系统生物学

陈 铭
浙江大学生命科学学院
2025年10月21日

本章内容

- Systems Biology Overview
- Network Representation
- Reconstruction of Biological Network
- Topological Analysis
- ODE & FBA
- ~~Petri net Modeling~~
- Synthetic Biology
- Bioinformatics Engineering

1 Systems Biology Overview

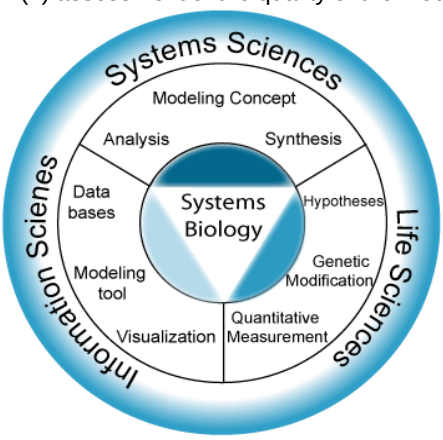
The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves:

(1) collection of large sets of experimental data 数据

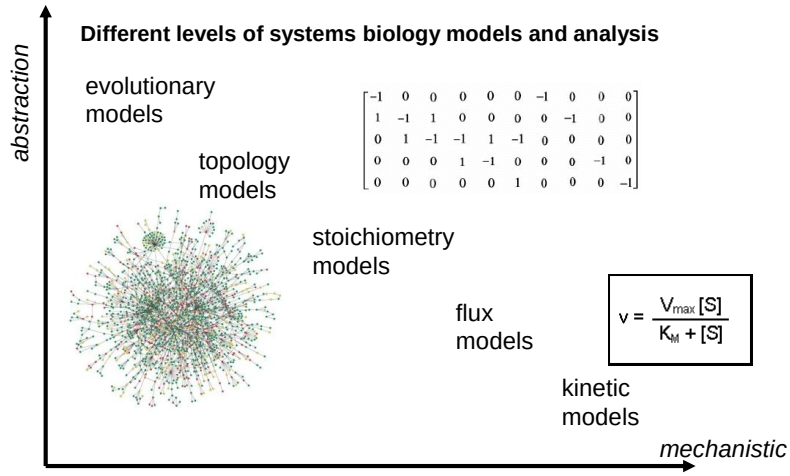
(2) proposal of mathematical models that might account for at least some significant aspects of this data set 模型

(3) accurate computer solution of the mathemcatical equations to obtain numerical predictions, 模拟

(4) assessment of the quality of the model by comparing numberical simulations with the experimental data. 验证



Leroy Hood, 1999



Systems Biology Historical Perspective

- Systems Theory

General Systems Theory
Karl Ludwig von Bertalanffy
1930s

Cybernetics
Norbert Wiener
1948

Biochemical systems theory
Michael Savageau
1960s

Metabolic control analysis
1970s

Hypercycles
Manfred Eigen
1979

2000
Human genome sequence

1869
Isolation of DNA

1953
Double-helix structure of DNA

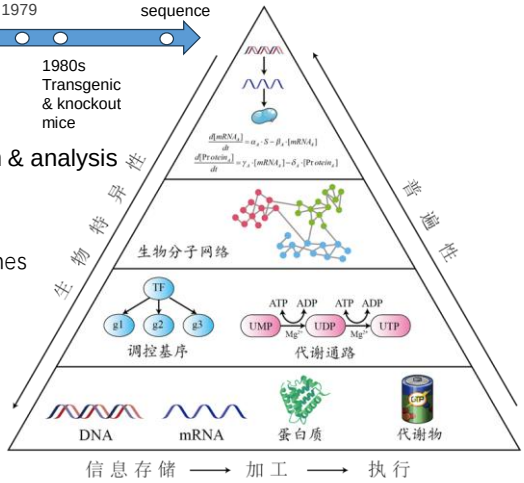
1980s
Transgenic & knockout mice

Early work suffered from inadequate data

- Molecular biology

Further advances require data integration & analysis

- Systems biology
  - Represents integration of the systems & molecular approaches
  - Motivated by need to relate genotype → phenotype
  - Enabled by high throughput measurement technologies & advances in computer hardware & algorithms



Systems Biology Definitions

Leroy Hood

- As global a view as possible
- Fundamentally quantitative
- Different scales integrated

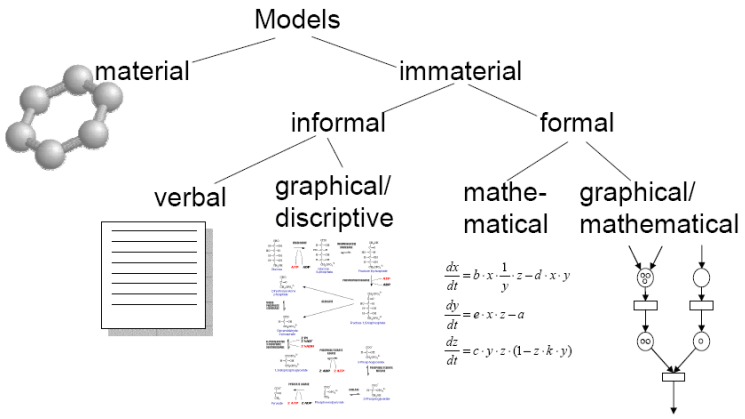
Hiroaki Kitano

- „Aims at systems-level understanding [which] requires a set of principles and methodologies that links the behaviors of molecules to systems characteristics and functions“

- **Data Integration**
  - Gene, mRNA, protein, small biological molecules, etc.
  - Different levels: from gene to cell, tissue, individual…
  - Different methods or approaches
  - Interdisciplinary work
- **Modeling and Simulation**
- **Dynamic Behavior Prediction**

Model Classification/types of models

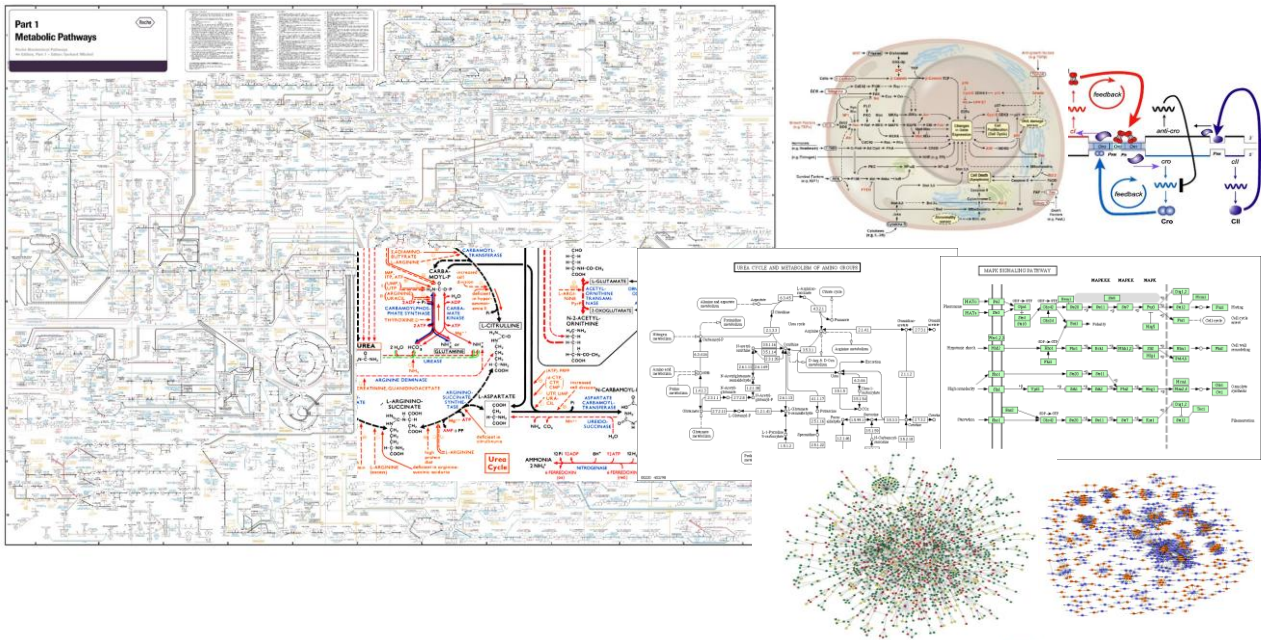
1. Conceptual or verbal - descriptions in a natural language.
2. Diagrammatic - graphical representations of the objects and relations (e.g., physiological diagrams of metabolic pathways such as the Krebs cycle).
3. Physical - a real, physical mock-up of a real system or object (a "tinker-toy" model of DNA or 3D structure of protein).
4. Formal - mathematical (usually using algebraic or differential equations).



