



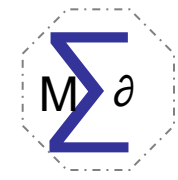
系统生物学 (Systems Biology)

马彬广



系统生物学建模概述

(第二讲)



纲要



- 系统与生物系统
- 模型与建模
- 系统生物学中的可计算模型举例
- 复杂性与系统分析



系统



□ 系统是相互联系、相互作用的诸元素的综合体 ---贝塔朗菲

□ 系统定义

如果对象集 S 满足以下两个条件：

(1) S 中至少包含两个不同的对象；

(2) S 中的对象按一定的方式相互联系在一起；

则称 S 为一个系统，称 S 中的对象为系统的组分，即组成部分。

-----苗东升《系统科学精要》

□ 系统特征：（1）多元性；（2）相关性；（3）整体性；

思考：什么是非系统？自然界中是否存在非系统？



生物系统



生物系统是指来自有机（organic）界的系统。

What is Life?

Life as we know. （如吾所识的生命，根据经验定义）

Life as it could be. （如其所能的生命，根据特征定义）

尚未解决的问题：

生命从那个层次开始？（cell?）

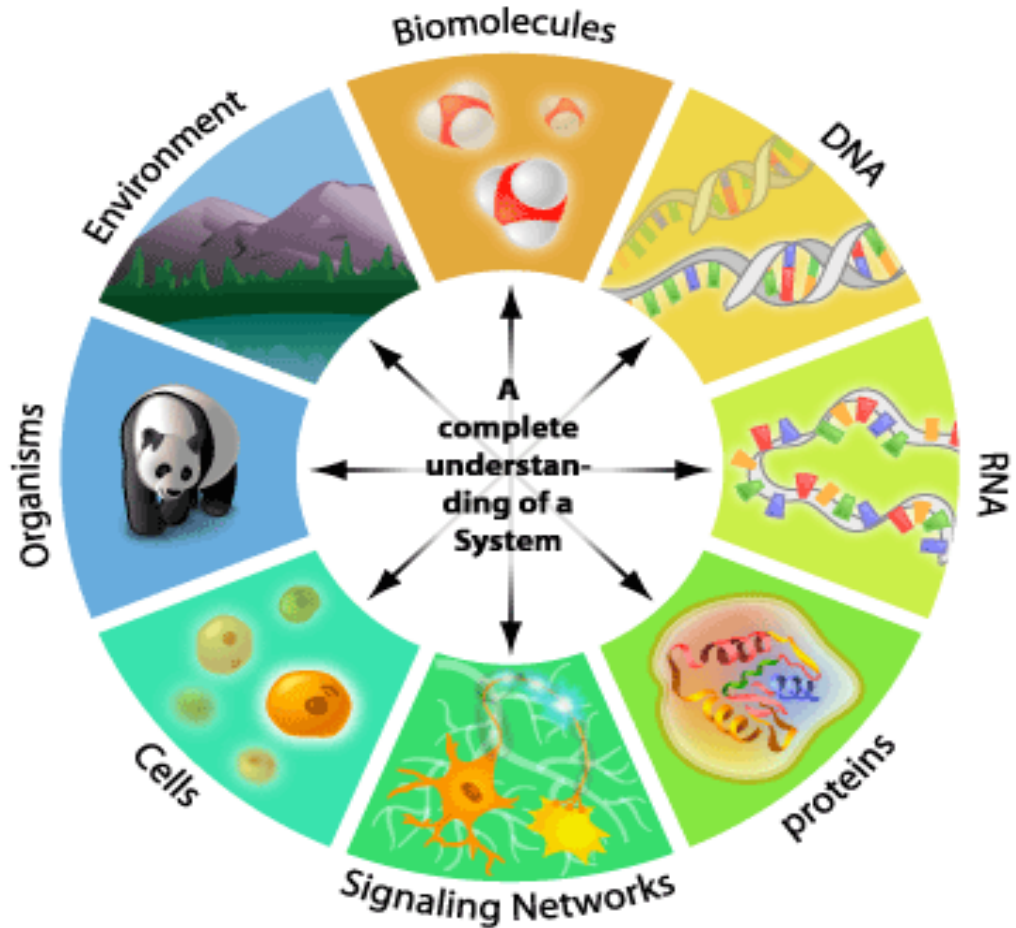
生命特征的最小子集？（self-production?）



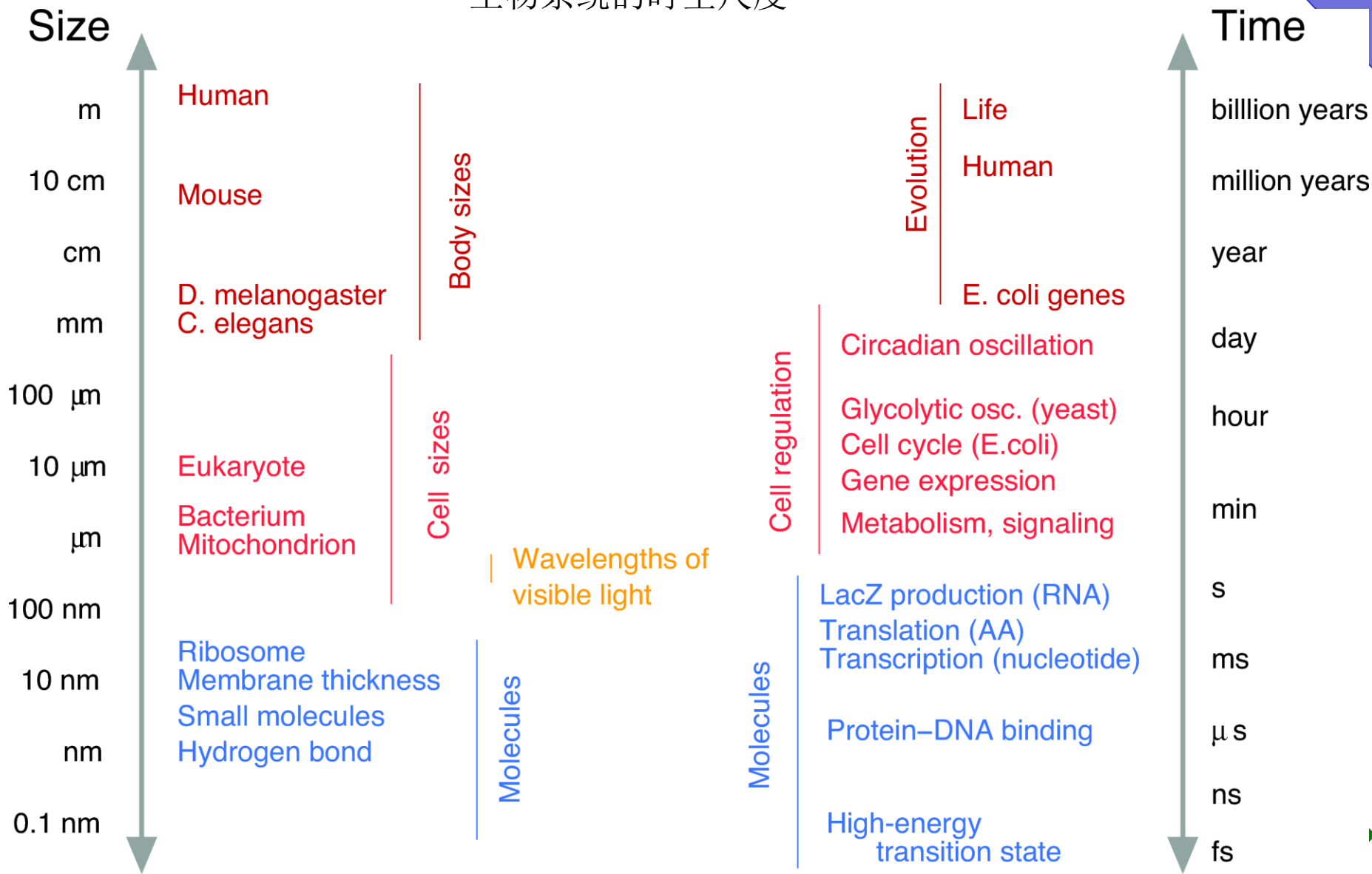
生命系统的层次



- ...
- 生物圈
- 生态系统
- 群体
- 个体
- 器官
- 组织
- 细胞
- 分子
- 原子
- ...



