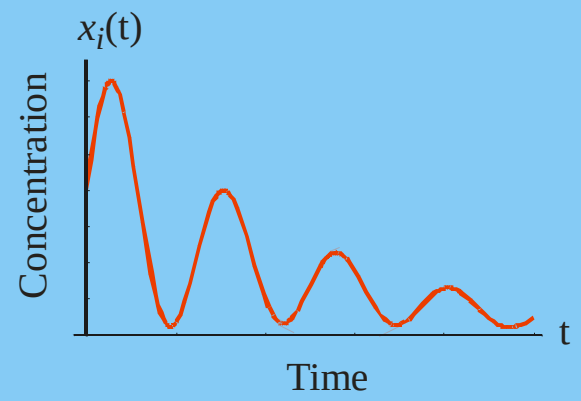
Boundary conditions:

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

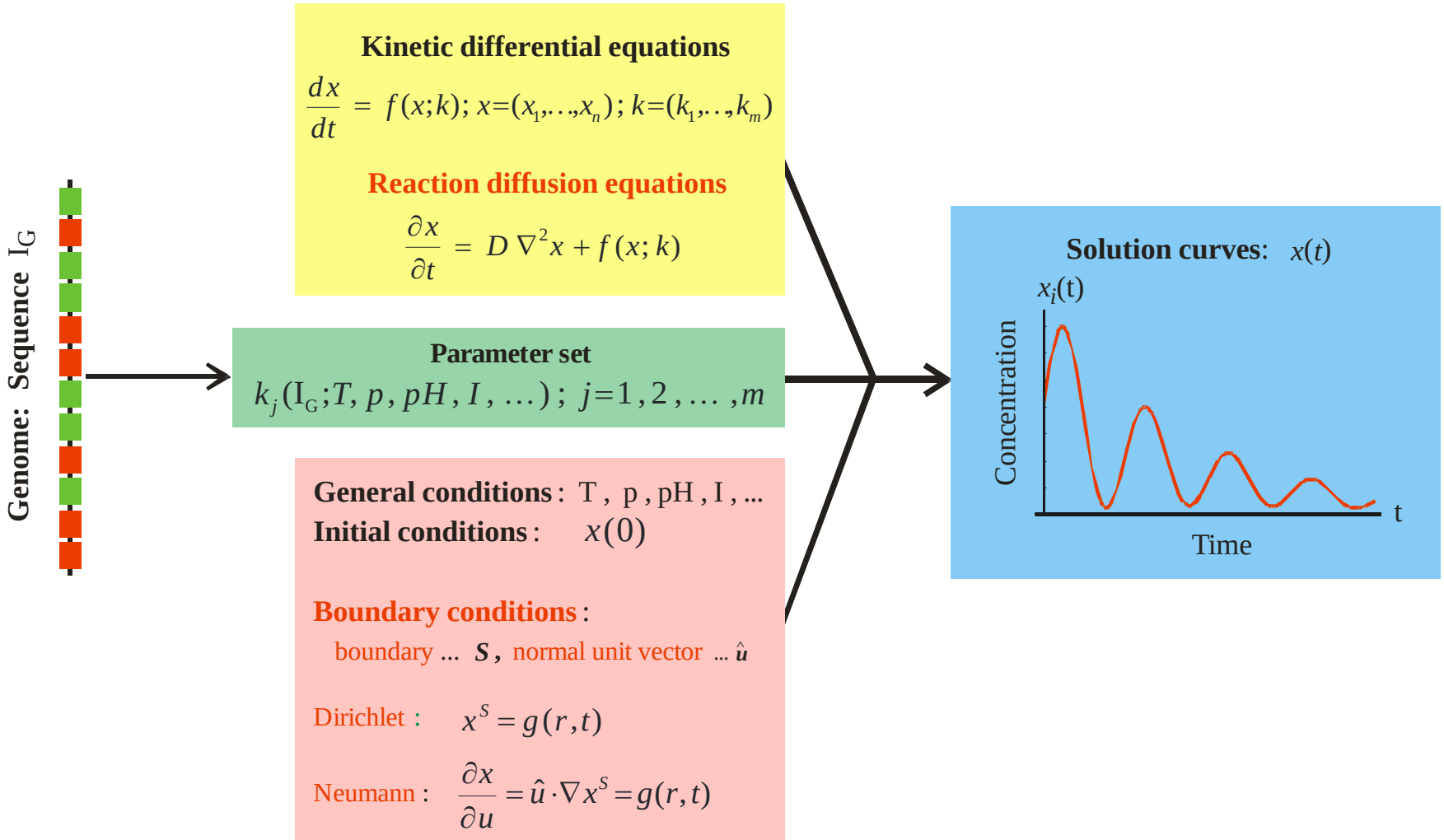
Dirichlet : $x^S = g(r,t)$

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 :

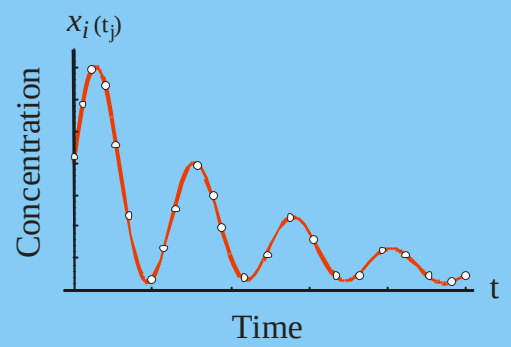
boundary ... S , normal unit vector... $\hat{u}$

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

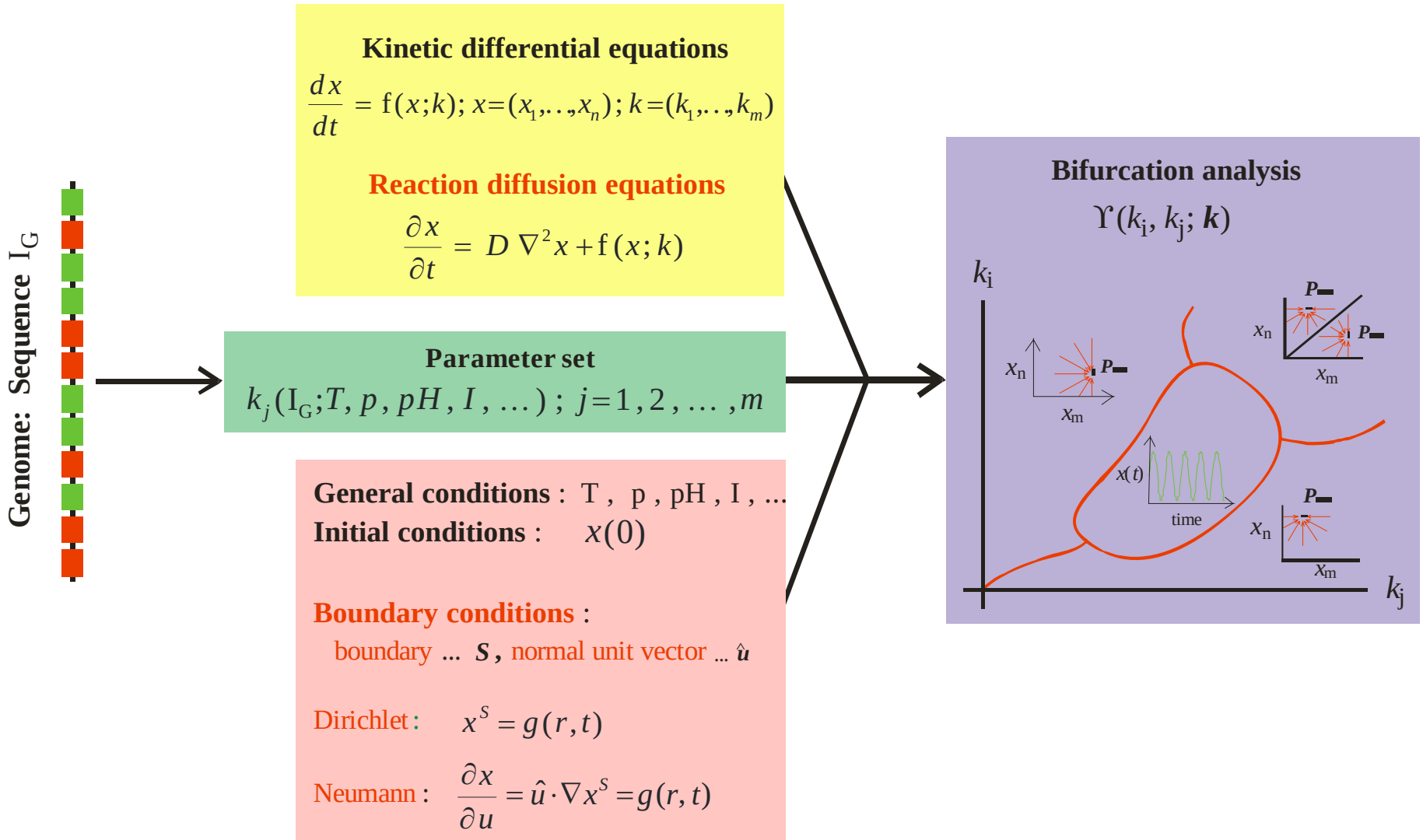
Neumann : $\frac{\partial x}{\partial u} = \hat{u} \cdot \nabla x^S = g(r,t)$

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:

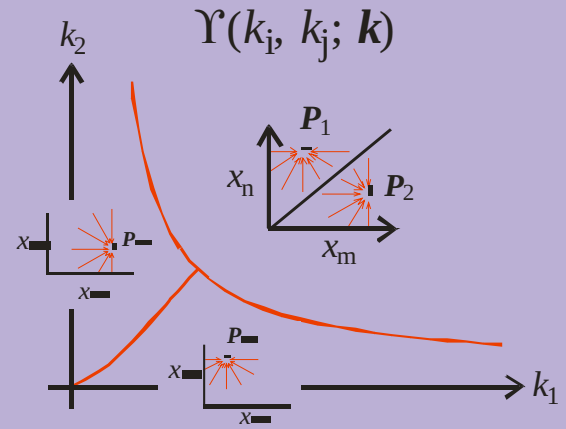
boundary ... S , normal unit vector... $\hat{u}$