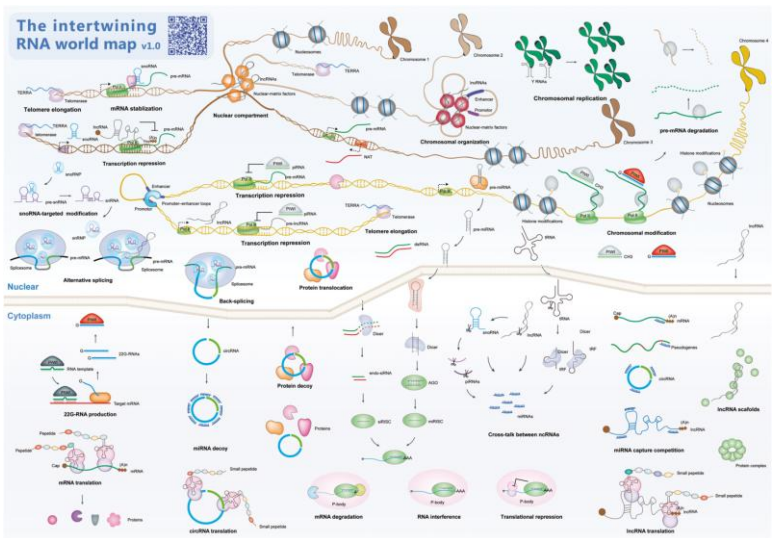
Wolfgang Prange, 2004

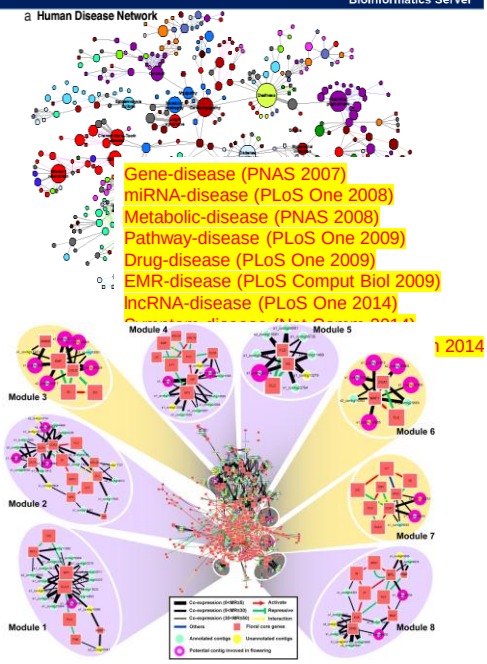
2 Network Representation

Network Representation



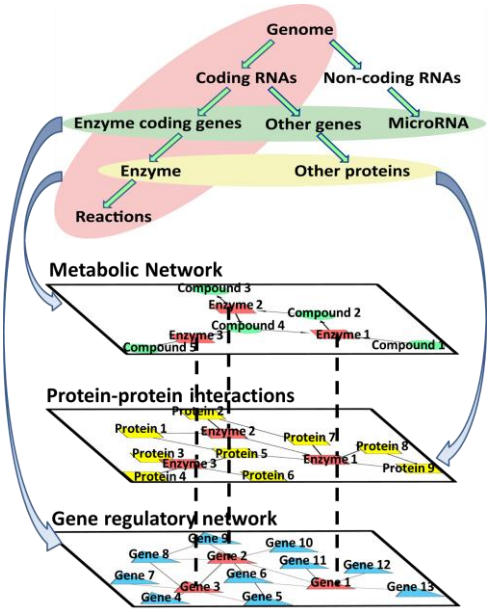
Network Representation

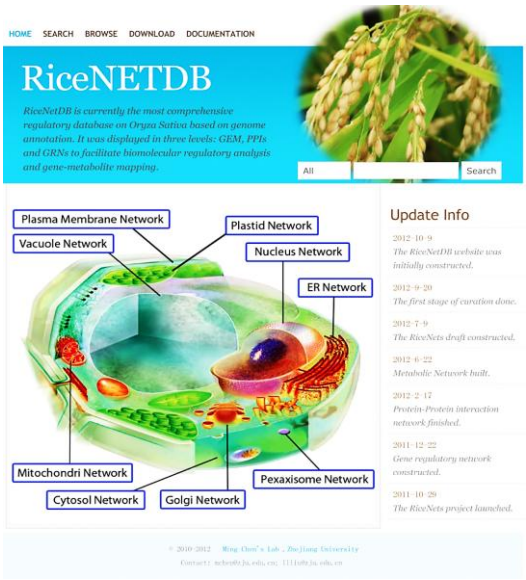




Gene-disease (PNAS 2007)
miRNA-disease (PLoS One 2008)
Metabolic-disease (PNAS 2008)
Pathway-disease (PLoS One 2009)
Drug-disease (PLoS One 2009)
EMR-disease (PLoS Comput Biol 2009)
lncRNA-disease (PLoS One 2014)

Network Representation

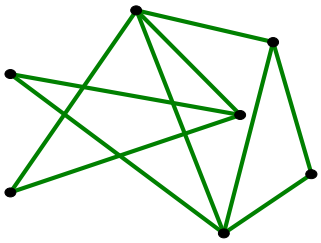




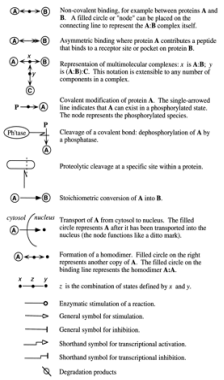
Pathway presentation/graphical notations

- Network
  - A linked list of interconnected nodes.
  - Node
    - Gene, protein, peptide, other biomolecules.
  - Edges
    - Biological relationships, etc., interactions, regulations, reactions, transformations, activation, inhibitions.

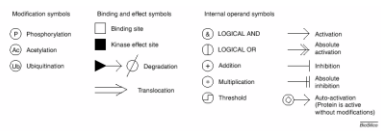
Graph = (Vertices, Edges)



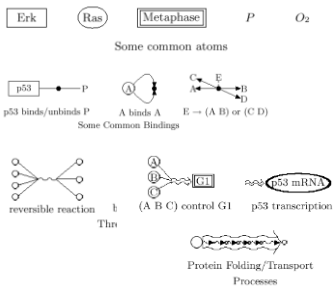
Kohn K.W. (1999). Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. *Mol. Biol.Cell. 10, 2703-2734.*



Kitano H. (2003). A graphical notation for biochemical networks. *BIOSILICO Vol. 1. No. 5.*

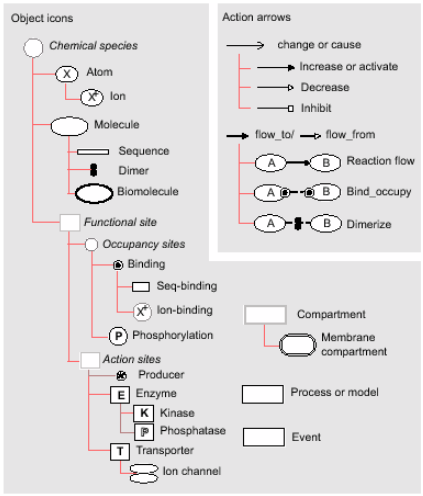


R. Maimon and S. Browning (2001). Diagrammatic Notation and Computational Grammar for Gene Networks. *Proceedings of the International Conference on Systems Biology.* 2001.

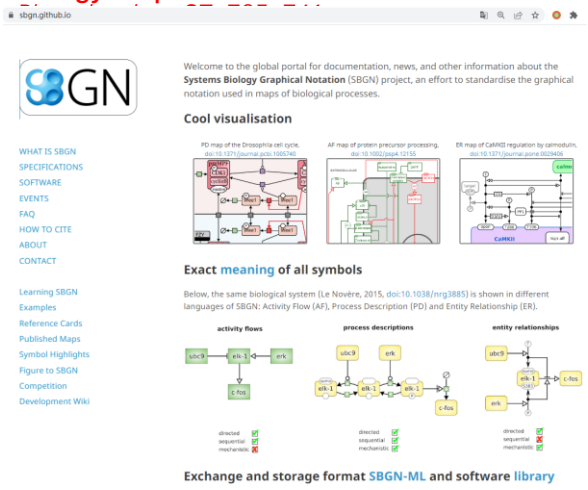


Some graphical notations for biopathways

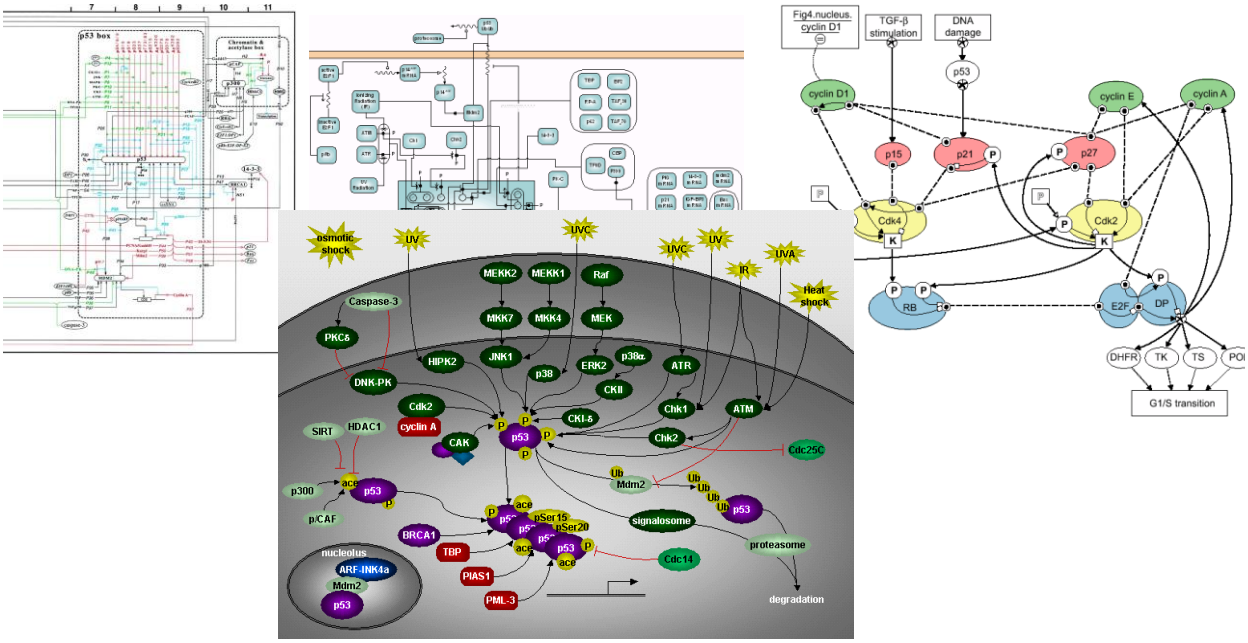
Cook D.L. et al. (2001). A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biol. 2. RESEARCH 0012.*