生物系统的时空尺度



© 2010 Wiley-VCH, Weinheim
 Klipp - Systems Biology
 ISBN: 978-3-527-31874-2 fig-01-01

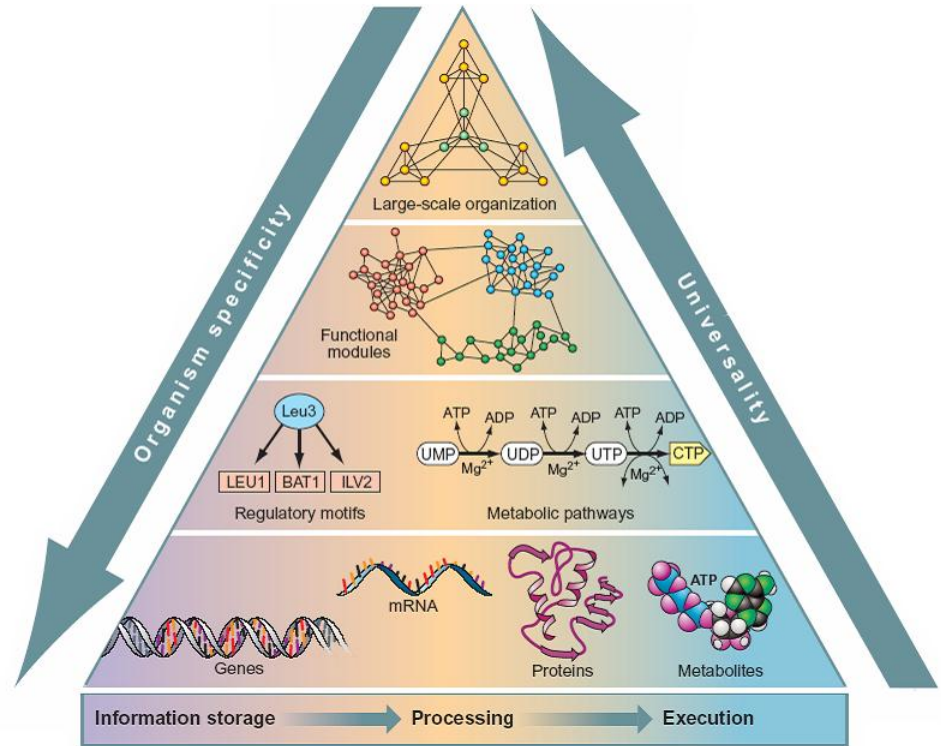
<http://bionumbers.hms.harvard.edu/>



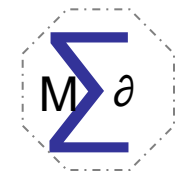
生命系统的特征



- 开放性 (Open)
- 复杂性 (Complex)
- 层级性 (Hierarchical)
- 动态性 (Dynamic)
- 稳健性 (Robust)
- 适应性 (Adaptive)
- 进化能力 (Evolvability)



Oltvai & Barabasi, Life's complexity pyramid, *Science* 2002



纲要



- 系统与生物系统
- 模型与建模
- 系统生物学中的可计算模型举例
- 复杂性与系统分析



模型与建模



什么是模型？

- 模型是与某物（原型）类似的替代物。
- 模型可以是具体的，也可以是抽象的。
- 模型的特点：（1）有效性；（2）可操作性；（3）简单性。
- 模型的极限是原型。

如何建模？

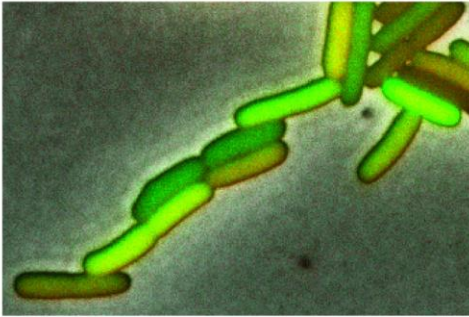
- 大而小之
- 小而大之
- 繁而简之
- 实而虚之
- 虚而实之



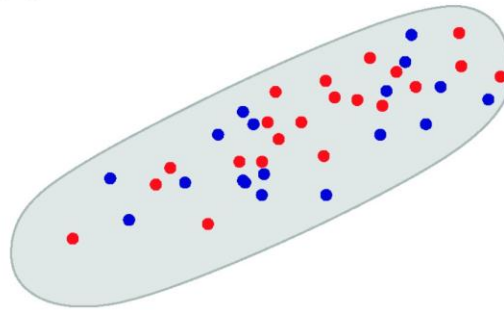
生物细胞的建模流程



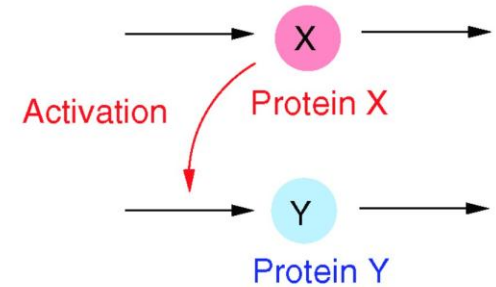
(a) Biological system



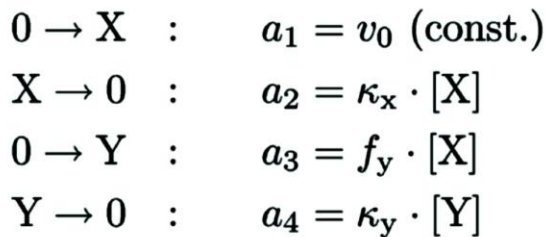
(b) Mental model



(c) Model scheme



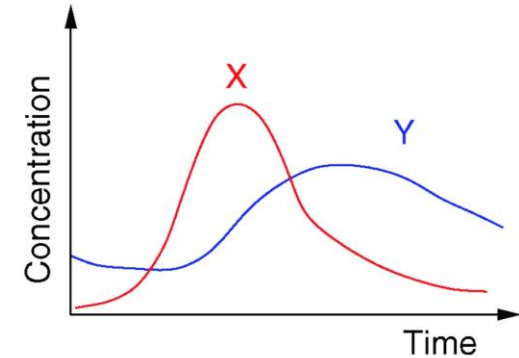
(d) Process model



(e) Dynamical model

$$\begin{aligned} dx/dt &= v_0 - \kappa_x x \\ dy/dt &= f_y x - \kappa_y y \\ x(0) &= x_0 \\ y(0) &= y_0 \end{aligned}$$

(f) Quantitative results



© 2010 Wiley-VCH, Weinheim
Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-01-02



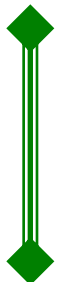
建模中的注意事项



- 主观性：建模是一个主观选择的过程，多层面、多手段。
- “黑箱”与“白箱”：
 - 黑箱建模，只需拟合输入输出关系；
 - 白箱建模，需要弄清系统的工作机制。
- “通用性”与“专用性”：
 - 有些模型，适用于各种生物学过程；
 - 有些模型，则适用于特定的生物学过程。
- “简单性”与“细节性”：简单性保证可解；细节性保证预测能力。
- 谚语：“一切模型皆是错的，但某些是有用的” --- George Box.

可解算模型 (computable models)

- 网络模型 (组分结构模型)
 - 静态网络
 - 生灭网络
 - 传递网络 (Petri Net)
- 方程模型 (数量关系模型)
 - 代数方程
 - 微分方程
 - 随机方程
- 自动机模型 (计算机模型)
 - 元胞自动机
 - Multi-Agent
 - 对象化模型 (分类+消息)





计算模型的好处



- 使概念精确化，假设严格化。
- 突显知识上的漏洞，未定的组份或作用必须被确定下来。
- 模型可以单独运行，不依赖于被建模对象的状态。
- 可以对时空尺度进行缩放（通过变化步长和单位）。
- 模型求解的算法和计算机程序具有一定的通用性，可反复使用(reusability)。
- 计算模型的建立和求解的代价，一般比做实验要小。
- 计算模型替代实验，减少动植物使用的伦理问题，对环境无污染。
- 模型结果的呈现方式，数学化，便于推广，可视化，直观理解。
- 计算模型可给出定量的、可验证的预测。



计算模型对实验的辅助作用



- 只要模型充分有效，则可以预演各种可能出现的情形。
- 可以追踪实验上无法测量的成分随时间的变化过程。
- 可以施加真实系统中无法施加的扰动（**Perturbations**）。
- 精确地改变某一种组分的含量，而保持其它组分的量不变，这在真实的实验系统里通常是做不到的。
- 可根据不同条件，反复运行建立的模型程序。

总之，计算模型和实验系统比较起来，提供了更大的自由度和灵活性。



计算模型中的基本概念



- 模型的范围（model scope）：只研究系统的某些方面的特征，简化的艺术。
- 模型的表述（model statements）：动力学方程，约束条件，随机因素等。
- 系统状态（System State）：即，系统某一刻的快照，用一组状态变量的取值来表征。（确定性模型，当前状态蕴含未来系统状态变换的所有信息；随机模型，当前状态蕴含系统未来状态的概率）
- 系统的变量，参数和常数：变量，随时在变；参数，虽条件而变；常数，自始至终不变。（变量、参数和常数的相对性，虽模型不同，而性质不同）
- 模型的行为（Model Behavior）：取决于内、外两方面的因素，模型的结构（变量，参数和常数的相互关系）决定模型的行为（Output），但不同结构的模型可能产生类似的行为，故此，仅考察系统的输出，往往无法确定系统的内部构成。
- 模型的分类（Model Classification）：定性与定量，确定与随机，离散与连续，可逆性，周期性。