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

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

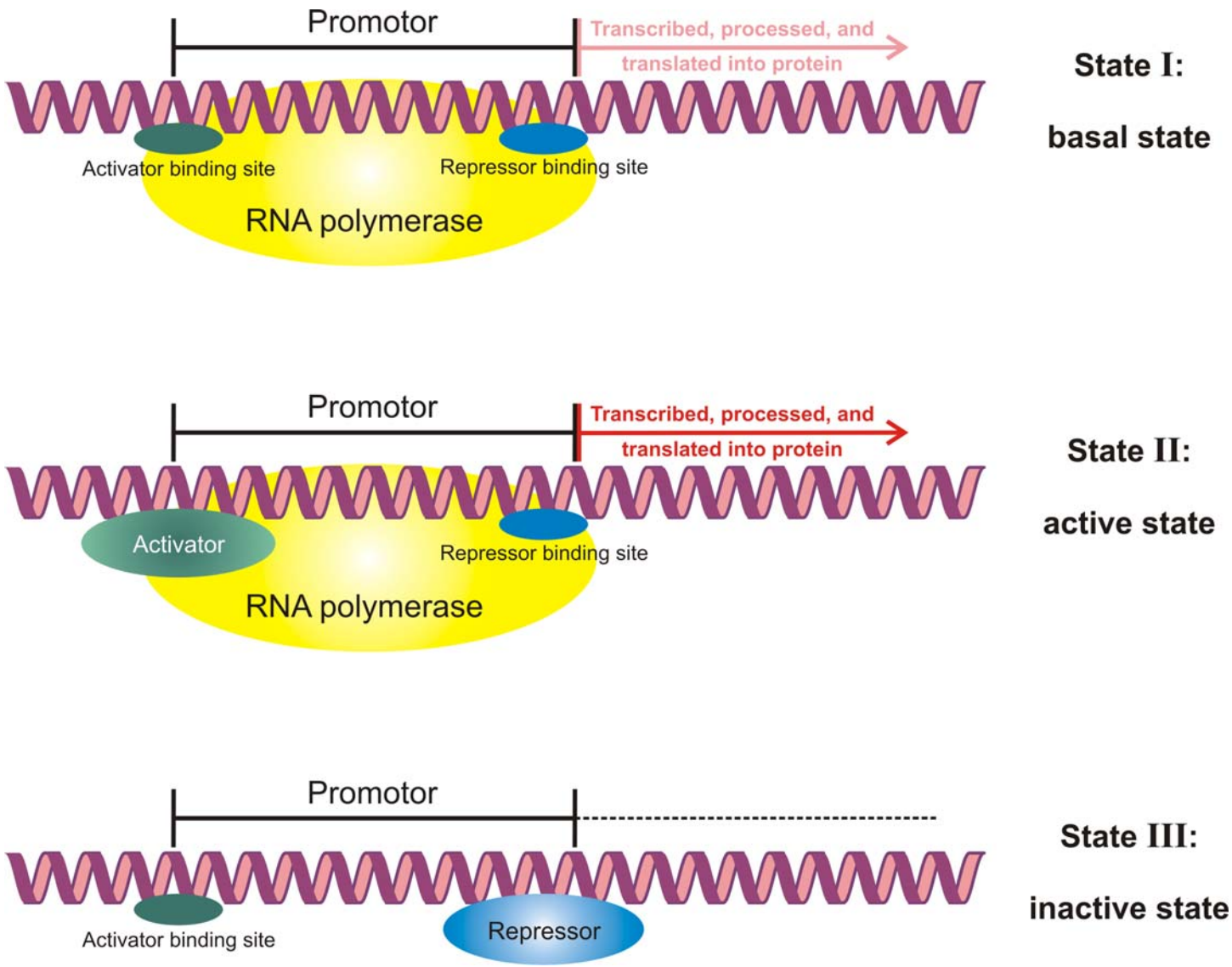
Bifurcation pattern

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



The inverse problem of bifurcation analysis (Level II)

1. Complex networks in cellular regulation
2. Experimental data and modeling in biology
3. Parameter determination and reverse engineering
4. **Gene regulation dynamics**
5. Inverse bifurcation analysis
6. Current challenges in biology



Three states of a gene regulated by activator and repressor

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 246 (2007) 395–419
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Dynamic patterns of gene regulation I: Simple two-gene systems

Stefanie Widder[a], Josef Schicho[b], Peter Schuster[a,c,*]

[a]*Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
[b]*RICAM—Johann Radon Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*
[c]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 24 February 2006; received in revised form 7 January 2007; accepted 8 January 2007
Available online 16 January 2007

Abstract

Regulation of gene activities is studied by means of computer assisted mathematical analysis of ordinary differential equations (ODEs) derived from binding equilibria and chemical reaction kinetics. Here, we present results on cross-regulation of two genes through activator and/or repressor binding. Arbitrary (differentiable) binding function can be used but systematic investigations are presented for gene–regulator complexes with integer valued Hill coefficients up to $n = 4$. The dynamics of gene regulation is derived from bifurcation patterns of the underlying systems of kinetic ODEs. In particular, we present analytical expressions for the parameter values at which one-dimensional (transcritical, saddle-node or pitchfork) and/or two-dimensional (Hopf) bifurcations occur. A classification of regulatory states is introduced, which makes use of the sign of a 'regulatory determinant' D (being the determinant of the block in the Jacobian matrix that contains the derivatives of the regulator binding functions): (i) systems with $D<0$, observed, for example, if both proteins are activators or repressors, to give rise to one-dimensional bifurcations only and lead to bistability for $n \geqslant 2$ and (ii) systems with $D>0$, found for combinations of activation and repression, sustain a Hopf bifurcation and undamped oscillations for $n>2$. The influence of basal transcription activity on the bifurcation patterns is described. Binding of multiple subunits can lead to richer dynamics than pure activation or repression states if intermediates between the unbound state and the fully saturated DNA initiate transcription. Then, the regulatory determinant D can adopt both signs, plus and minus.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Basal transcription; Bifurcation analysis; Cooperative binding; Gene regulation; Hill coefficient; Hopf bifurcation