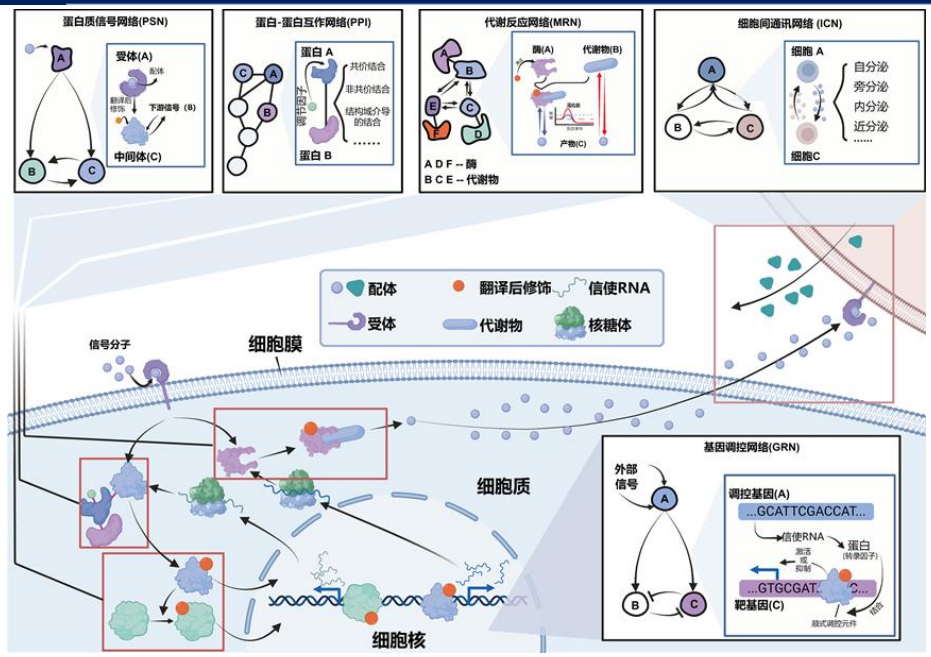
- **Database specific notations**. KEGG/Metabolic pathways; GeneNet; TRANSPATH …
- Nicolas Le Novère (2009). **The Systems Biology Graphical Notation.** *Nature*



Networks: p53, The gatekeeper of death



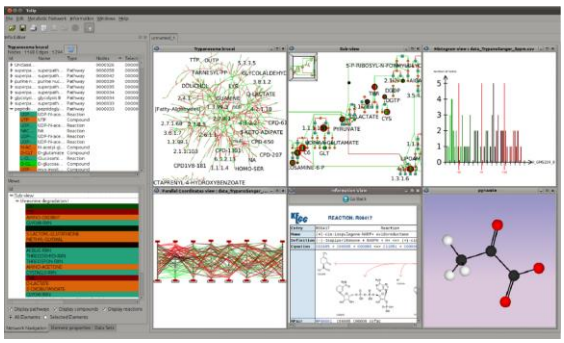
Integrated Networks



Network Visualization Tools and Layouts

- Cytoscape
  - plugins
- GraphViz
- Systrip
- Vanted
- Jdesigner
- CellDesigner
- Cell Illustrator
- iPATH2
- …

Overview of Systrip a visual environment for the analysis of time-series data in the context of biological networks.



Network layout（网络布局）：

1. Basic Structural (基础结构)
圆形 → Circular
力导向 → Force-Directed
簇 → Cluster
星型 → Star
网格 → Grid
二分图 → Bipartite

2. Hierarchical & Radial (层次与径向)
层次 → Hierarchical
Subtype: 树状 → Tree
径向 → Radial

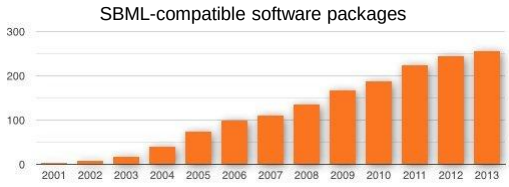
3. Matrix & Tabular (矩阵与表格)
矩阵 → Matrix

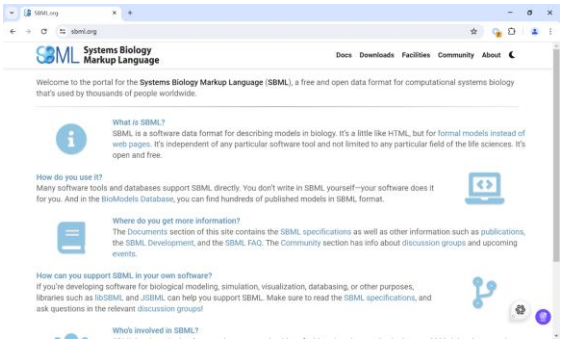
4. Flow & Volume (流向与流量)
流向（针对有向无环图） → Flow (DAG-Specific)
桑基图（流量） → Sankey Diagram (Flow Volume)
弦图（双向流量） → Chord Diagram

5. Spatial & Geographic (空间与地理)
地理 → Geographic

Standardisation

- BioPAX: Biological Pathways Exchange
  - Tools: Cytoscape, VisANT, BioUML, …
- PSI-MI: Proteomics Standards Initiative-Molecualr Interactio
  - Databases: BIND, DIP, HPRD, IntAct, MINT
- SBML: **Systems Biology Markup Language**
- CellML: language for specifying and exchanging biological processes
- SBGN: http://www.sbgn.org/





SBML: Systems Biology Markup Language

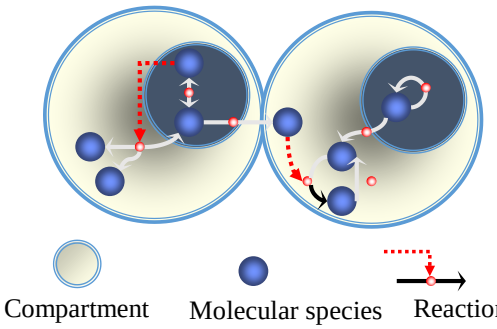
SBML Systems Biology Markup Language

无法在不同的模拟和分析工具之间交换模型，其根源在于缺乏描述模型的通用格式。因此研究者们决定开发一种简单的基于xml的语言，用于在模拟/分析工具之间表示和交换模型:系统生物学标记语言(SBML)。

为了组织架构变更和版本化，SBML的开发是分层的。SBML的主要版本被称为level，代表了语言的组成和结构的实质性变化。SBML级别是共存的。例如，SBML level3不会使level 2过时，并且仍然继续使用2级兼容的型号和软件工具。

What Kind of Models Can You Express in SBML?

- Focus: systems of biochemical reactions
- Models can also include:
  - Compartments
  - Rules/constraints
  - Discrete events



SBML的语法结构

SBML的结构包括大量的组件：reactants species,product species,reactions,rate laws,以及在rate laws中的参数。为了能够分析或者模拟这些网络，还要添加一些显性的组件，包括：species的compartments和各种量的units。SBML顶层的组件由这些组件组成。

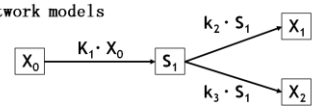
beginning of model definition
- *list of function definitions (optional)*
- *list of unit definitions (optional)*
- *list of compartments (optional)*
- *list of species (optional)*
- *list of parameters (optional)*
- *list of initial assignments (optional)*
- *list of rules (optional)*
- *list of constraints (optional)*
- *list of reactions (optional)*
- *list of events (optional)*

end of model definition

浙江大学 ZHEJIANG UNIVERSITY SBML
BiS @ZJU Bioinformatics Server

Systems Biology Markup Language (SBML)

- Biochemical network models



- A model is described using a list of components:
  - Beginning of model definition
    - List of unit definitions (optional)
    - List of compartments
    - List of species
    - List of parameters (optional)
    - List of rules (optional)
    - List of reactions
  - End of model definition

Example (cont.)

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="1" version="1">
 <model name="simple">
  <listOfCompartments>
   <compartment name="c1" />
  </listOfCompartments>
  <listOfSpecies>
   <specie name="X0" compartment="c1"
                boundaryCondition="true"
            initialAmount="1"/>
   <specie name="S1" compartment="c1"
                boundaryCondition="false"
            initialAmount="0"/>
   <specie name="X1" compartment="c1"
                boundaryCondition="true"
            initialAmount="0"/>
   <specie name="X2" compartment="c1"
                boundaryCondition="true"
            initialAmount="0.23"/>
  </listOfSpecies>
```

Example (cont.)

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="1" version="1">
 <model name="simple">
  <listOfCompartments>
   <compartment name="c1" />
  </listOfCompartments>
  <listOfSpecies>
   <specie name="X0" compartment="c1"
                boundaryCondition="true"
            initialAmount="1"/>
   <specie name="S1" compartment="c1"
                boundaryCondition="false"
            initialAmount="0"/>
   <specie name="X1" compartment="c1"
                boundaryCondition="true"
            initialAmount="0"/>
   <specie name="X2" compartment="c1"
                boundaryCondition="true"
            initialAmount="0.23"/>
  </listOfSpecies>
```

9

Example (cont.)