计算模型中的基本概念



- 系统的稳态（**Steady States**）：系统的状态变量的取值不再随时间而改变的状态。动力学系统的终结状态（时间趋于无穷时的状态）之一（另外两种终结状态是振荡和混沌）。
- 系统的稳态实际上是一种抽象，因为，万物流逝；对于快速进行的过程（**fast processes**）来说，经历一段时间之后，回进入变化很缓慢的准稳态（**quasi-steady state**）；而对于慢过程来说，在考察时间较短时，可以看成是常量。
- 故此，现实地看，每一个稳态，可以看作是一个准稳态镶嵌在一个更大一点的非静止的环境之中，即，稳态是对前述状态的一种理论上的抽象。
- 尽管是一种理想化的抽象，稳态的概念对于研究和分析系统的行为，非常有用，既可以解释系统的规律，又可以简化数学运算。
- 系统生物学中，其它理想化的简化方式，还包括：用**Boolean Network**表述基因调控，用质量作用定律描述酶动力学等。



计算模型中的基本概念



建模的**不唯一性**:

- 同一生物系统，可以用不同的实验方法进行研究，从而得到不同的数据和模型。
- 同一种建模手段，可用于描述不同的生物学过程。
- 模型之间也有包含和扩展等逻辑关系。
- 建模的不唯一性，带来的好处是，可以用不同手段，从不层面反映系统的特征；
- 坏处是，给模型整合带来很大的困难，比如，从个别的代谢途径的模型，整合出全细胞代谢的模型就很困难。



建模需要数据整合



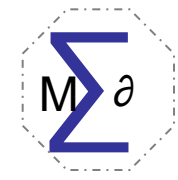
- 高通量技术和各种组学的开展，为系统生物学建模提供数据。
- 数据交换需要规范化的格式，模型描述也需要统一的语言，这些需求催生了各种标准化协议。很多“Minimum-information about xxxx”的陈述，表示对某些实验或建模的最小信息需求。例如：

MIAME: Minimum information about a microarray experiment.

MIRIAM: Minimum information requested in the annotation of biochemical models.

等。

- 模型描述语言主要是基于XML实现的，如SBML（Systems Biology Markup Language），CellML等。
- 更高级的数据整合，是要把分层次的信息之间的关联找出来，构建统一的模型。



纲要



- 系统与生物系统
- 模型与建模
- 系统生物学中的可计算模型举例
- 复杂性与系统分析

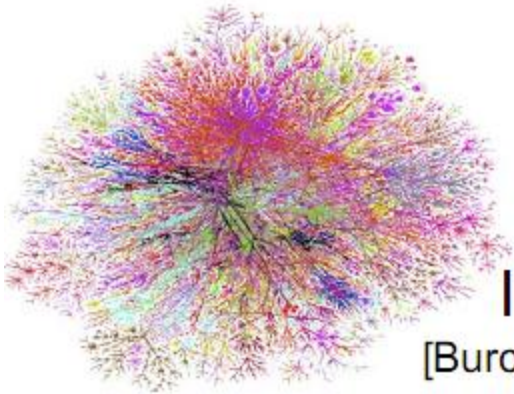


模型举例

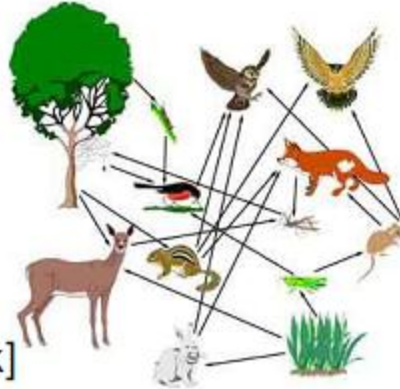


- 网络模型
- 动力学模型
- 元胞自动机模型

Networks as a universal language



Internet
[Burch & Cheswick]



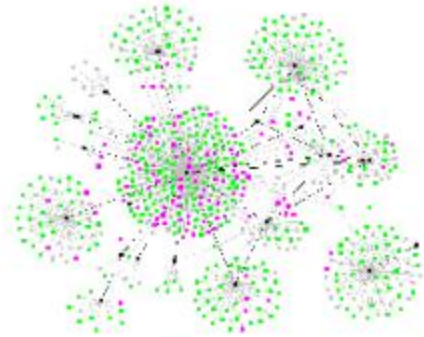
Food Web



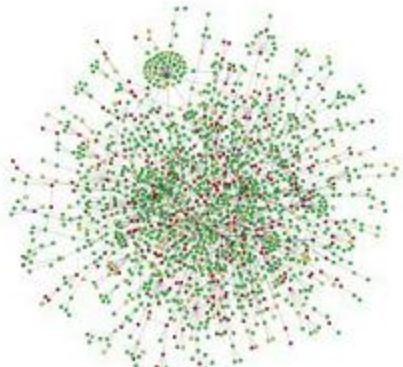
Electronic
Circuit



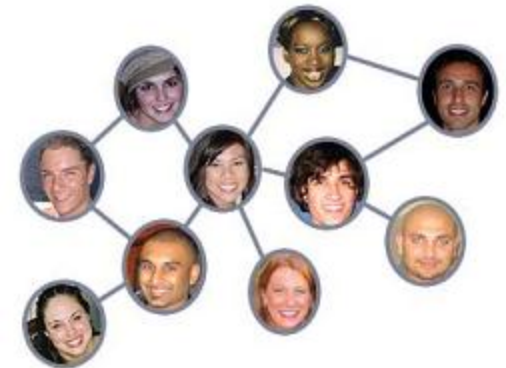
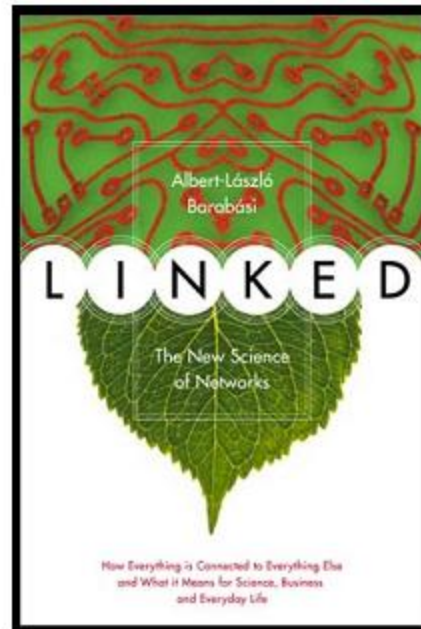
Neural Network
[Cajal]



Disease
Spread
[Krebs]



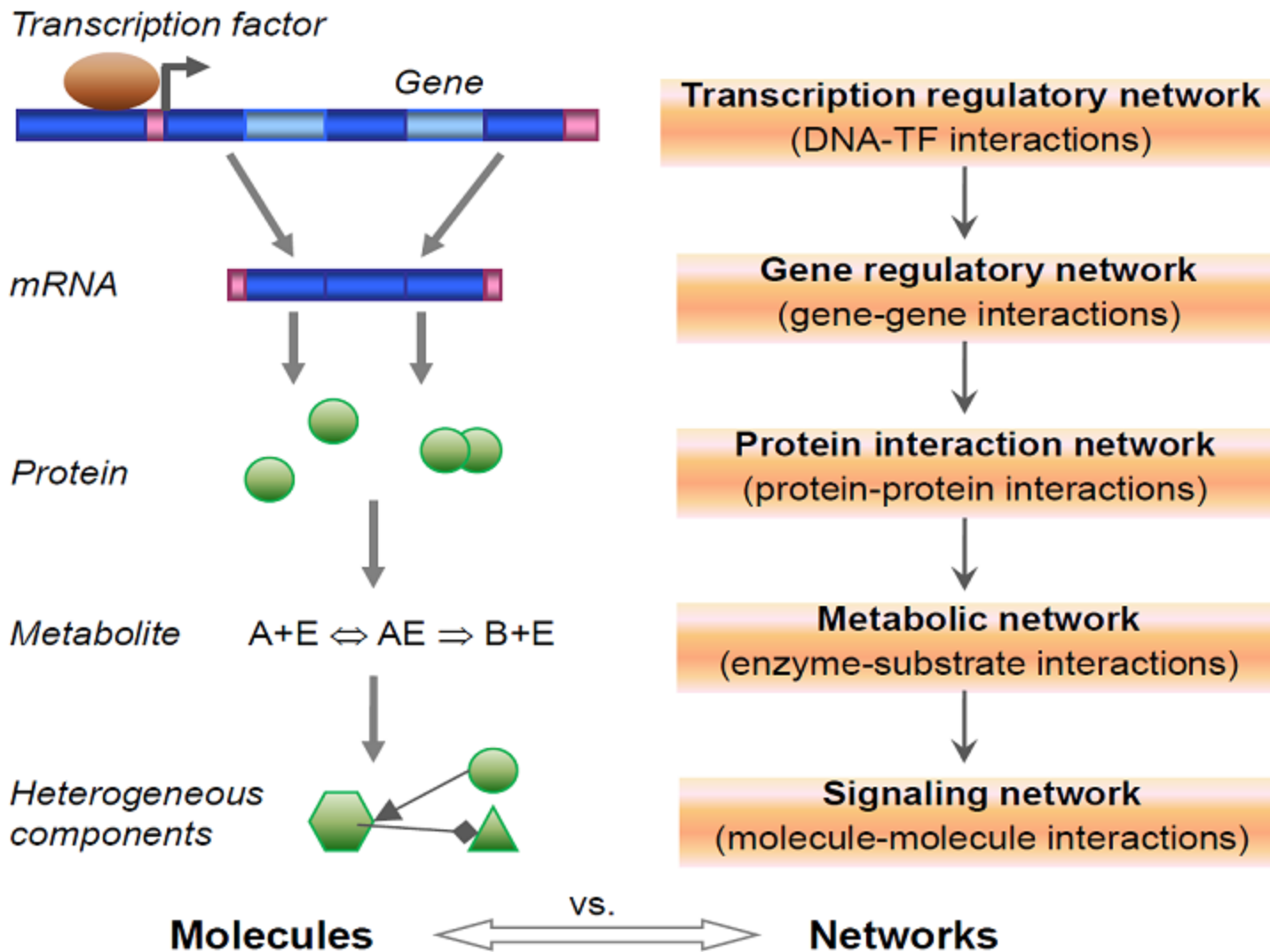
Protein
Interactions
[Barabasi]