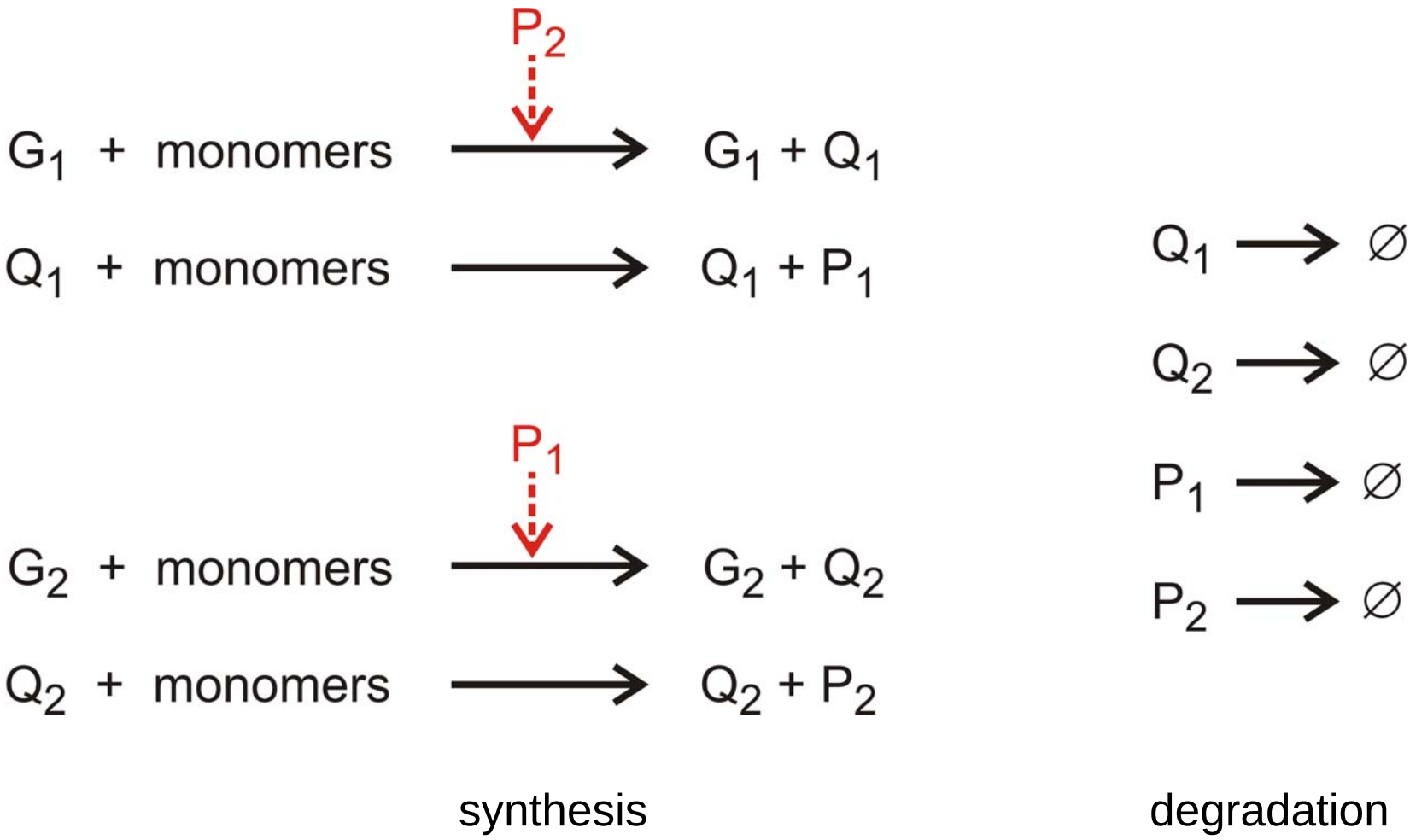
1. Introduction

Theoretical work on gene regulation goes back to the 1960s (Monod et al., 1963) soon after the first repressor protein had been discovered (Jacob and Monod, 1961). A little later the first paper on oscillatory states in gene regulation was published (Goodwin, 1965). The interest in gene regulation and its mathematical analysis never ceased (Tiwari et al., 1974; Tyson and Othmer, 1978; Smith, 1987) and saw a great variety of different attempts to design models of genetic regulatory networks that can be used in systems biology for computer simulation of *gen*(etic and met)*abolic* networks.[1] Most models in the literature aim at a minimalist dynamic description which, nevertheless, tries to account for the basic regulatory functions of large networks in the cell in order to provide a better understanding of cellular dynamics. A classic in general regulatory dynamics is the monograph by Thomas and D'Ari (1990). The currently used mathematical methods comprise application of Boolean logic (Thomas and Kaufman, 2001b; Savageau, 2001; Albert and Othmer, 2003), stochastic processes (Hume, 2000) and deterministic dynamic models, examples are Cherry and Adler (2000), Bindschadler and Sneyd (2001) and Kobayashi et al. (2003) and the recent elegant analysis of bistability (Craciun et al.,

*Corresponding author. Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.
Tel.: +43 1 4277 52743; fax: +43 1 4277 52793.
E-mail address: pks@tbi.univie.ac.at (P. Schuster).

[1]Discussion and analysis of combined genetic and metabolic networks has become so frequent and intense that we suggest to use a separate term, *genabolic networks*, for this class of complex dynamical systems.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.01.004

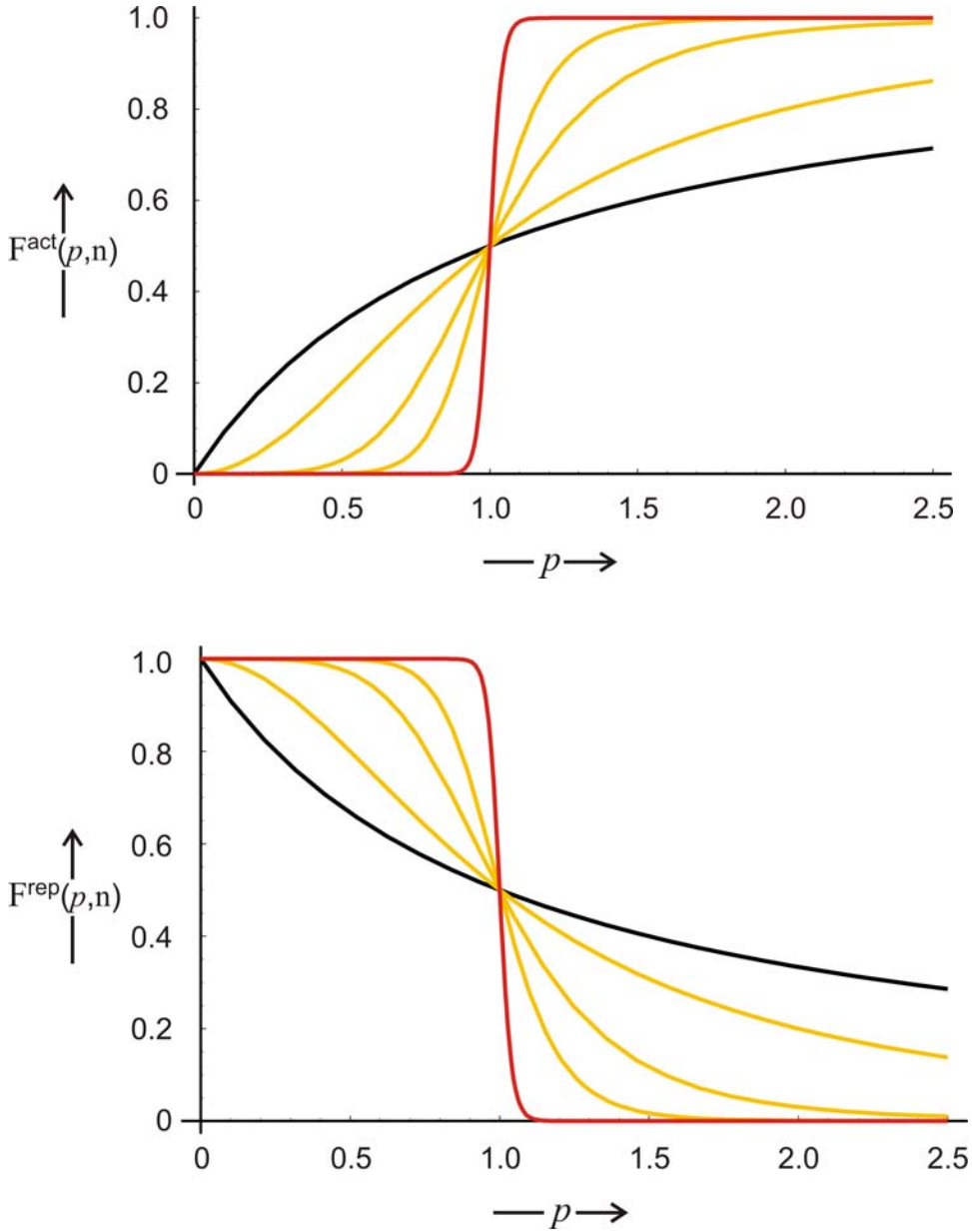


Cross-regulation of two genes

$$\text{Activation}: \quad F_i(p_j) = \frac{p_j^{\mathrm{n}}}{K + p_j^{\mathrm{n}}}$$

$$\text{Repression}: \quad F_i(p_j) = \frac{K}{K + p_j^{\mathrm{n}}}$$

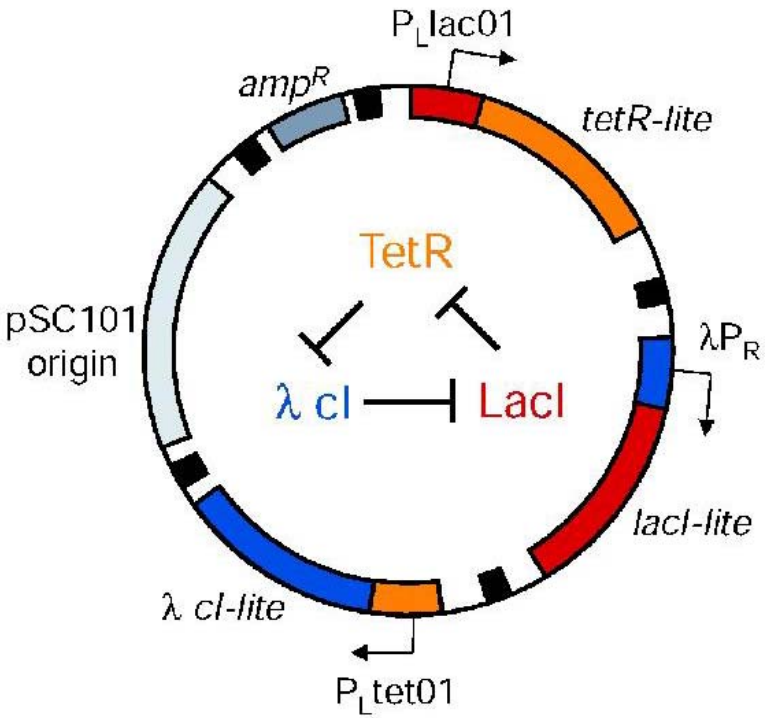
$$i, j = 1,2$$

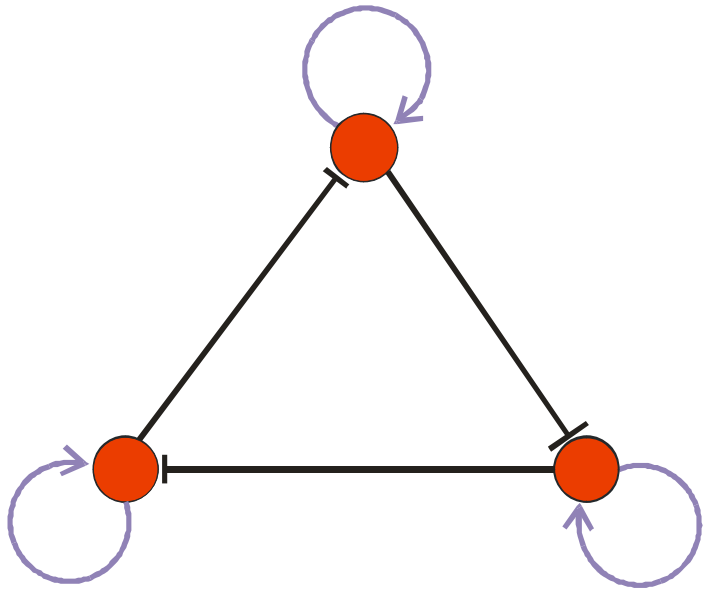


Gene regulatory binding functions

Hill coefficient: n	Act.-Act.	Act.-Rep.	Rep.-Rep.
1	S , E	S	S
2	E , B(E,P)	S	S , B(P_1,P_2)
3	E , B(E,P)	S , O	S , B(P_1,P_2)
4	E , B(E,P)	S , O	S , B(P_1,P_2)