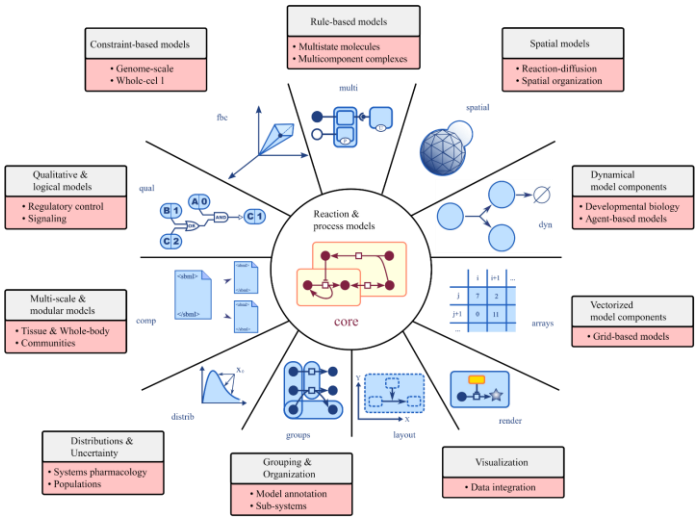
```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="1" version="1">
 <model name="simple">
  <listOfCompartments>
   <compartment name="c1" />
  </listOfCompartments>
  <listOfSpecies>
   <specie name="X0" compartment="c1"
                boundaryCondition="true"
            initialAmount="1"/>
   <specie name="S1" compartment="c1"
                boundaryCondition="false"
            initialAmount="0"/>
   <specie name="X1" compartment="c1"
                boundaryCondition="true"
            initialAmount="0"/>
   <specie name="X2" compartment="c1"
                boundaryCondition="true"
            initialAmount="0.23"/>
  </listOfSpecies>
```

4

Example (cont.)

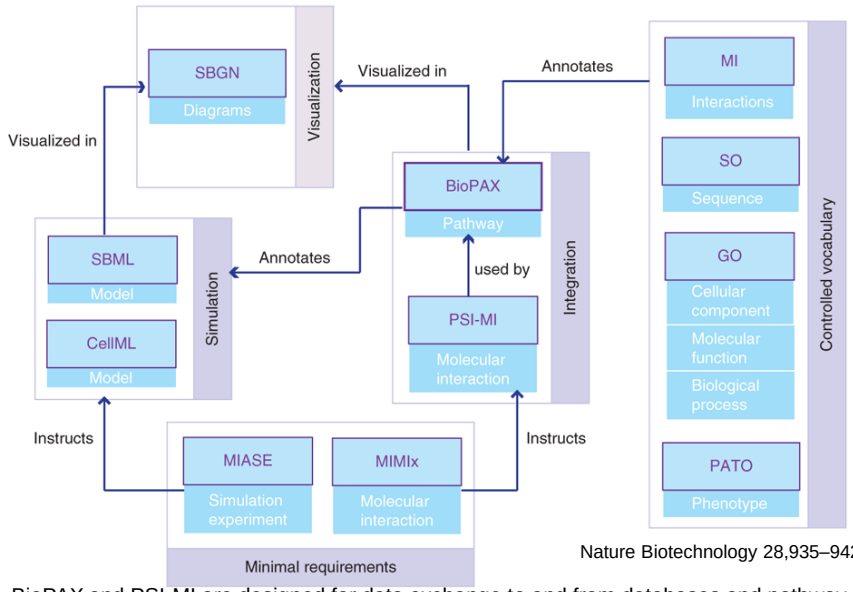
```
<listOfReactions>
 <reaction name="reaction_1" reversible="false">
  <listOfReactants>
   <specieReference specie="X0" stoichiometry="1"/>
  </listOfReactants>
  <listOfProducts>
   <specieReference specie= S1  stoichiometry="1"/>
  </listOfProducs>
  <kineticLaw formula="k1 * X0">
   <listOfParameters>
    <parameter name="k1" value="0"/>
   </listOfParameters>
  </kineticLaw>
 </reaction>

 <reaction name="reaction_2" reversible="false">
  <listOfReactants>
   <specieReference specie="S1" stoichiometry="1"/>
  </listOfReactants>
  . . .
 ..
```

Application Support for SBML



- DARPA BioSPICE uses SBML as the model definition language
- Applications supporting SBML:
  - BASIS — UK
  - Bio Sketch Pad — BBN
  - CellDesigner — ERATO Kitano Symbiotic Systems Project
  - Cellerator — NASA JPL & University of California Irvine
  - Cytoscape — Institute for Systems Biology & MIT
  - Gepasi — Virginia Tech
  - Jarnac — Keck Graduate Institute
  - JDesigner — Keck Graduate Institute
  - JigCell — Virginia Tech
  - NetBuilder — University of Hertfordshire
  - SigPath — Mount Sinai
  - StochSim — Cambridge University
  - Virtual Cell — University of Connecticut Health Center
  - WinSCAMP — Keck Graduate Institute
- Others coming: E-CELL, Dbsolve, GSK PWF, more

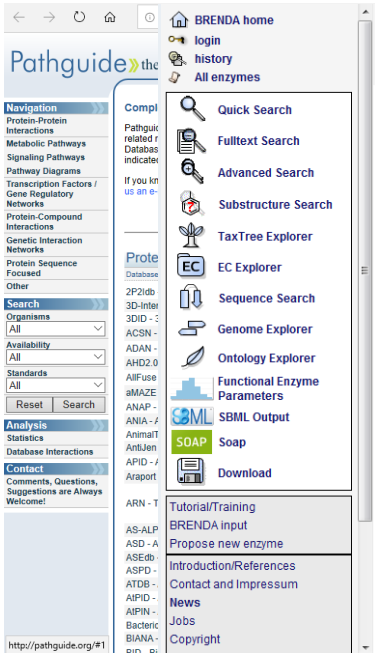
Standardisation

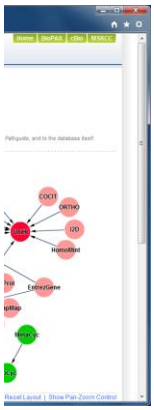


Nature Biotechnology 28,935–942(2010)

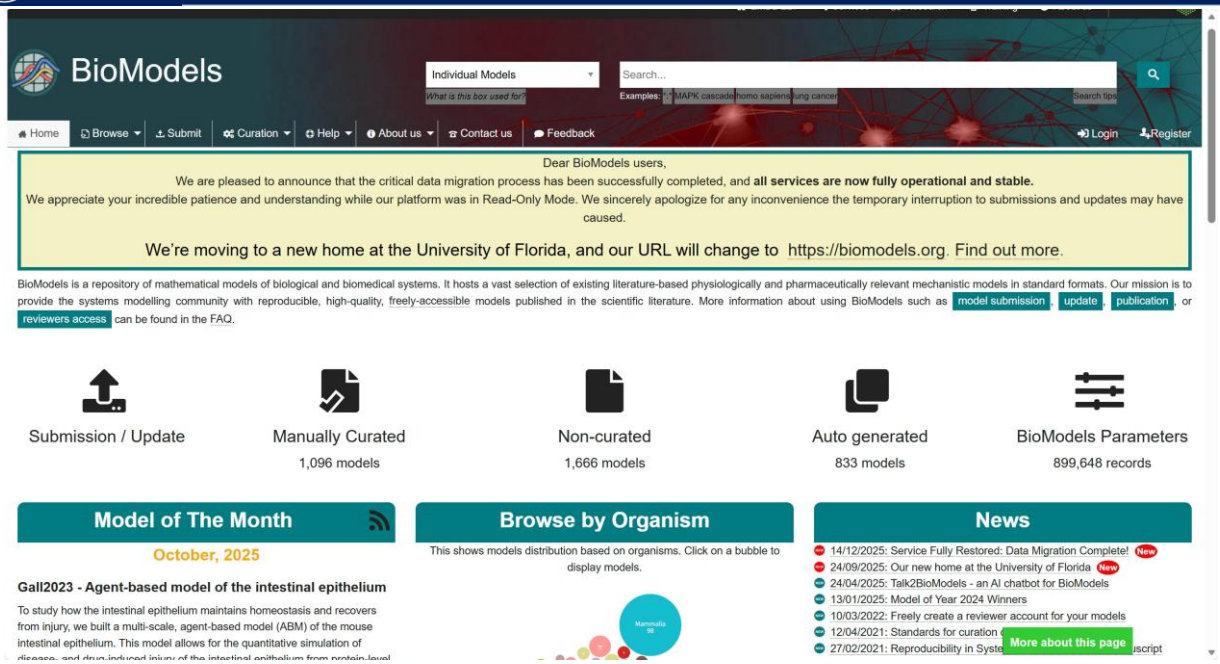
BioPAX and PSI-MI are designed for data exchange to and from databases and pathway and network data integration. SBML and CellML are designed to support mathematical simulations of biological systems and SBGN represents pathway diagrams.