Social Network

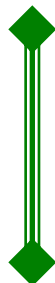
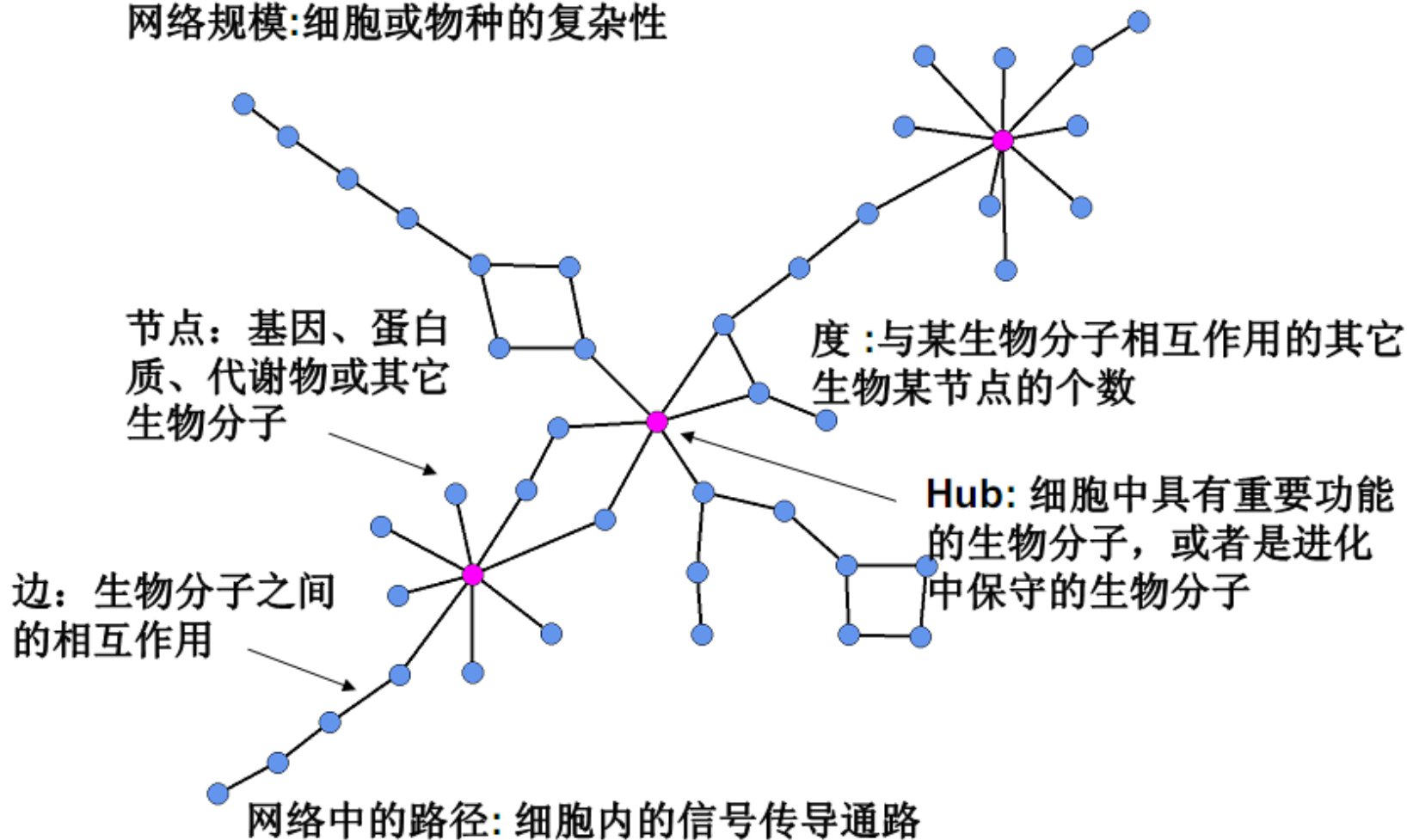


Biomolecular Networks ?