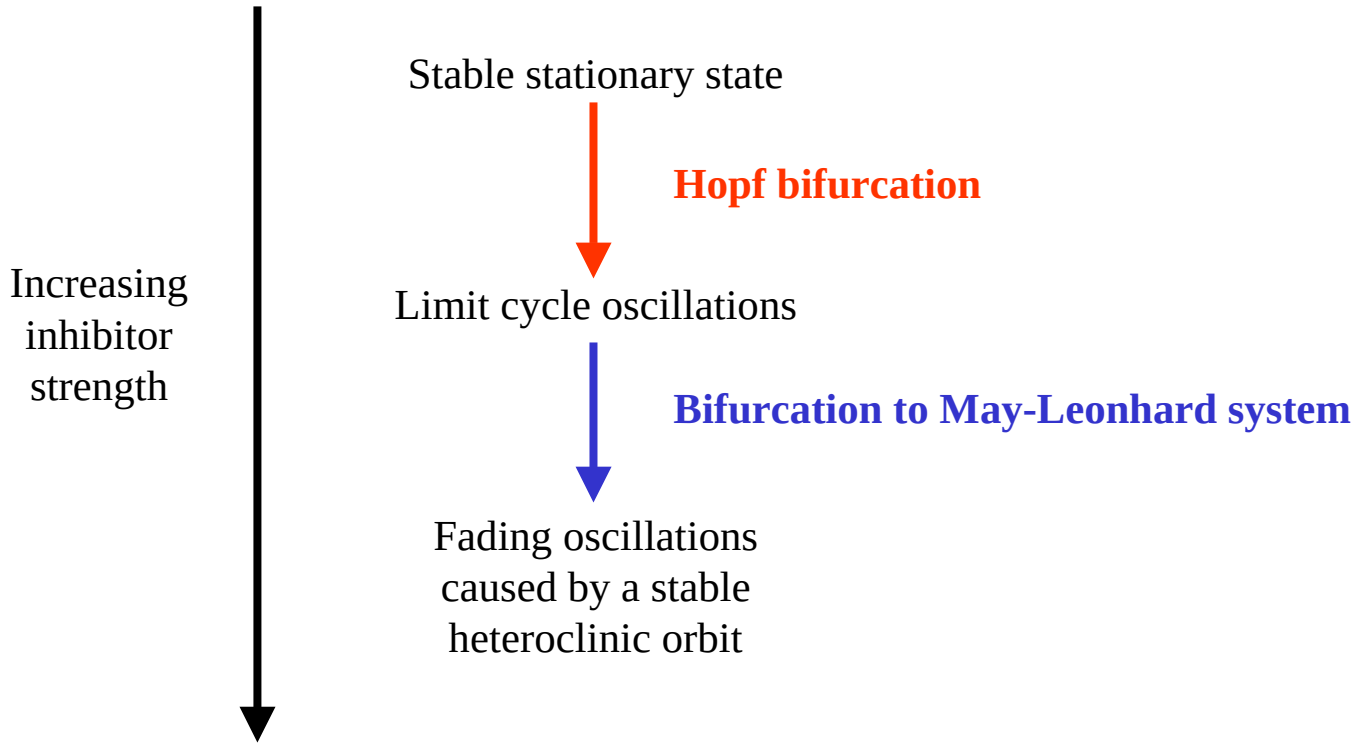
S ..... stable point attractor
E ..... extinction
O ..... oscillations
B ..... bistability

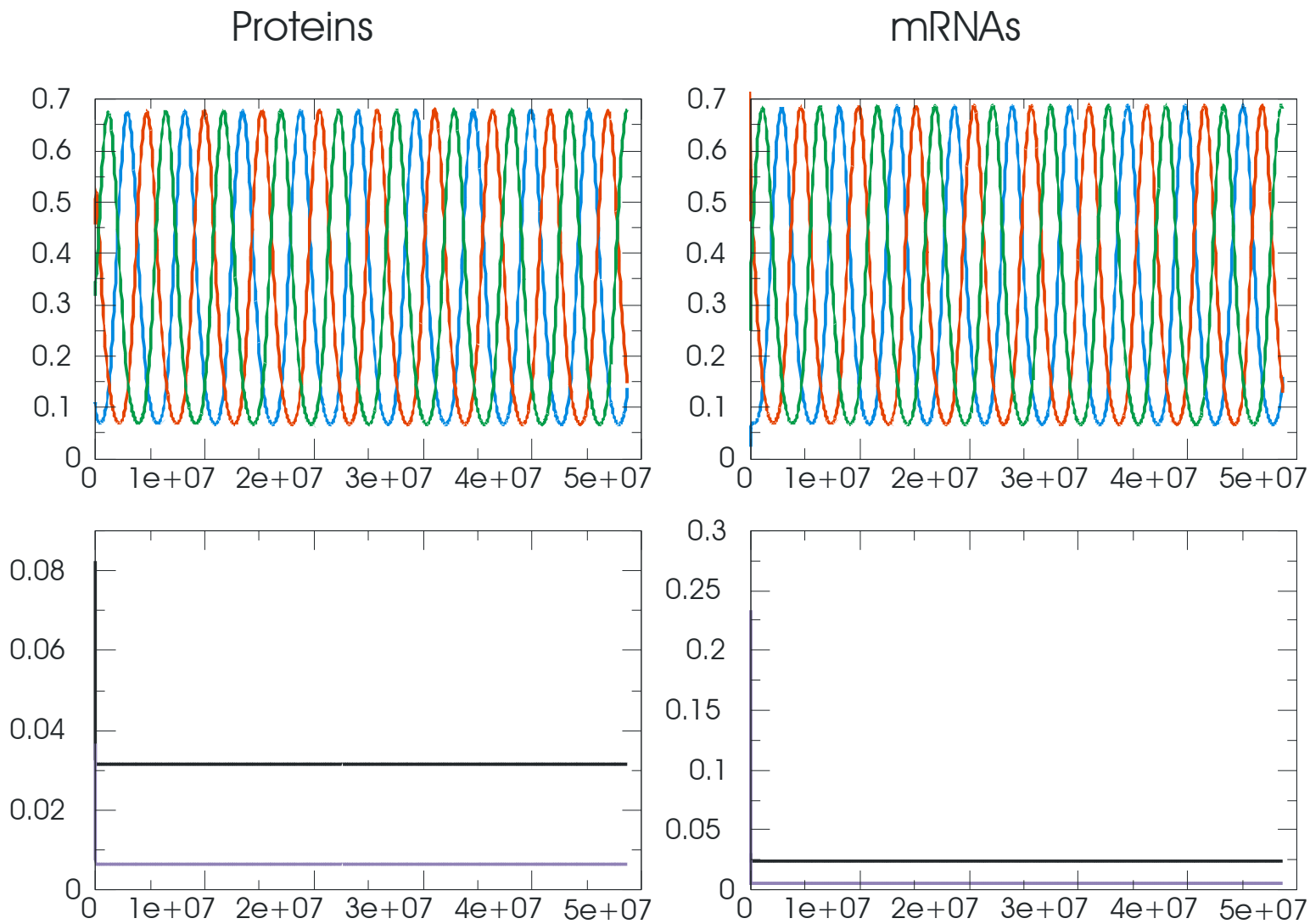




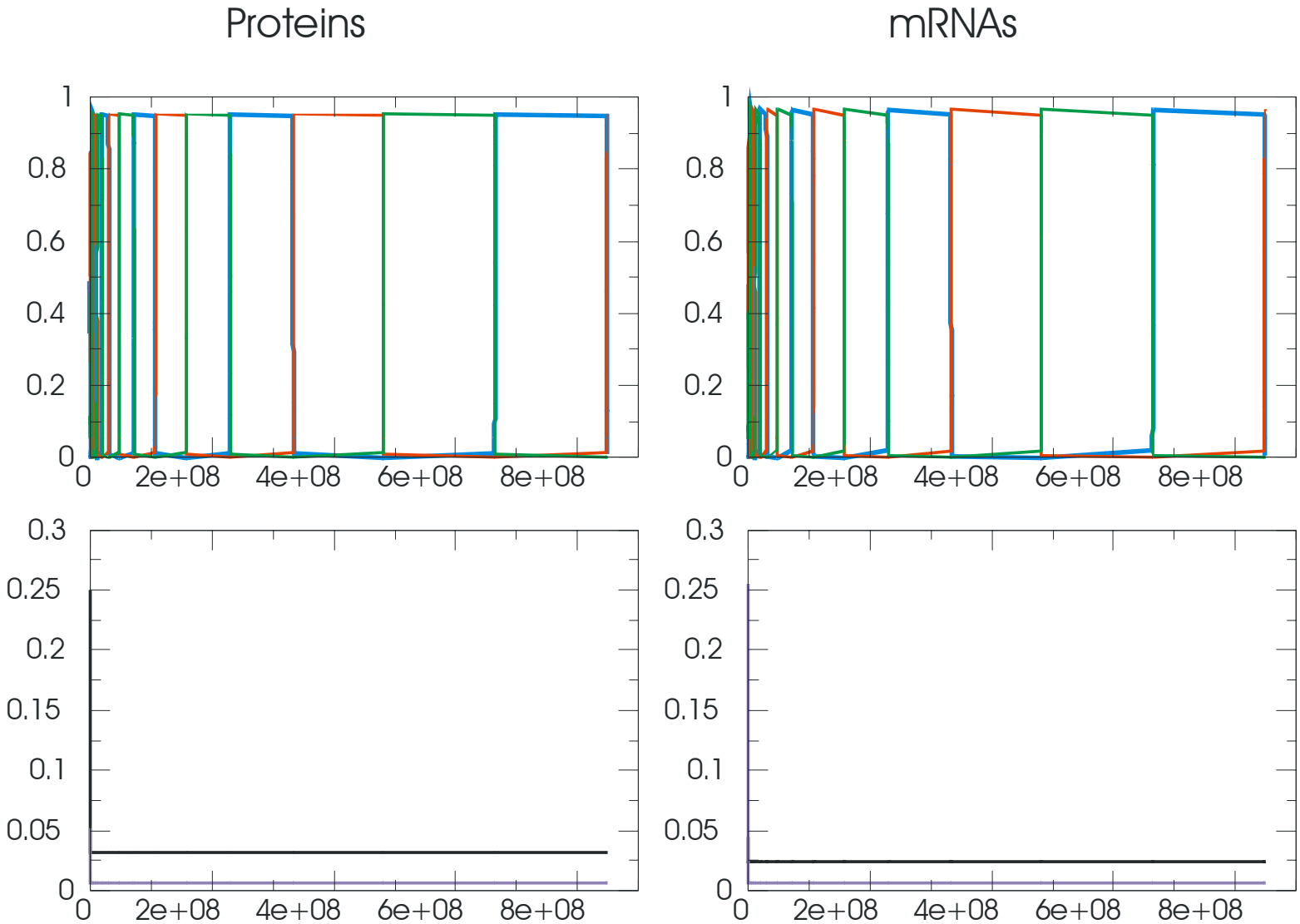
An example analyzed and simulated by MiniCellSim