BioPathways/BioNetworks Resources



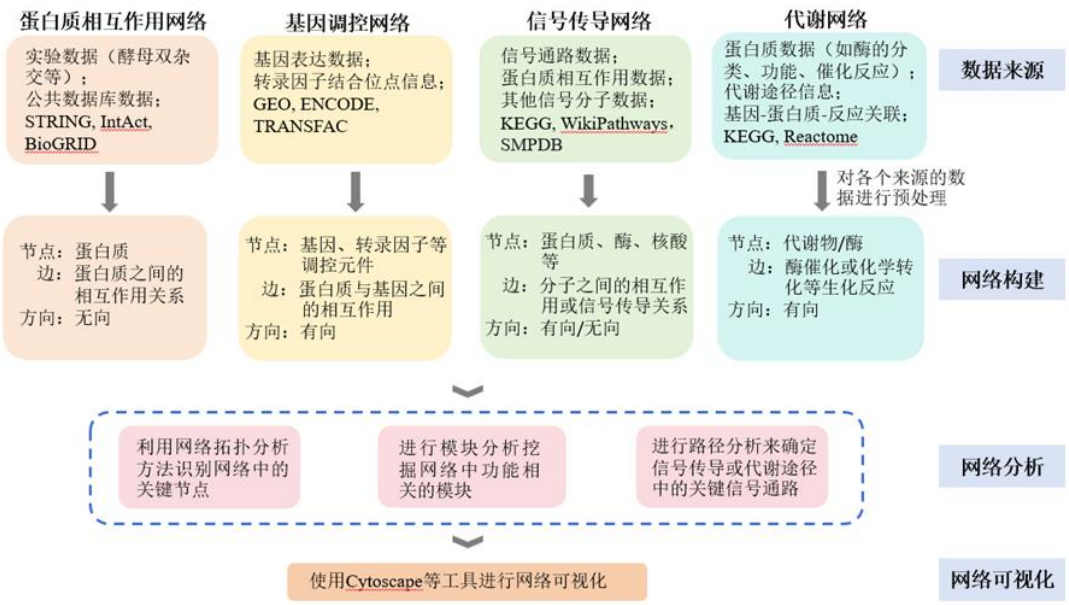


BioModels

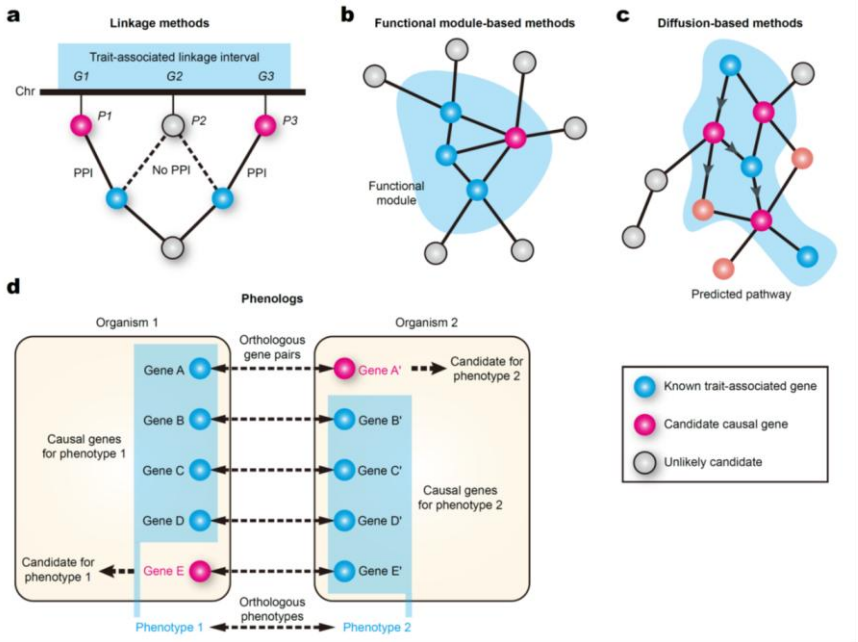


3 Reconstruction of Biological Network

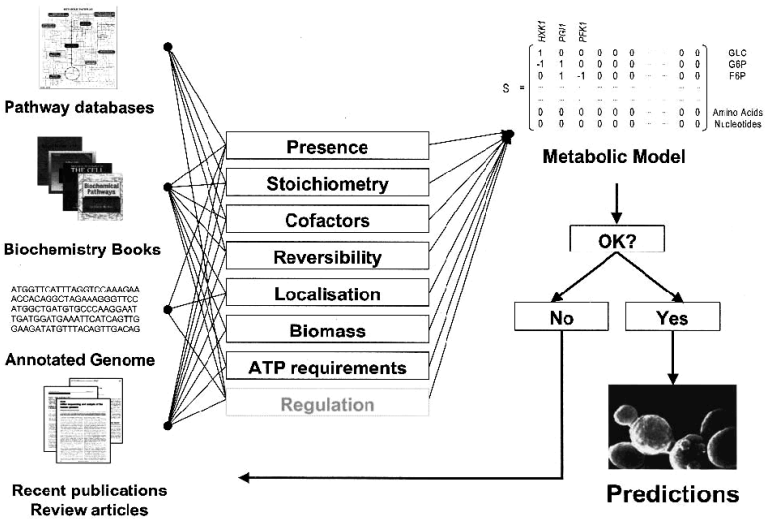
Pipelines for biological network reconstruction



FBA application



Genome-scale Metabolic Network Reconstruction



- Reconstruction of Amino Acid Biosynthesis Pathways from the Complete Genome Sequence. Bono, *Genome Res.* 1998
- A Framework of Automated Reconstruction of Microbial Metabolic Pathways. Bansal, *bibe*, 2000
- Reconstruction of Metabolic Pathways by the Exploration of Gene Expression Data with Factor Analysis. David A. Henderson, Dissertation, 2001
- Genome-scale Reconstruction of the *S. cerevisiae* Metabolic Network. Foerster, *Genome Research*, 2002
- PathFinder: reconstruction and dynamic visualization of metabolic pathways. Goesmann, *Bioinformatics*, 2002

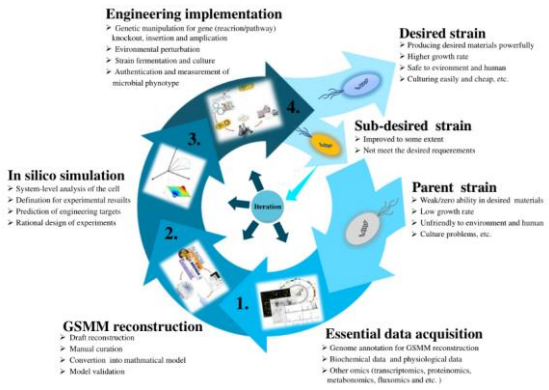
Genome-scale Metabolic Network Reconstruction

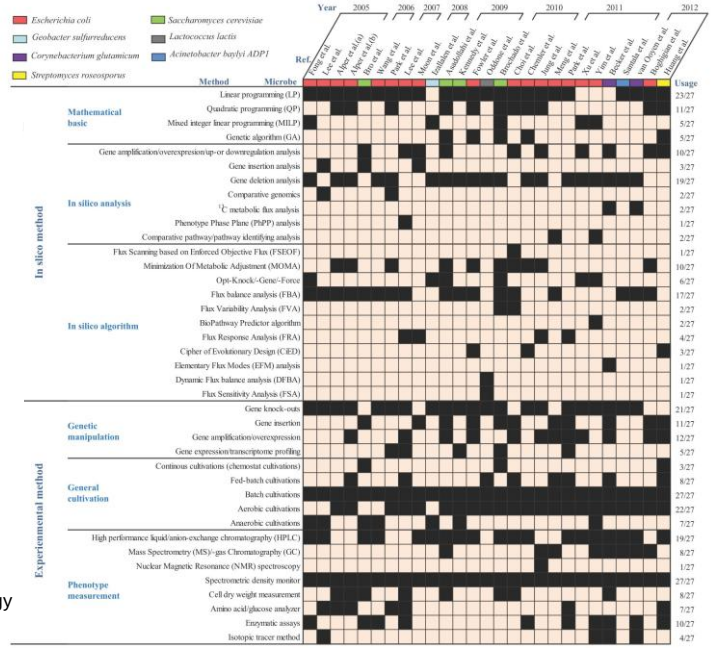
Since 2003

- Metabolic Pathway Reconstruction for Malaria Parasite *Plasmodium falciparum* → Limviphuvadh, *Genome Informatics*, 2003
- *Ab initio* reconstruction of metabolic pathways →Boyer, *Bioinformatics*, 2003
- Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms → Ma, *Bioinformatics*, 2003
- Automated metabolic reconstruction for *Methanococcus jannaschii* → Tsoka, *Archaea*, 2004
- Reconstruction of regulatory and metabolic pathways in metal-reducing δ-proteobacteria → Rodionov, *Genome Biology,* 2004
- Computational prediction of human metabolic pathways from the complete human genome → Romero, *Genome Biology*, 2004
- …