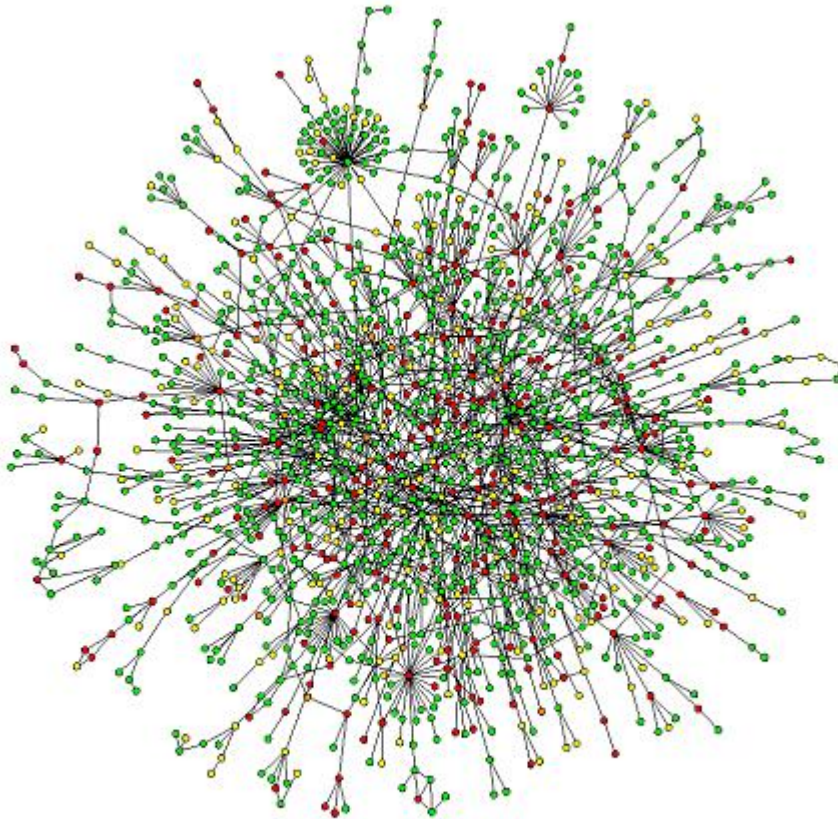
生物学语言下的网络

网络规模:细胞或物种的复杂性



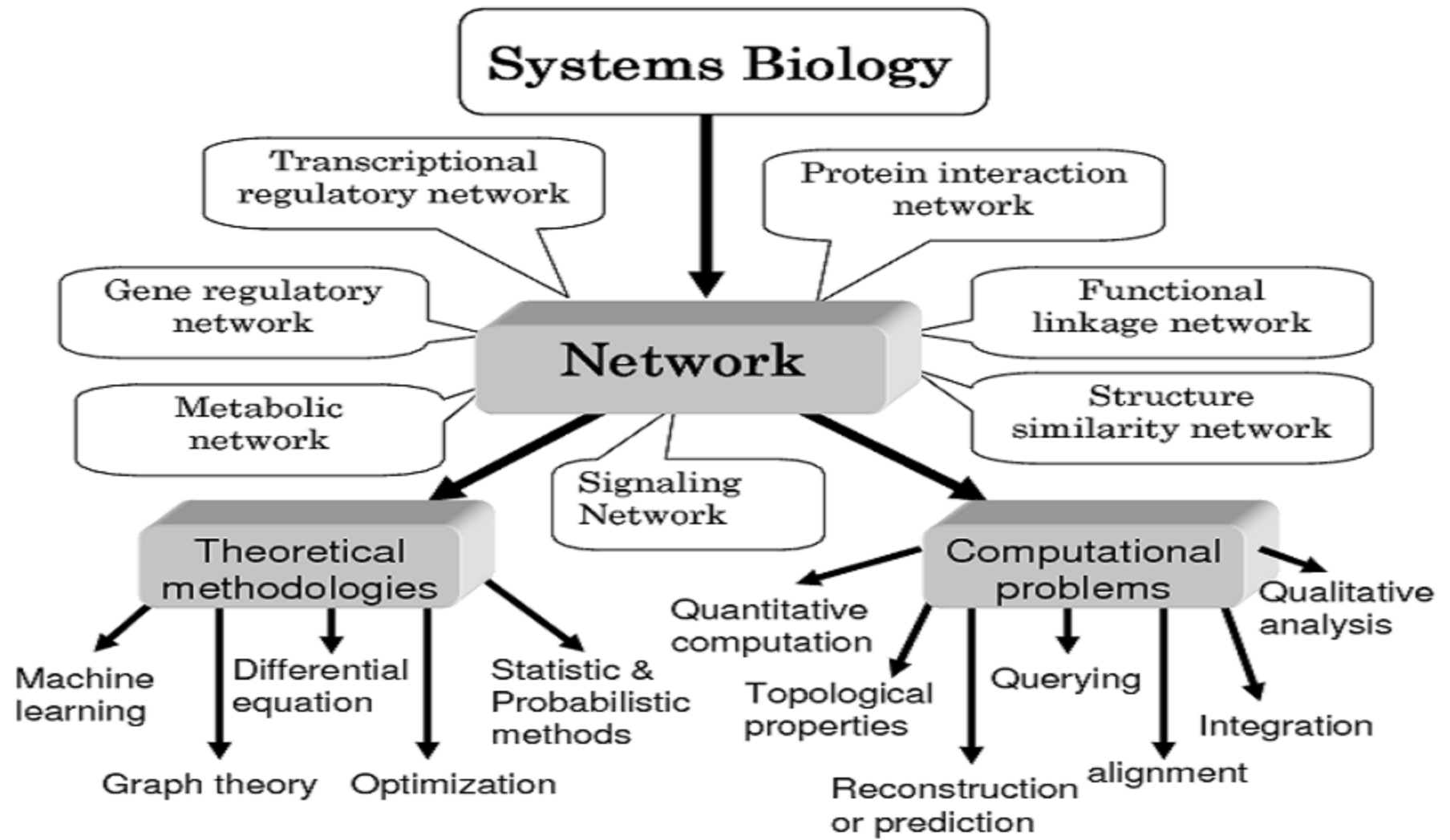


网络拓扑结构分析



Yeast protein interaction network

- Degree Distribution
- Clustering Analysis
- Local Density
- Centrality Analysis
- Connectivity Analysis
- Modularity Analysis
- Spectral Analysis
- Robustness Analysis



Network Systems Biology



模型举例



- 网络模型
- 动力学模型
- 元胞自动机模型

Quantitative Kinetic Models. The TIME



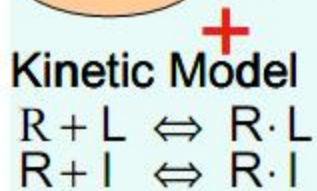
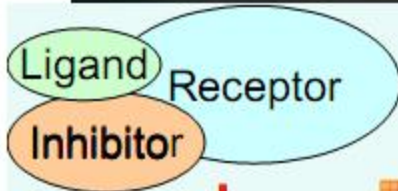
Kinetic models - time dependency incorporated

- Kinetic behaviour (rate laws) added to static model

Kinetic constants by fitting to experimental data

Mathematical model

- Time variation of all concentrations and fluxes can be simulated
- Model analyses possible: sensitivity, linear stability, bifurcation, and asymptotic analysis

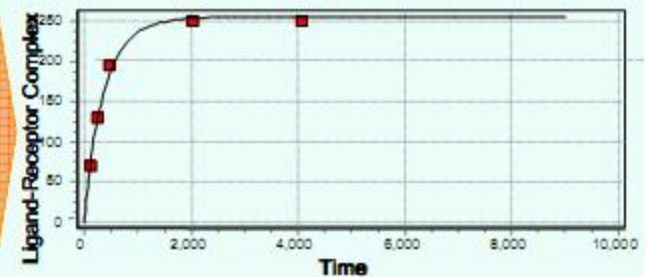


Static model

Mathematical Model

$$\begin{aligned} [R]' &= -k_1[R][L] + k_2[RL] - k_3[R][I] + k_4[RI] \\ [RL]' &= k_1[R][L] - k_2[RL] \\ [RI]' &= k_3[R][I] - k_4[RI] \\ [L]' &= -k_1[R][L] + k_2[RL] \\ [I]' &= -k_3[R][I] + k_4[RI] \\ L_0 &= [L] + [RL] \\ I_0 &= [I] + [RI] \\ R_0 &= [R] + [RL] + [RI] \end{aligned}$$

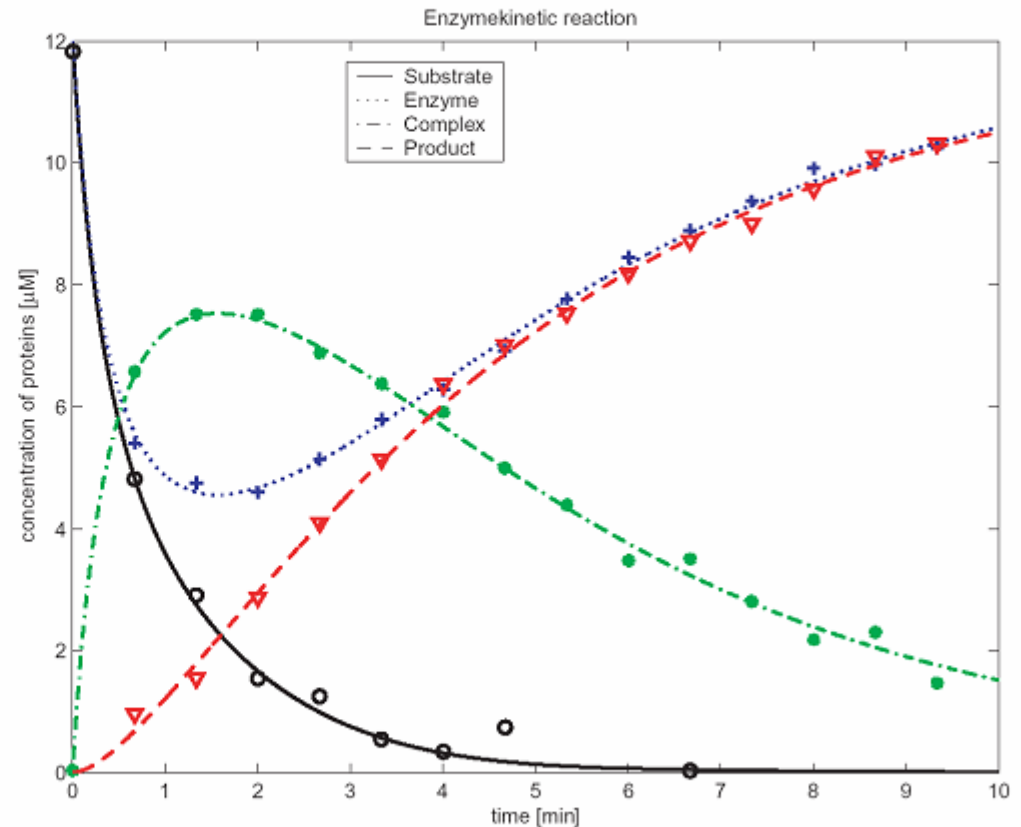
Numerical Simulation



Bioinformatics, 1999, Vol 15, 749-758,

Model Analysis. Dynamics

- Steady State
- Damped oscillations
- Triggers
- Switches
- Oscillation
- Coupled Oscillations
- Chaos
- ???



Deterministic Models

- Goal
 - model mesoscale system
 - average case behavior
- Three forms
 - ODE system
 - ODE compartment system
 - PDE (rare!)
- data limited...

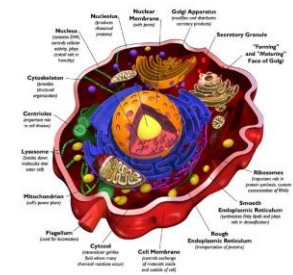
$$\frac{dC}{dt} = (\text{generation}) - (\text{consumption})$$

lumped cell



$$\frac{dC_i}{dt} = (\text{generation}) + (\text{flux in}) - (\text{consumption}) - (\text{flux out})$$

cell compartments



$$\frac{\partial \rho_D}{\partial t} = D_D \frac{\partial^2 \rho_D}{\partial x^2} - \frac{\sigma_1 \rho_D}{1 + \sigma_1' \rho_e} + \sigma_2 \rho_e \rho_d \quad (1)$$

$$\frac{\partial \rho_d}{\partial t} = \frac{\sigma_1 \rho_D}{1 + \sigma_1' \rho_e} - \sigma_2 \rho_e \rho_d \quad (2)$$

$$\frac{\partial \rho_E}{\partial t} = D_E \frac{\partial^2 \rho_E}{\partial x^2} - \sigma_3 \rho_D \rho_E + \frac{\sigma_4 \rho_e}{1 + \sigma_4' \rho_D} \quad (3)$$

$$\frac{\partial \rho_e}{\partial t} = \sigma_3 \rho_D \rho_E - \frac{\sigma_4 \rho_e}{1 + \sigma_4' \rho_D} \quad (4)$$



continuous time & space
(MinCDE oscillation)