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

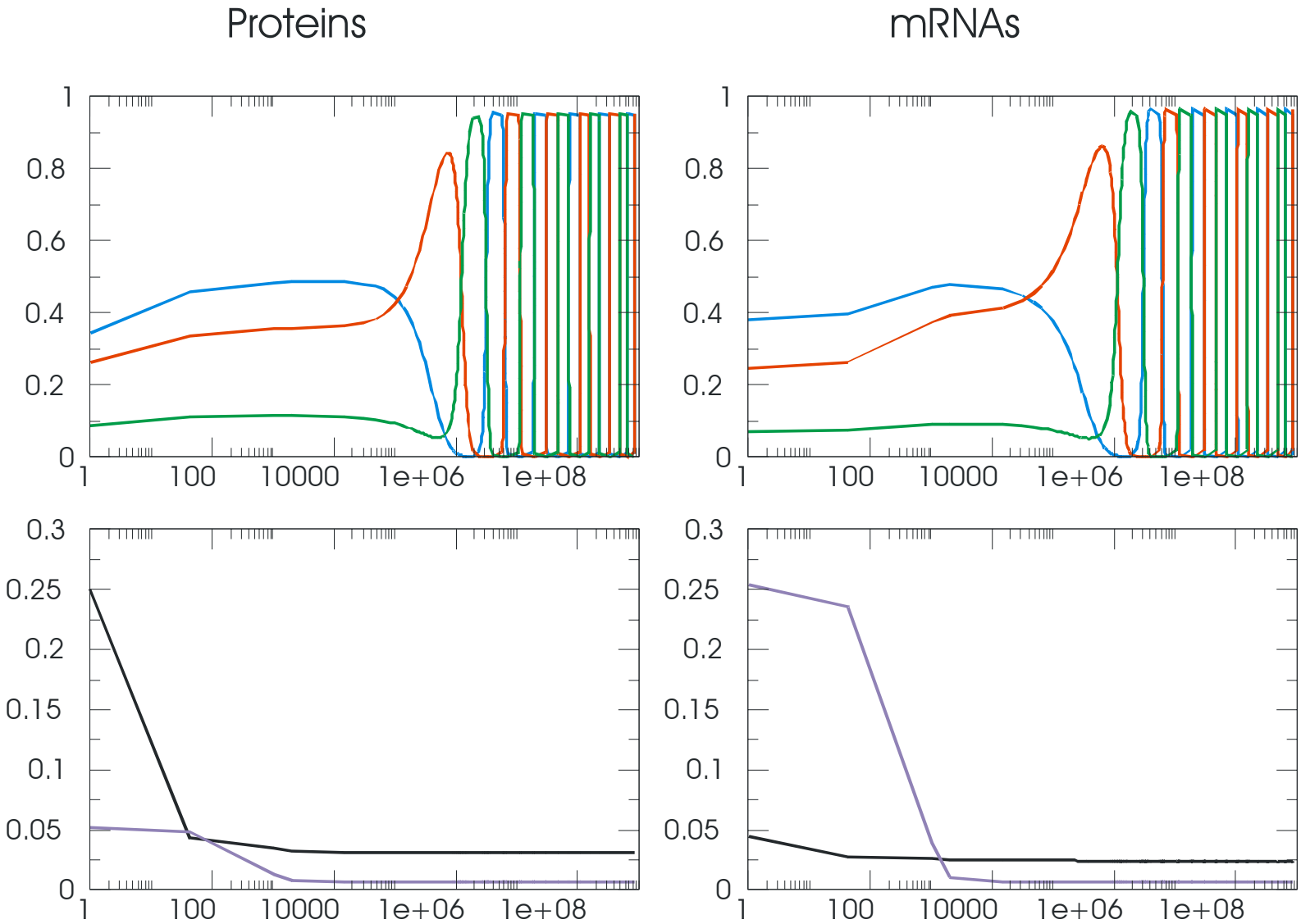

Increasing
inhibitor
strength
Stable stationary state
Hopf bifurcation
Limit cycle oscillations
Bifurcation to May-Leonhard system
Fading oscillations
caused by a stable
heteroclinic orbit