Genome-scale Metabolic Network Reconstruction

重建GMSS网络流程

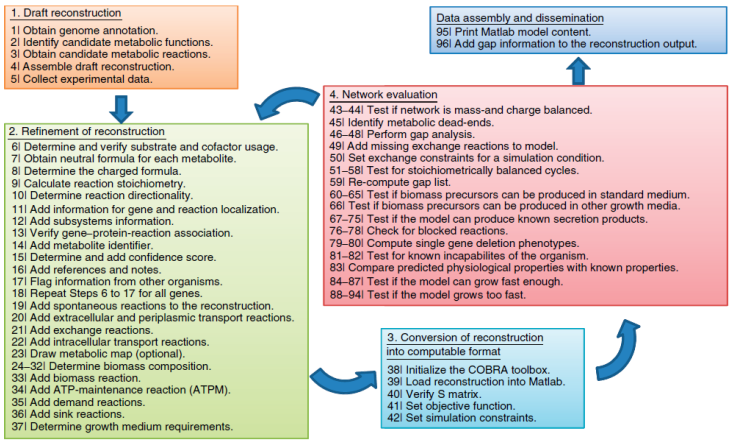




Xu et al. 2012
Applied Microbiology and Biotechnology

Genome-scale Metabolic Network Reconstruction

A protocol for generating a high-quality genome-scale metabolic reconstruction



Reconstruction of biochemical networks in microorganisms

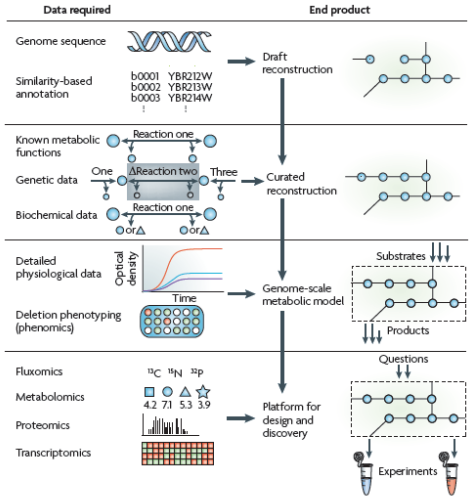
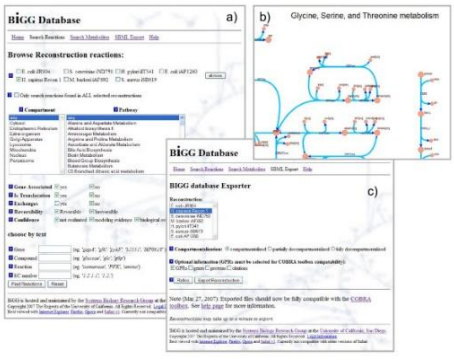


Figure 1 | **Phases and data used to generate a metabolic reconstruction.**

Feist et al. Nature Reviews Microbiology, 2009, 7:129-143

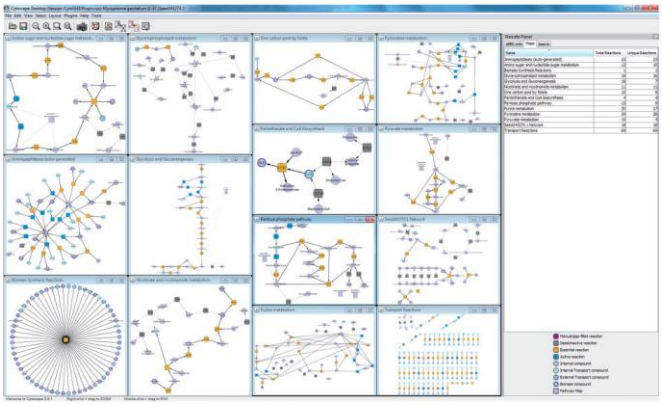
GMSS Tools

- SEED, Model SEED -> CytoSEED (Cytoscape plugin)
- BiGG



The BiGG knowledgebase can be accessed through a web browser.

Jan Schellenberger et al. BMC Bioinformatics, 2010



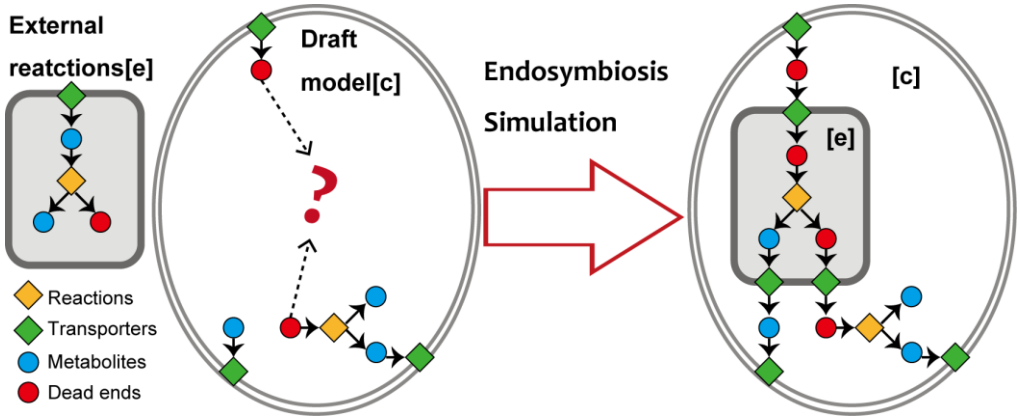
CytoSEED interface displaying a refined visualization of the metabolic model for *Mycoplasma genitalium* G-37

Matthew DeJongh et al. Bioinformatics, 2012

MGSS: Gap Filling

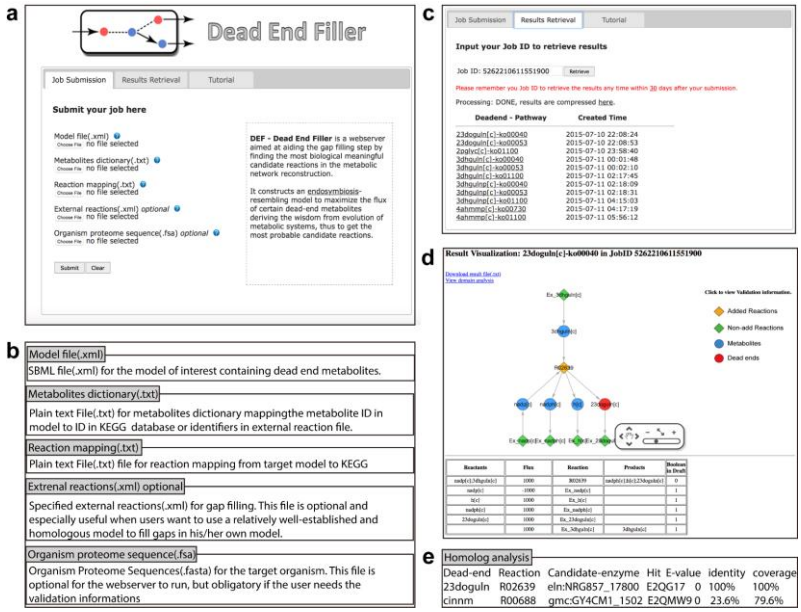
Gap filling for reconstructed metabolic model

- DEF: an automated dead-end filling approach based on quasi-endosymbiosis



DEF Web Server

http://bis.zju.edu.cn/DEF/



Methods for gene regulatory network modeling

- Ruled based formalism
- 基于规则的形式方法
- Correlation networks
- 关联网络
- Boolean networks
- 布尔网络
- Bayesian networks
- 贝叶斯网络
- Stochastic equations
- 随机方程
- Differential equations
- 微分方程
- Petri nets
- （派翠）Petri网
- 。。。

模型方法	优点	缺点
聚类方法	可获得共同表达的基因簇、相似基因组功能模块；	无法给出基因间具体的调控关系；基因表达相似的具有相同的功能的前提假设被质疑；
布尔网络模型	简单、离散性的网络；适于构建大规模网络	导致实验信息丢失，构建粗糙的网络；网络的准确度低；
概率布尔网络模型	处理不确定性； 可展现更多调控细节	不适于大规模网络；高计算复杂度
贝叶斯网络模型	可处理实验数据的噪声；体现真实调控过程的不确定性；构建更加准确的基因调控网络	忽略调控过程的时序关系；无法体现基因调控的反馈
动态贝叶斯模型	体现基因调控过程的时序性；允许存在环及回路来体现调控的反馈现象	计算开销大；适合小规模网络
微分方程模型	体现调控过程的时序性以及动力学特征；构建更加精细的网络	计算开销大，会造成维数灾难问题；适合小规模网络

gene regulatory network modeling

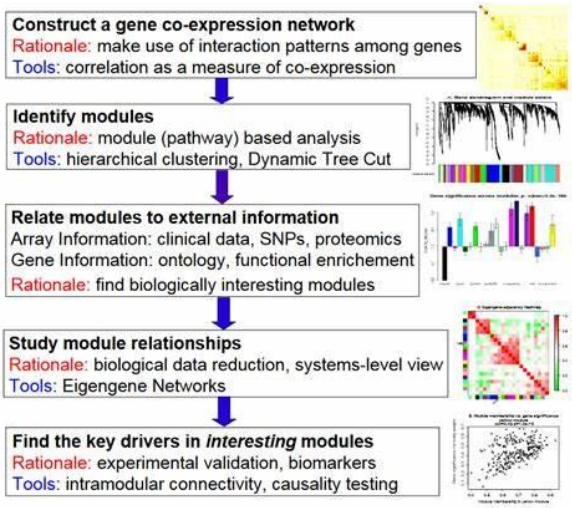
基于表达谱数据重构基因调控网络

基于共表达网络的方法	共表达网络是一种利用基因表达式数据中基因间共同上调和下调的模式构建的网络。	可以通过提取不同样本中相同的共同基因组，分析它们的共同调控模式来重构基因调控网络。
基于因果推断的方法	因果推断是一种通过识别基因之间存在的因果依赖关系来重构基因调控网络的方法。	通常基于基因表达中的时间顺序关系，并使用基于因果推断的评估方法来探索基因间的调控关系。
基于复杂系统理论的方法	复杂系统理论将基因调控网络看作一个复杂的动态系统，通过建立模型来探索基因之间的调控关系。	通常基于微分方程或随机网络模型，并利用基因表达数据的时间序列数据来参数化这些模型。
基于信息论的方法	信息论是一种识别基因间调控关系的方法，通过计算基因之间的信息交流来重构基因调控网络。	通常基于信息熵和条件熵，以及相关性和互信息等概念来计算基因之间的关联。

主要算法和工具：WGCNA

WGCNA是一种常见的基于共表达网络的算法，可以通过将共表达基因组分成不同的模块来识别基因之间的调控关系。这个方法通常使用R程序包来实现。

相关参考文献：Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9(1), 559.