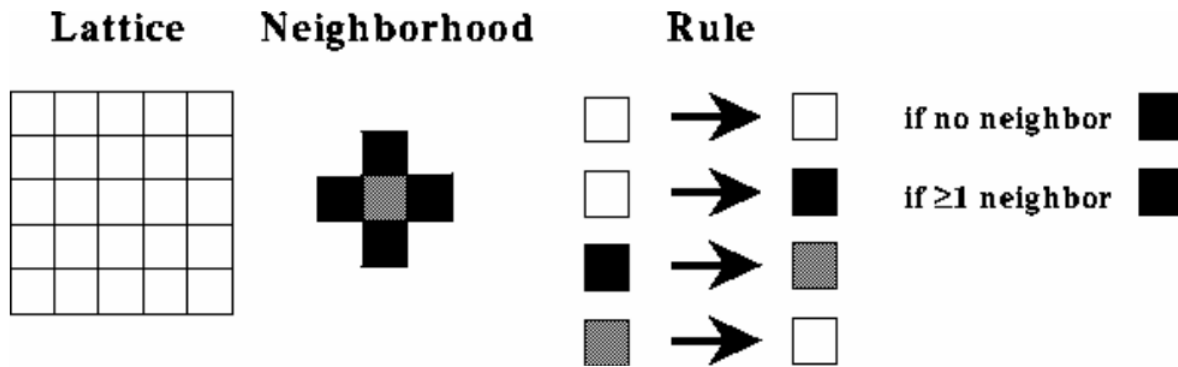


模型举例



- 网络模型
- 动力学模型
- 元胞自动机模型

Cellular Automata



Can be extended to 3D lattice



SimCell

SimCell

File Edit Run

Molecule1
Molecule2
Protein2
DNA
Protein3
Molecule2
Protein1
Membrane
Protein4

Protein Properties Card

Data Fields		Color
Name:	Protein4	[Color swatches]
# of Molecules:	10	[Color swatches]
Velocity:	3.0 mm/ms	[Color swatches]
Membrane Permeability:	Impermeable	[Color swatches]
Distribution:	Uniform	[Color swatches]
Location:	<input type="radio"/> In <input checked="" type="radio"/> Out <input type="radio"/> In-Out	[Color swatches]
Decay Rate:	0.0 mol/sec	[Color swatches]
Creation Rate:	0.0 mol/sec	[Color swatches]

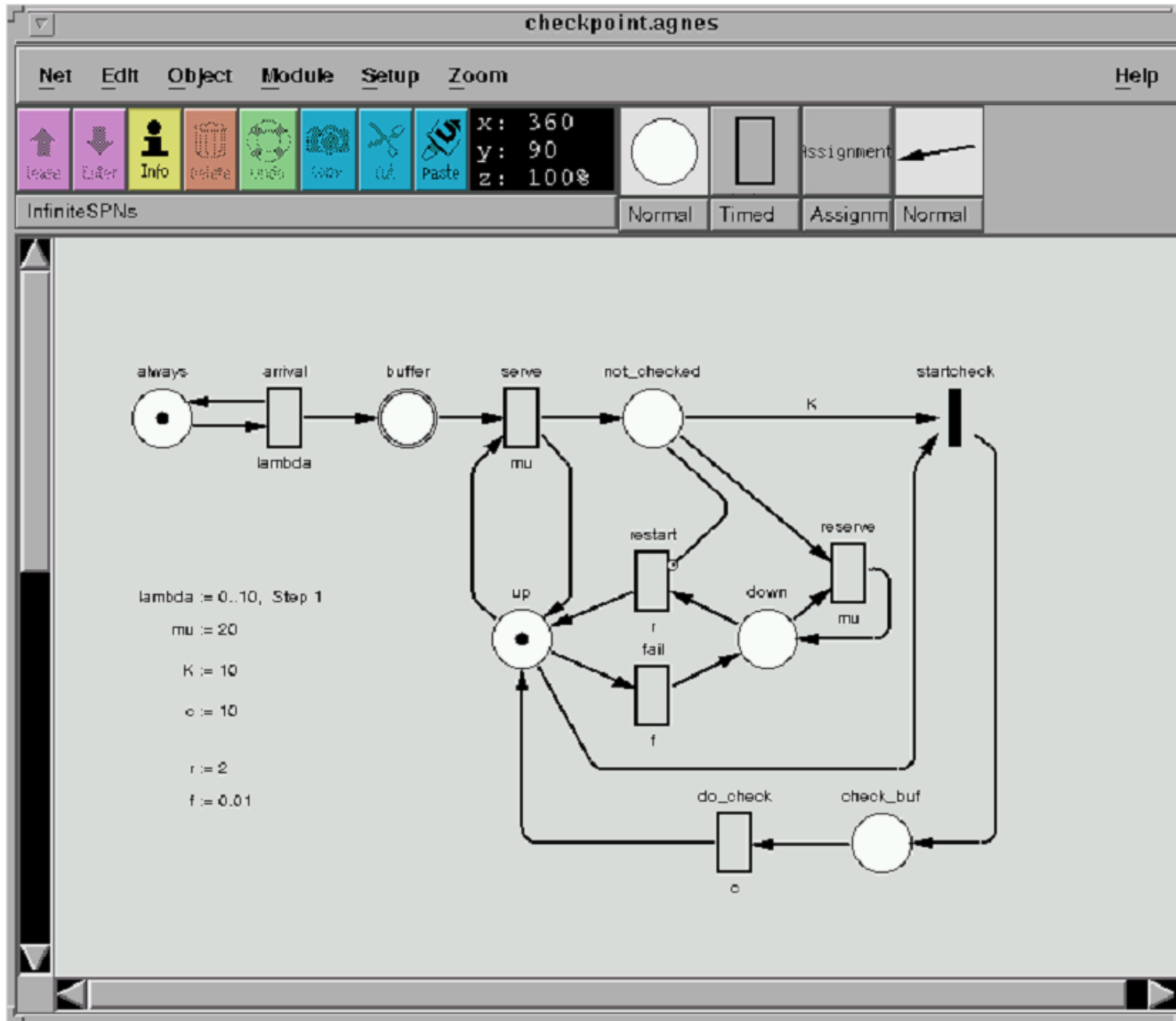
Simulation Speed

Time: 0

Run Stop Pause Quit Graphs

The image displays the SimCell software interface. The main window shows a 2D simulation of a cell with various components: Molecule1 (green), Molecule2 (blue), Protein1 (green), Protein2 (red), Protein3 (green), Protein4 (purple), and DNA (blue). A membrane is shown at the bottom. A 'Protein Properties Card' is open, showing settings for Protein4, including Name, # of Molecules (10), Velocity (3.0 mm/ms), Membrane Permeability (Impermeable), Distribution (Uniform), Location (Out), Decay Rate (0.0 mol/sec), and Creation Rate (0.0 mol/sec). A 'Simulation Speed' window shows a graph of simulation speed over time, with values fluctuating between 0 and 1000. The bottom window shows a grid view of the simulation with a 'Time: 0' label and buttons for Run, Stop, Pause, Quit, and Graphs.