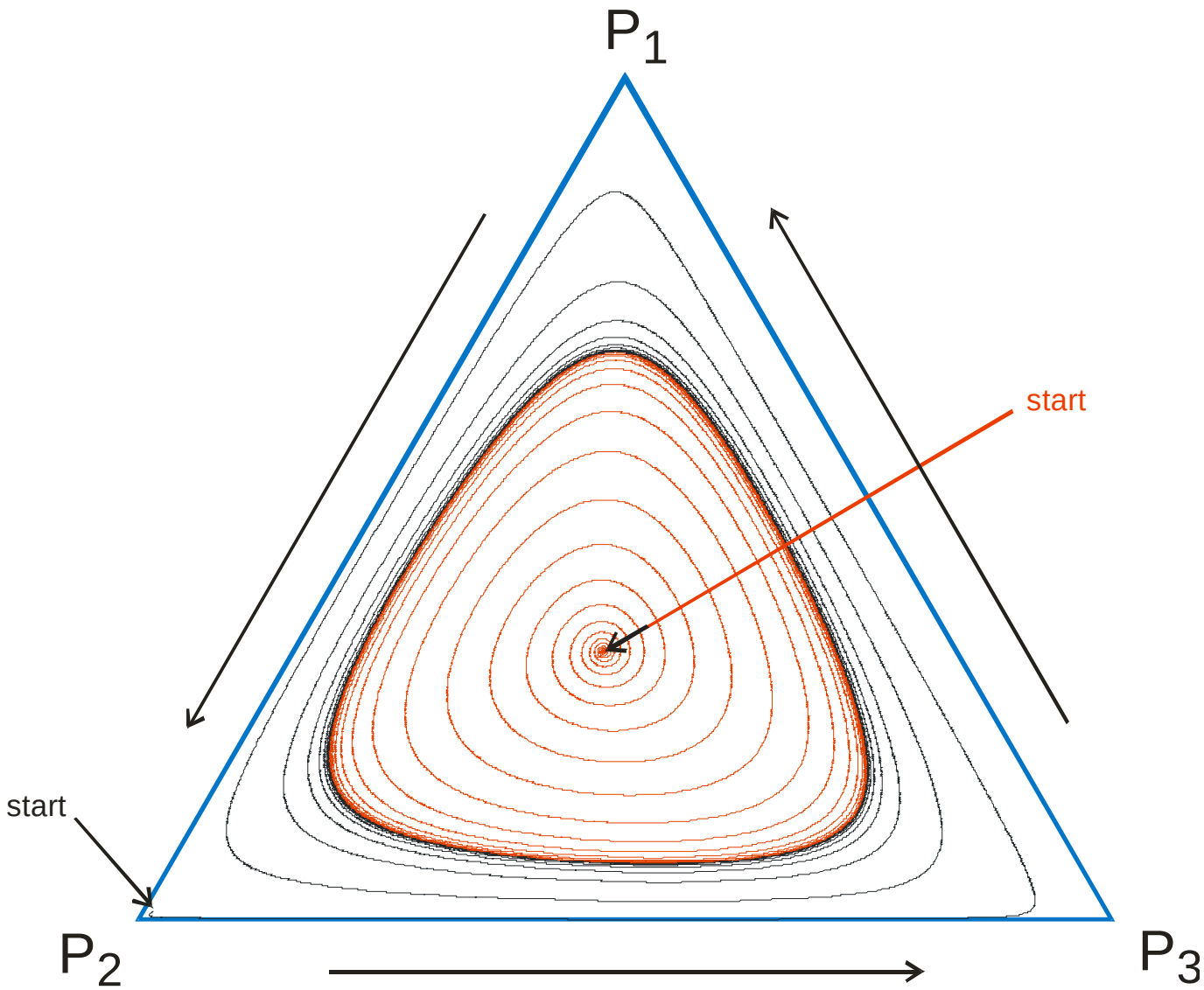
The repressilator limit cycle



The repressilator heteroclinic orbit

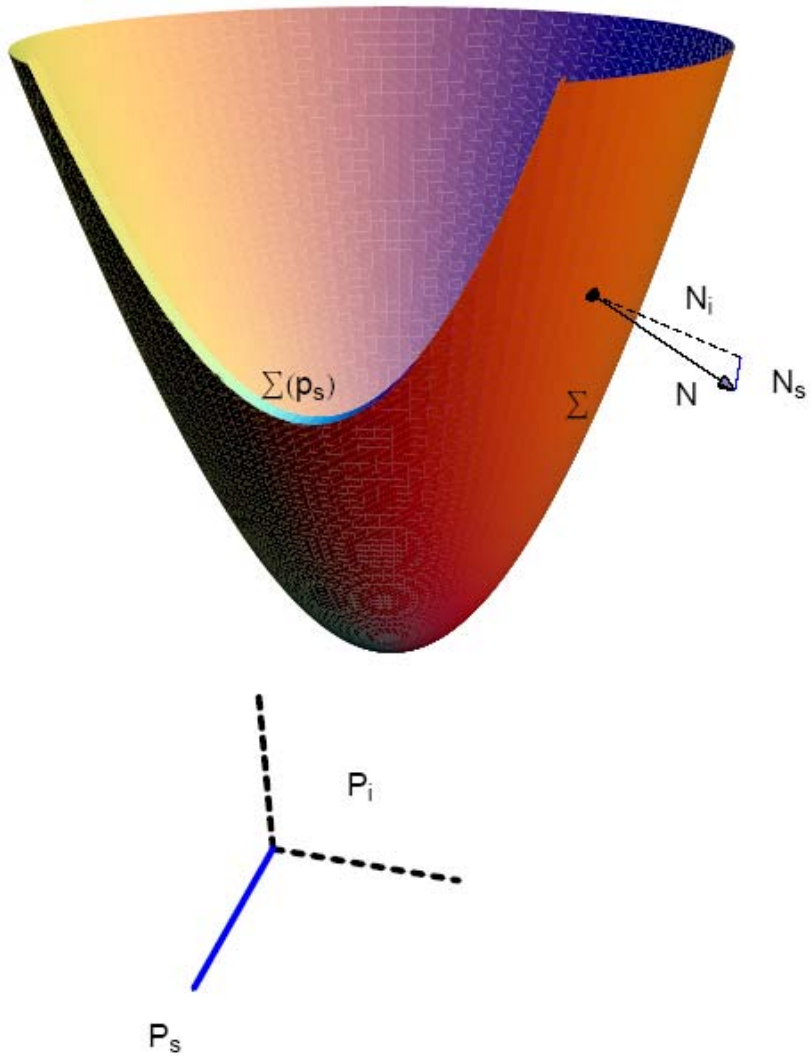


The repressilator heteroclinic orbit (logarithmic time scale)

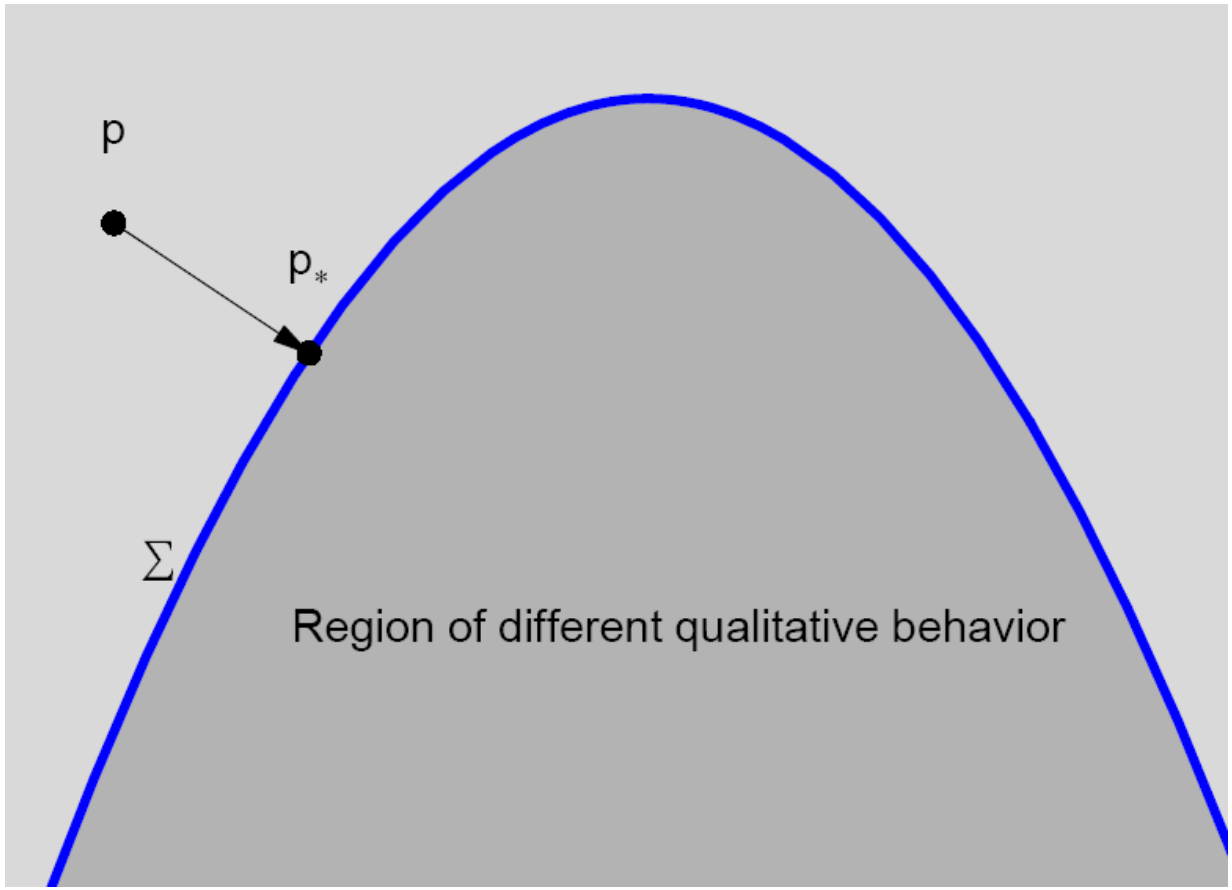


The repressilator limit cycle

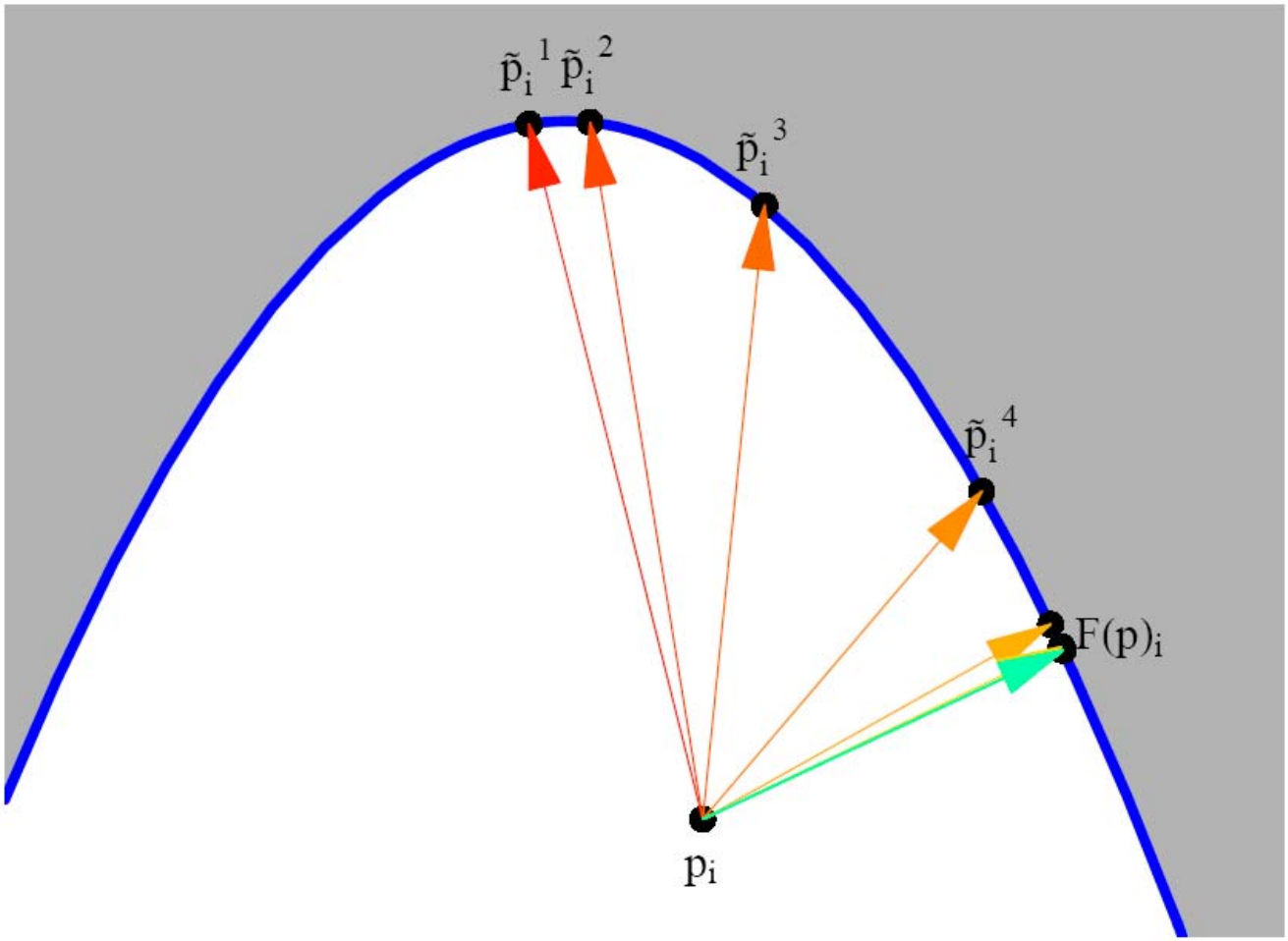
1. Complex networks in cellular regulation
2. Experimental data and modeling in biology
3. Parameter determination and reverse engineering
4. Gene regulation dynamics
5. **Inverse bifurcation analysis**
6. Current challenges in biology