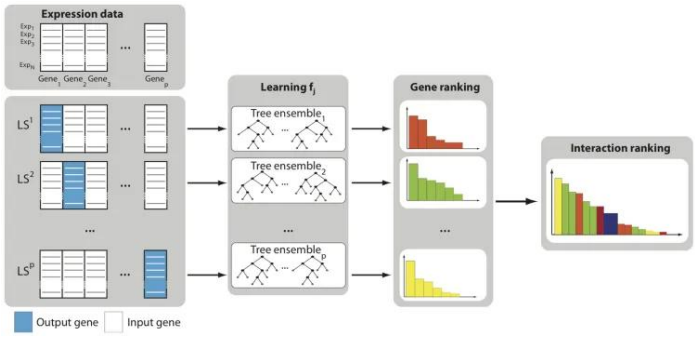


浙江大学 ZHEJIANG UNIVERSITY gene regulatory network modeling BiS @ZJU Bioinformatics Server

主要算法和工具： GENIE3

GENIE3是一种基于因果推断的算法，通过使用树回归模型来识别基因之间的因果依赖关系。这个方法通常使用Python程序包来实现。

相关参考文献：Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One, 5(9), e12776.

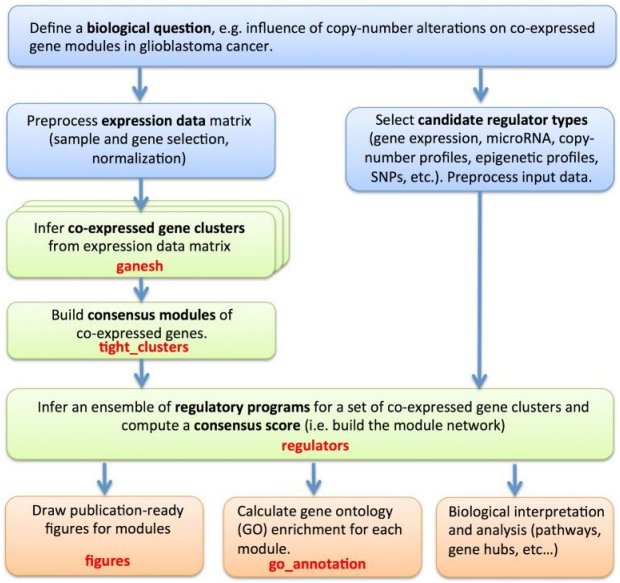


算法步骤:
目标基因选择：对数据集中的每个基因逐一进行分析，假设该基因为目标基因。
潜在调控因子选择：剩余的所有基因作为潜在的调控因子。
随机森林建模：通过构建一个随机森林模型（或其他树模型）来预测目标基因的表达水平，该模型的输入是潜在调控因子的表达数据。
重要性评分：根据随机森林模型，计算每个调控因子的特征重要性评分（Feature Importance Score），这反映了该基因作为调控因子的重要性。
网络构建：重复上述过程，对于每个目标基因，生成一个基因调控网络。最终网络由所有的特征重要性评分（即基因之间的调控关系）组成

主要算法和工具： Lemon-Tree

Lemon-Tree是一种基于复杂系统理论的算法，将基因调控网络看作一个动态的复杂系统，通过解决微分方程组来重构基因调控网络。这个方法通常使用Java程序包来实现。

相关参考文献：Zhang, Y., Liu, K., Liu, Z. P., Lai, L. and Huang, J. (2011). Lemon-Tree: a generic framework for construction of spatiotemporal gene regulatory networks. PLoS One, 6(10), e27156.

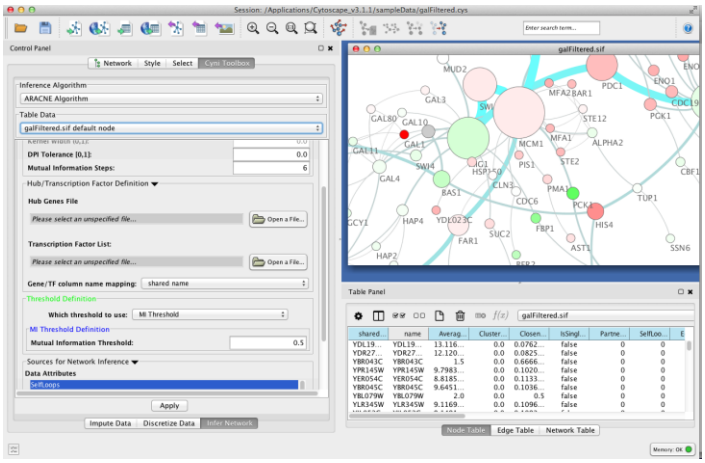


浙江大学 ZHEJIANG UNIVERSITY gene regulatory network modeling BiS @ZJU Bioinformatics Server

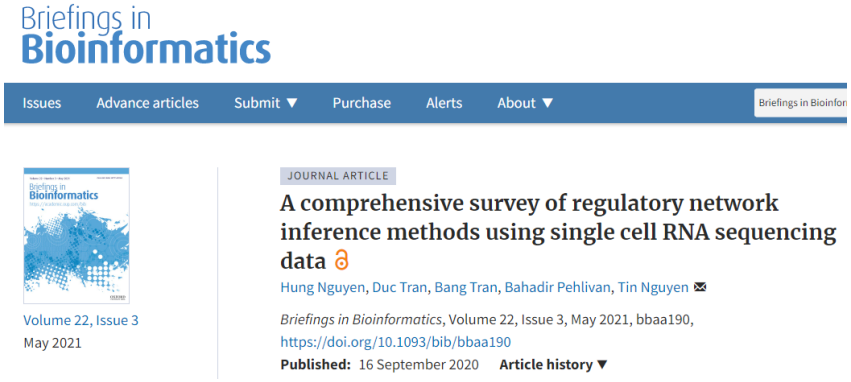
主要算法和工具：ARACNE

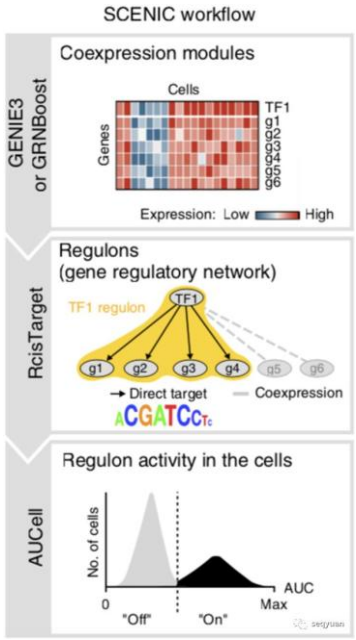
ARACNE是一种基于信息论的算法，使用互信息来量化基因之间的相互依赖关系。

相关参考文献：Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R. and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics, 7(Suppl 1), S7.