Petri Nets

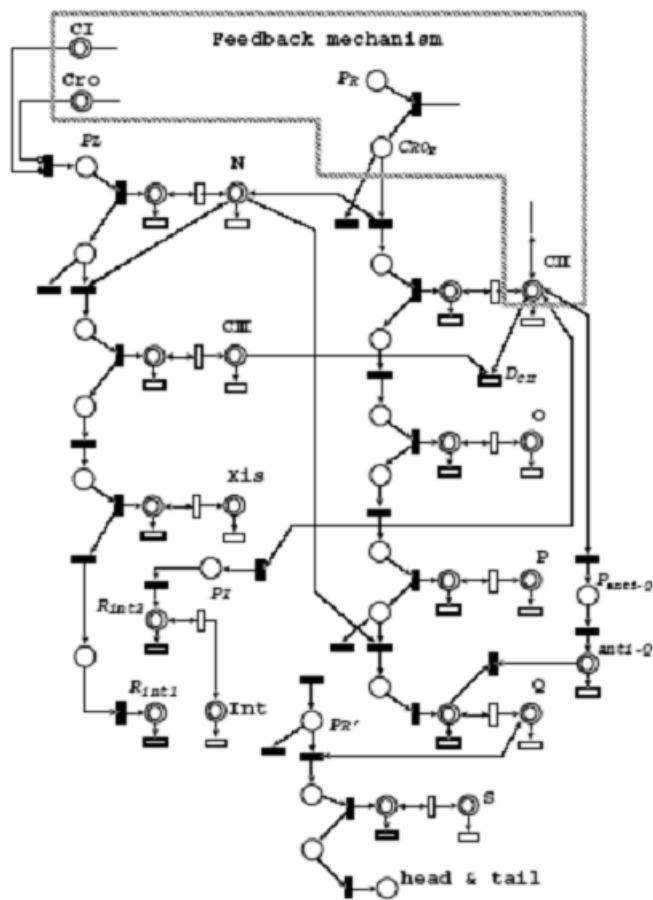


Petri Nets

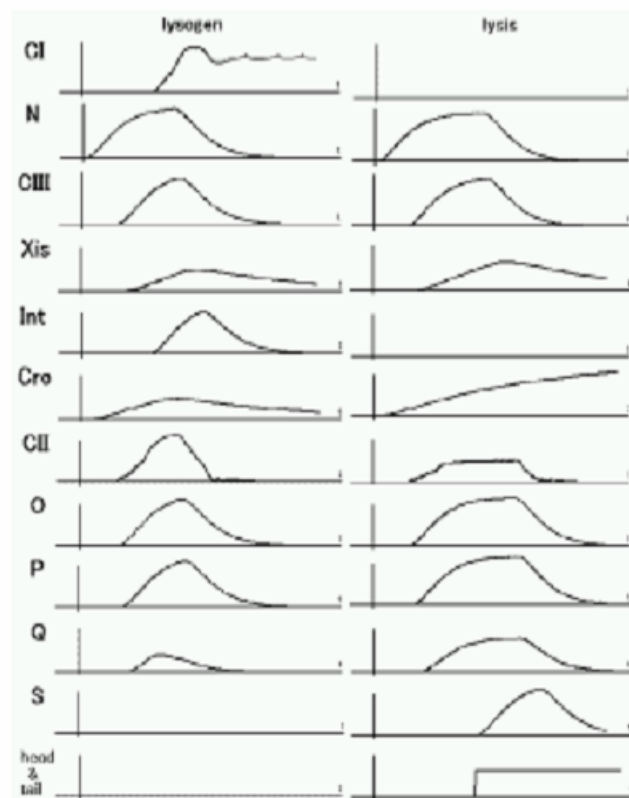
- A directed, bipartite graph in which nodes are either "places" (circles) or "transitions" (rectangles)
- A Petri net is marked by placing "tokens" on linked or connected places
- When all the places have a token, the transition "fires", removing a token from each input place and adding a token to each place pointed to by the transition (its output places)
- Petri nets are used to model concurrent systems, particularly network protocols w/o differential eqs.
- Hybrid petri nets allow modelling of continuous and discrete phenomena



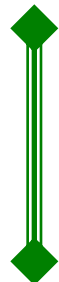
Hybrid Petri Nets



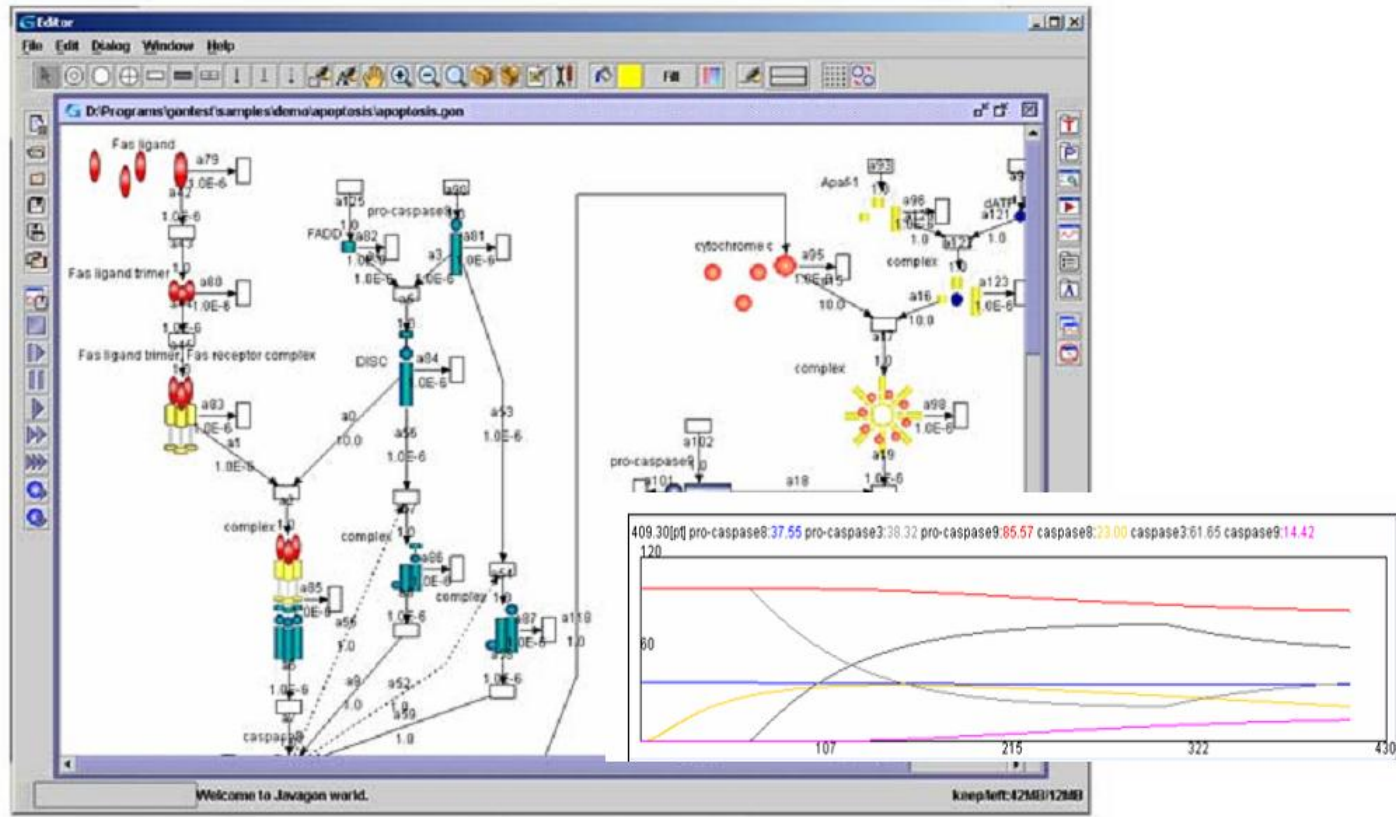
λ phage control circuit



Predicted protein expression



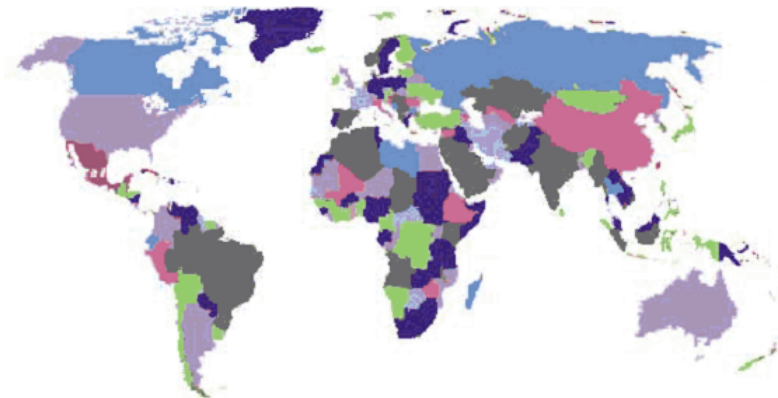
Cell Illustrator – An HPN with a GUI www.genomeobject.net



Now sold as a product by Gene Networks International - <http://www.gene-networks.com>

Nationalism in Simulation

- **Petri Nets – Germany, Japan**
- **Flux-Balance Analysis – USA**
- **Pi Calculus – France**
- **ODE's and PDE's – Japan, UK**
- **Agent-Based methods (CA) - Canada**



Cell: The Unit of Life

- **Reproduction** by cell division
- **Metabolism**, including **taking in raw materials, building cell components, converting energy, molecules and releasing by-products**
- **Response to external and internal stimuli** such as changes in **temperature, pH or levels of nutrients**
- **Growth, development, and maintenance of homeostasis,...**

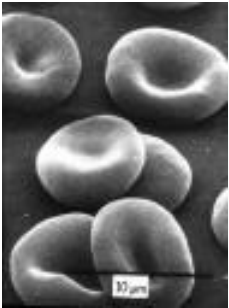




Successful Models

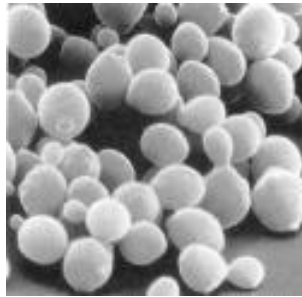


Red Blood Cell



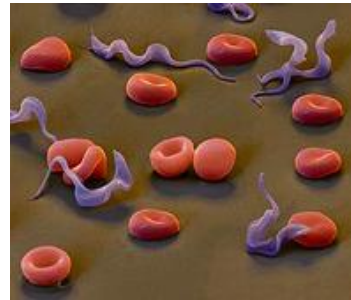
Mulquiney, Joshi, Heinrich, ...