The bifurcation manifold



Defininition of the forward operator $F(p)$



Iterative solution for min J(p)

$$\begin{aligned}\dot{x}_i &= \beta_i(y_i - x_i) \\ \dot{y}_i &= \alpha_i \left(\frac{1-\delta_i}{1 + x_{i-1 \mod n}^{h_i}} + \delta_i \right) - y_i, \; i = 0, \cdots, n-1\end{aligned}$$

$$\alpha_i = \alpha, \beta_i = \beta, h_i = h, \delta_i = \delta$$

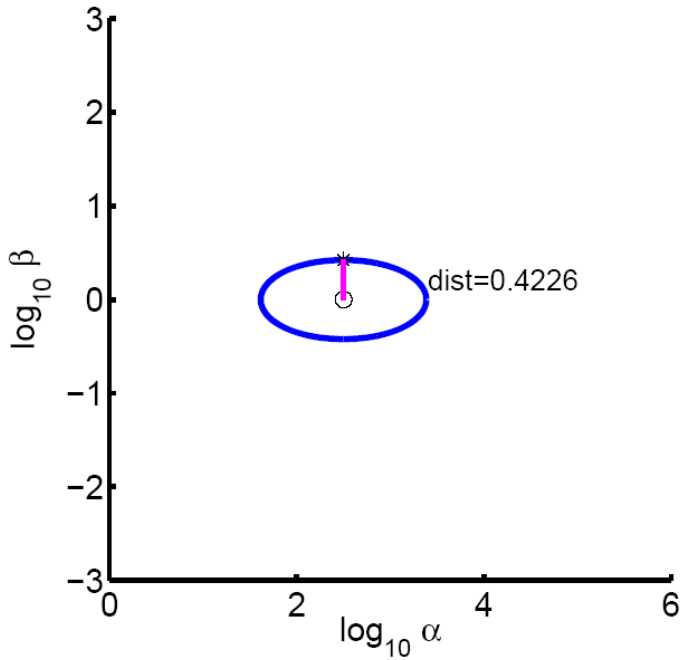
$$p_i = (\alpha, \beta)$$

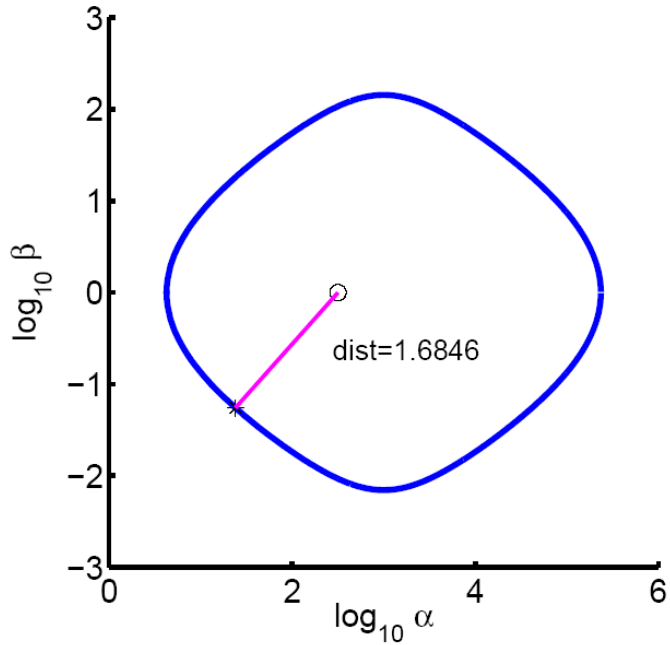
$$p_s = (\delta, h)$$

$$(10^{-4}, 0) \leq (\delta, h) \leq (10^{-1}, 2)$$

Inverse bifurcation analysis of the repressilator model

S. Müller, J. Hofbauer, L. Endler, C. Flamm, S. Widder, P. Schuster. A generalized model of the repressilator. *J. Math. Biol.* **53**:905-937, 2006.





Inverse bifurcation analysis of the repressilator model

J. Lu, H.W. Engl, P. Schuster. Inverse bifurcation analysis: Application to simple gene systems. *AMB Algorithms for Molecular Biology* **1**:11, 2006.

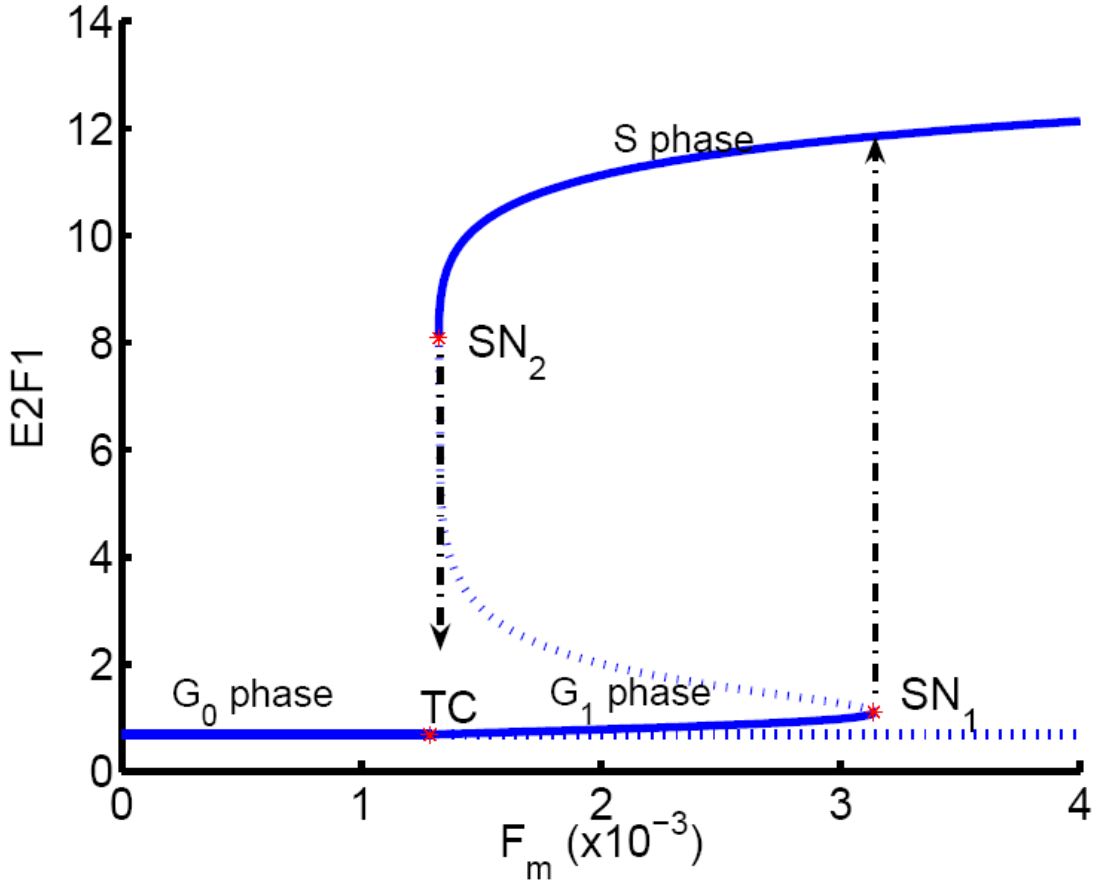
$$\frac{d}{dt}[\text{pRB}] = k_1 \frac{[\text{E2F1}]}{K_{m1} + [\text{E2F1}]} \frac{J_{11}}{J_{11} + [\text{pRB}]} - \phi_{\text{pRB}}[\text{pRB}]$$

$$\frac{d}{dt}[\text{E2F1}] = k_P + k_1 \frac{a^2 + [\text{E2F1}]^2}{K_{m2}^2 + [\text{E2F1}]^2} \frac{J_{12}}{J_{12} + [\text{pRB}]} - \phi_{\text{E2F1}}[\text{E2F1}]$$

$$\frac{d}{dt}[\text{AP1}] = F_m + k_{25}[\text{E2F1}] \frac{J_{15}}{J_{15} + [\text{p}RB]} \frac{J_{65}}{J_{11} + [\text{pRB}']} - \phi_{\text{AP1}}[\text{AP1}]$$

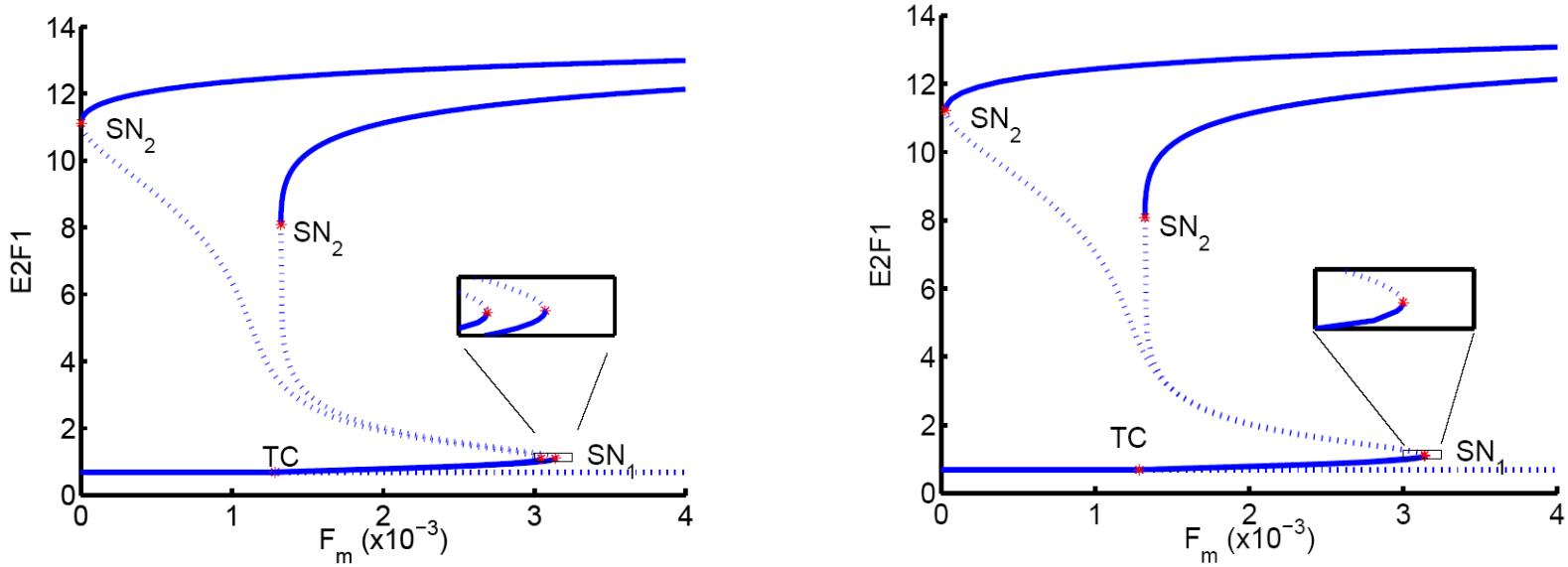
A simple dynamical cell cycle model

J.J. Tyson, A. Csikasz-Nagy, B. Novak. The dynamics of cell cycle regulation. *Bioessays* **24**:1095-1109, 2002



A simple dynamical cell cycle model

J.J. Tyson, A. Csikasz-Nagy, B. Novak. The dynamics of cell cycle regulation. *Bioessays* **24**:1095-1109, 2002



Inverse bifurcation analysis of a dynamical cell cycle model

J. Lu, H.W. Engl, P. Schuster. Inverse bifurcation analysis: Application to simple gene systems. *AMB Algorithms for Molecular Biology* **1**:11, 2006.

1. Complex networks in cellular regulation
2. Experimental data and modeling in biology
3. Parameter determination and reverse engineering
4. Gene regulation dynamics
5. Inverse bifurcation analysis
6. **Current challenges in biology**

Explanation of important global properties

homeostasis

robustness

stability against mutation

self-repair or regeneration

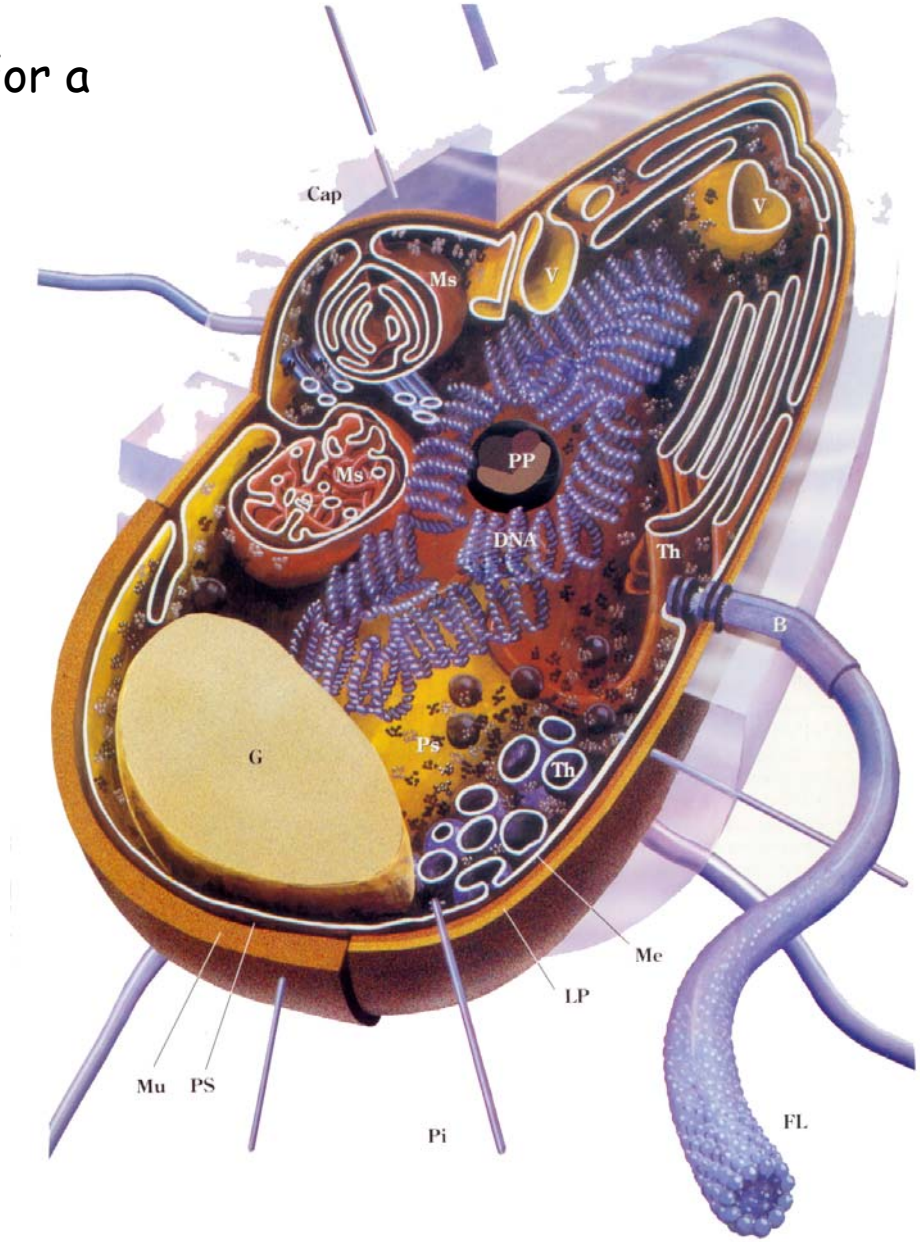
.....

The bacterial cell as an example for a simple form of autonomous life

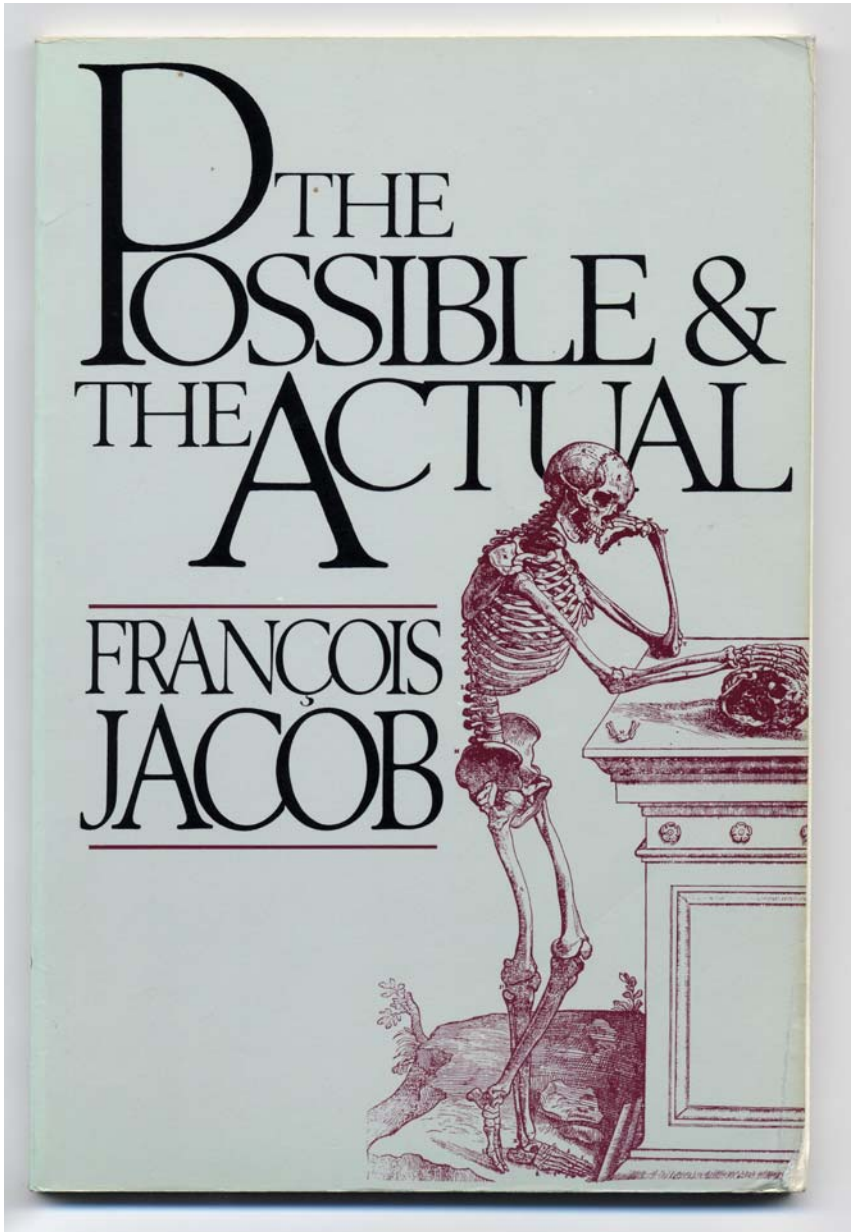
Escherichia coli genome:

4 million nucleotides

4460 genes



The structure of the bacterium *Escherichia coli*



Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

François Jacob. *The Possible and the Actual*. Pantheon Books, New York, 1982, and

Evolutionary tinkering. *Science* **196** (1977), 1161-1166.

The difficulty to define the notion of „gene".

Helen Pearson,
Nature **441**: 399-401, 2006

WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past[1]. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals[2]. If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist Laurence Hurst at the University of Bath, UK.

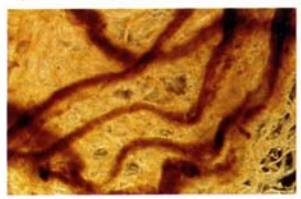
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

"We've come to the realization that the genome is full of overlapping transcripts."
— Phillip Kapranov

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the previously unimagined scope of RNA.



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

P. PLAILLY/SPL

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from[3].

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team[4], and one by geneticist Rotem Sorek[5], now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

ENCODE stands for
ENCyclopedia Of DNA Elements.

ENCODE Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
Nature **447**:799-816, 2007

Web-Page for further information:

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