Yeast Glycolysis



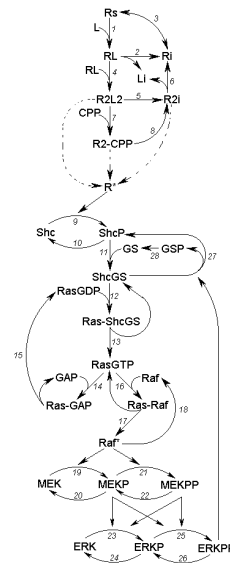
Bas Teusink

Trypanosoma Brucei



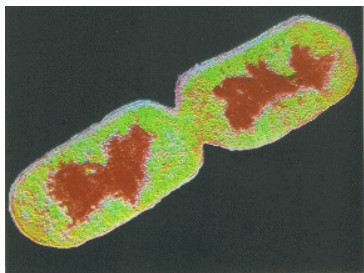
Barbara Bakker, Westerhoff and Cornish-Bowden

EGF Signaling Pathway



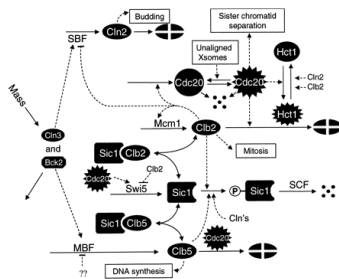
Frances Brightman et al

Chemotaxis, ecoli



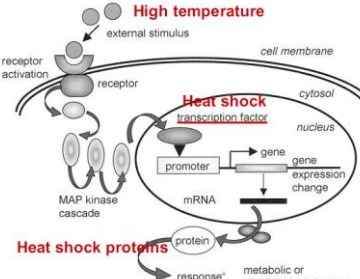
Many Contributors

Yeast Cell Cycle



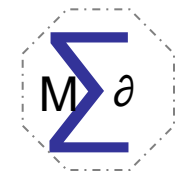
John Tyson et al

Yeast Heat Shock Response



Many Contributors

- Biochemical network
- HSP12
 - HSP26
 - HSP42
 - HSP60
 - HSP78
 - HSP82
 - SSA1
 - SSA2
 - SSA3
 - SSA4



纲要



- 系统与生物系统
- 模型与建模
- 系统生物学中的可计算模型举例
- 复杂性 with 系统分析



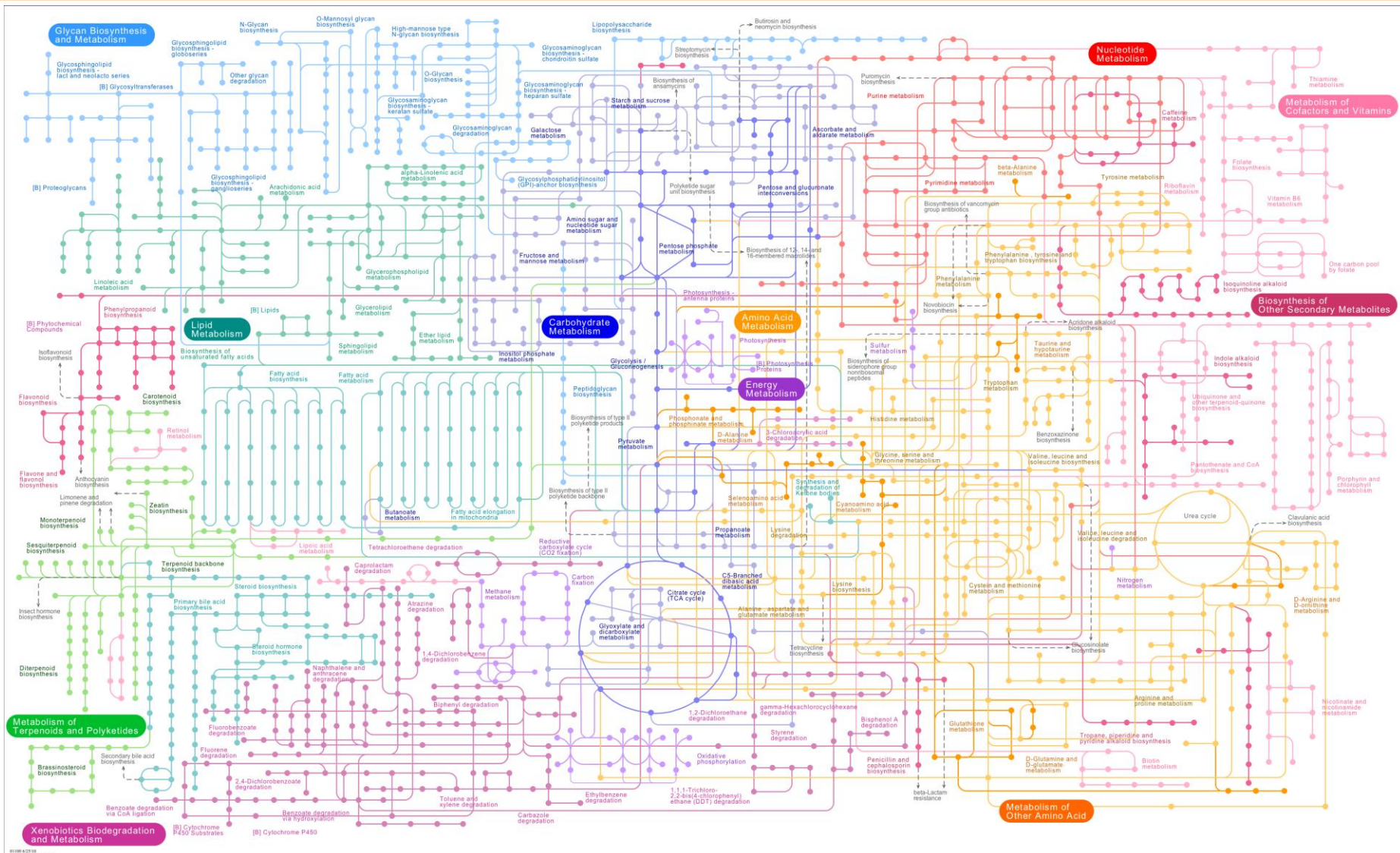
复杂性



- 据有人统计，迄今，各学科中关于复杂性的定义，已超过50种之多。
- 什么是复杂性：对一个系统的最短的描述也很长。
- 复杂性的来源：系统中大量多种类型的组分的非线性相互作用。
- 复杂性的表现：网络、聚类、分层、嵌套、反馈等。



复杂的代谢系统

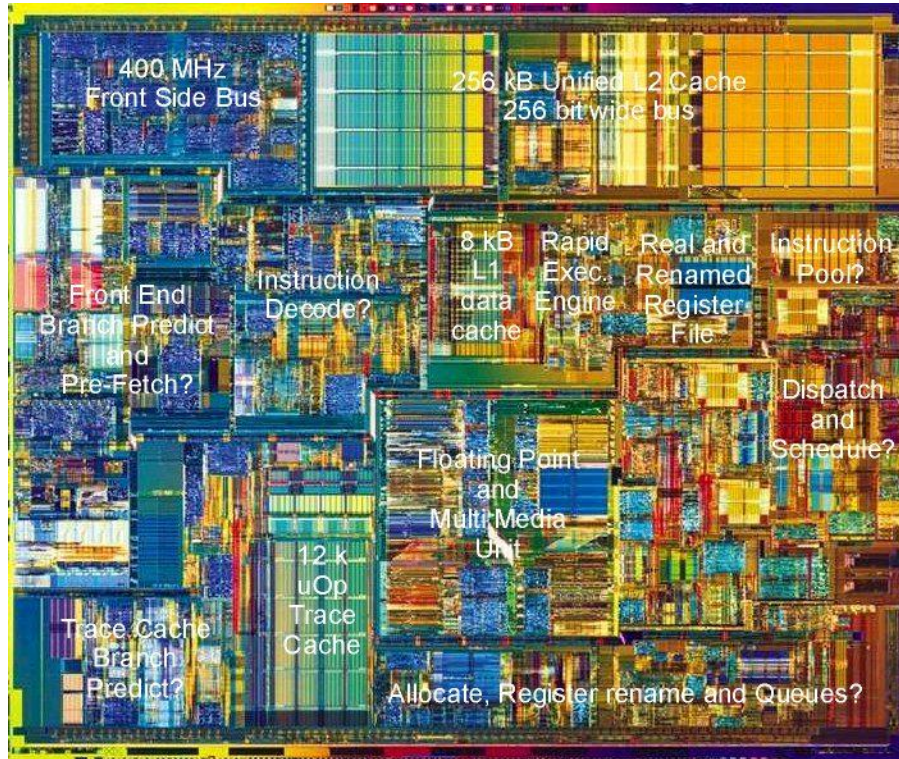




如何处理复杂性



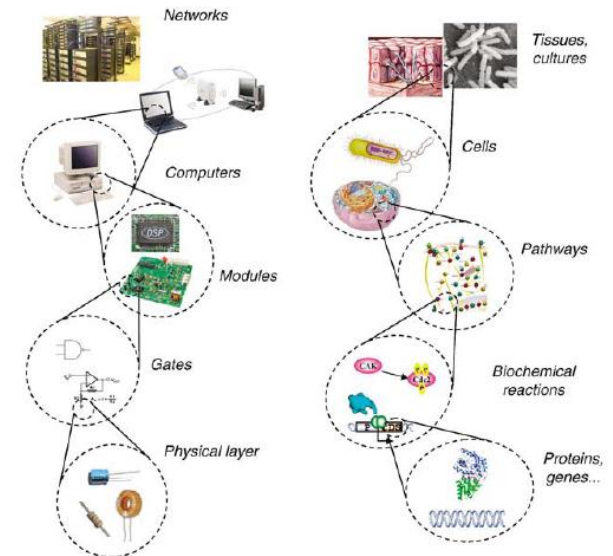
Man-Made Complexity



Intel Pentium 4
42 million transistors

原则:

- 抽象（分层设计）
- 模块化（接口设计）





系统水平的分析



- 稳定性分析 (Stability)
- 分叉分析 (Bifurcation)
- 鲁棒性分析 (Robustness)
- 反馈分析 (Feedback, \pm)
- 流 (平衡) 分析 (Flux, FBA)