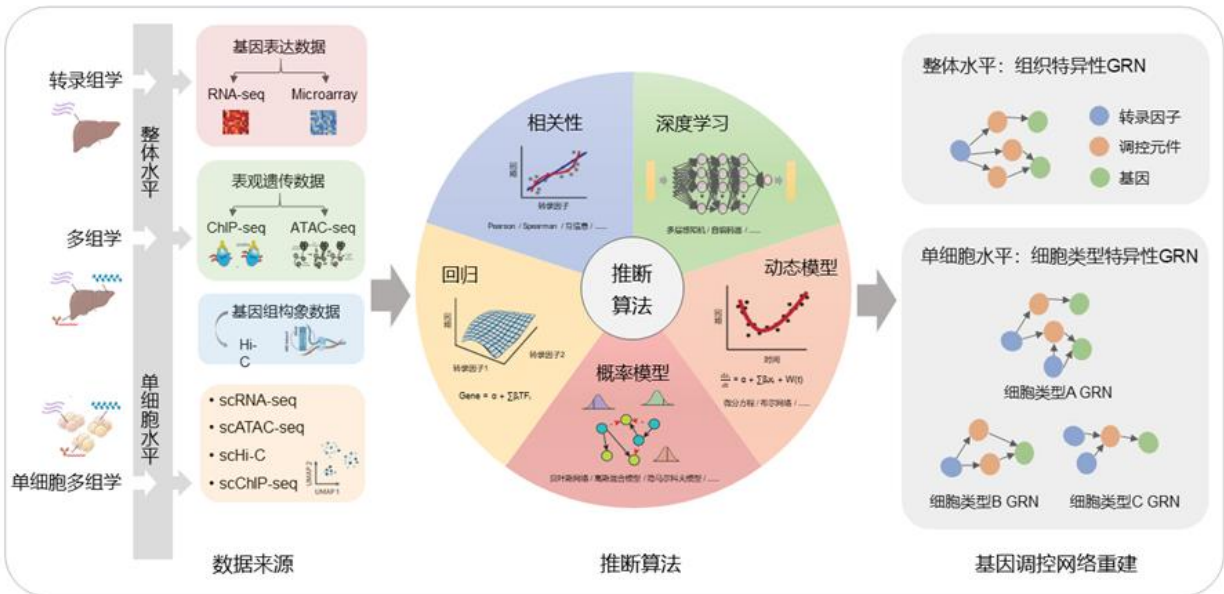
Single Cell GRNs



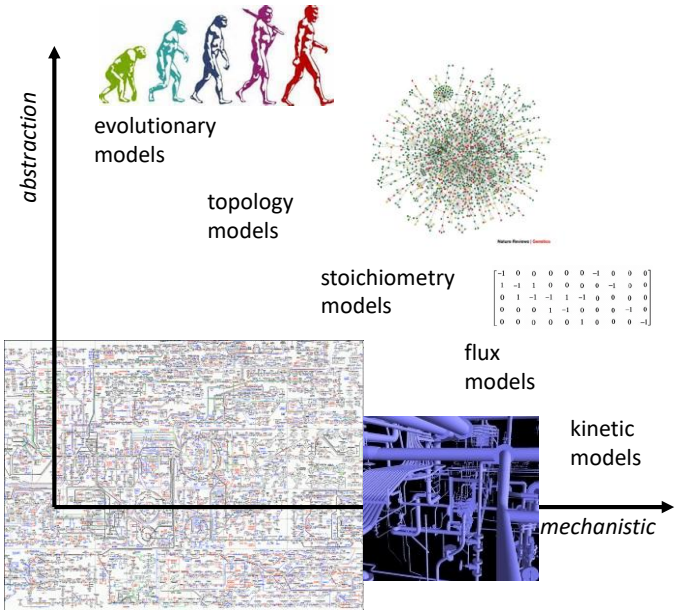


1.用GENIE3（随机森林）或GRNBoost (Gradient Boosting) 推断转录因子与候选靶基因之间的共表达模块。每个模块包含一个转录因子及其靶基因，纯粹基于共表达。
2.使用RcisTarget分析每个共表达模块中的基因，以鉴定enriched motifs；仅保留TF motif富集的模块和targets，每个TF及其潜在的直接targets gene被称作一个调节子（regulon）
3.使用AUCell评估每个细胞中每个regulon的活性，AUCell分数用于生成Regulon活性矩阵，通过为每个regulon设置AUC阈值，可以将该矩阵进行二值化（0|1，on|off），这将确定Regulon在哪些细胞中处于"打开"状态。

gene regulatory network reconstruction



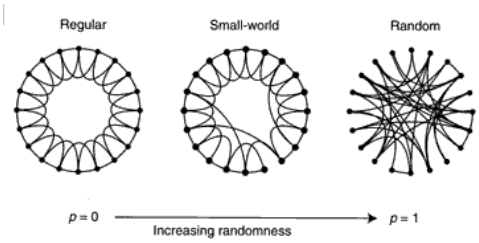
Different levels of systems biology models and analysis



4 Topological Analysis

Networks

Regular, random and real-world networks



D.J. Watts, S.H.Strogatz, Nature **393**, 440 (1998)

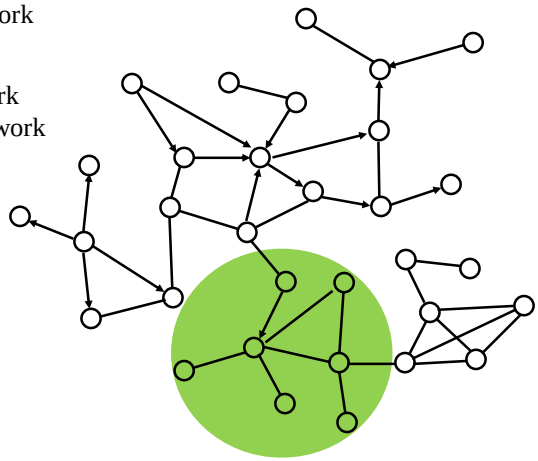
Real-world networks are likely to be small-world networks:

- The separation between any two randomly chosen nodes is very short.
- A high degree of clustering.
- Distrubution of connectivity in most real-world networks follows a power-law distribution $P(k) \sim k^{-\gamma}$ (scale-free property).

Network Analysis/Measures

- directed network
- undirected network

- weighted network
- unweighted network



- Degree k_i 度
- Degree distribution $P(k)$ 度分布
- Mean path length 平均路长
- Network Diameter 网络直径
- Clustering Coefficient 集聚系数

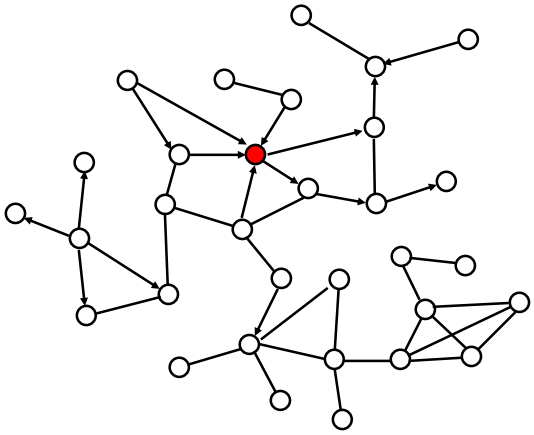
G(V,E); |V| = 32, |E| = 39

Cliques: protein complexes

Hubs: regulatory modules

Subgraphs: maximally weighted

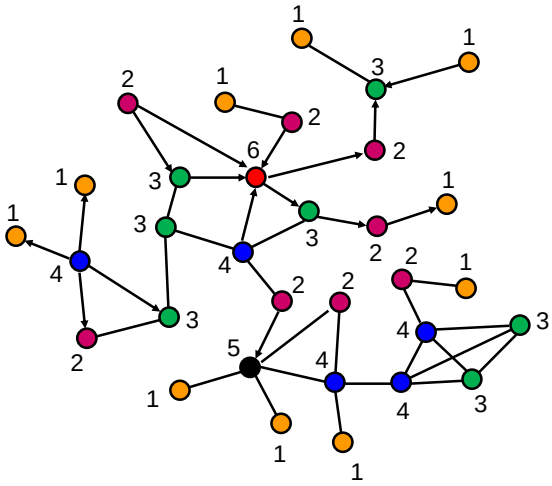
Network Analysis/Measures



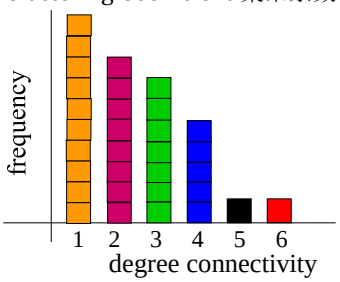
- Degree k_i 度
- Degree distribution $P(k)$ 度分布
- Mean path length 平均路长
- Network Diameter 网络直径
- Clustering Coefficient 集聚系数

K=6
K_{in}=4, K_{out}=2

Network Analysis/Measures



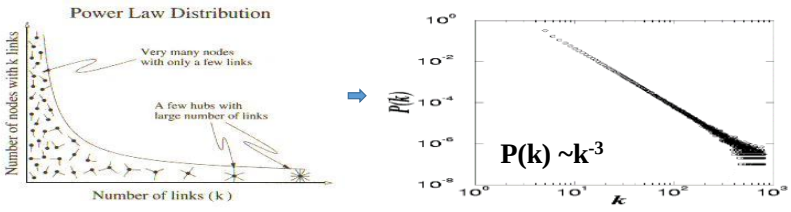
- Degree k_i 度
- Degree distribution $P(k)$ 度分布
- Mean path length 平均路长
- Network Diameter 网络直径
- Clustering Coefficient 集聚系数

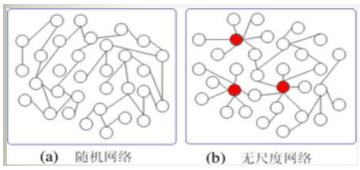


Scale-free network

无标度网络Scale-free network

- 真实网络几乎都具有小世界效应，大量真实网络的结点度服从幂率分布。 $P(k) \sim k^{-\gamma}$





无标度网络与随机网络的对比：(a)中的随机网络，大部分结点都连出2到3条边，0条与1条边的和4条边的都很少，而(b)中的无尺度网络，大部分结点连1条边，少数结点（红色）连有大量边。

Scale-free networks BA模型

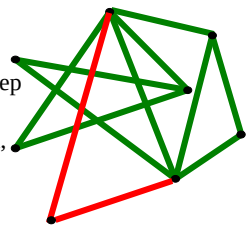
Albert-László Barabási与Réka Albert在1999年的论文中提出了一个模型来解释复杂网络的无标度特性，称为BA模型。这个模型基于两个假设：

增长模式：不少现实网络是不断扩大不断增长而来的，例如互联网中新网页的诞生，人际网络中新朋友的加入，新的论文的发表，航空网络中新机场的建造等等。

优先链接模式：新的结点在加入时会倾向于与有更多链接的结点相连，例如新网页一般会有到知名的网络站点的链接，新加入社群的人会想与社群中的知名人士结识，新的论文倾向于引用已被广泛引用的著名文献，新机场会优先考虑建立与大机场之间的航线等等。Rich gets richer!

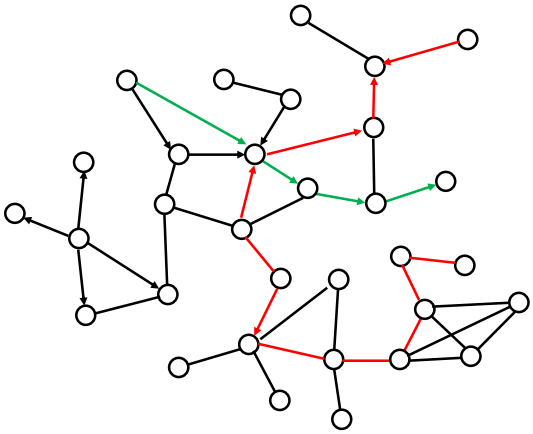
The algorithm

- Growth: starting with a small number of nodes m_0, at every timestep adding a new node with m (<=m_0) edges to the exsiting nodes.
- Preferential attachment: when choosing the nodes to be connected, the probability of being connected depends on the its degree k_i/sum(k_j).



A.-L.Barabási, R. Albert, Science **286,** 509 (1999)

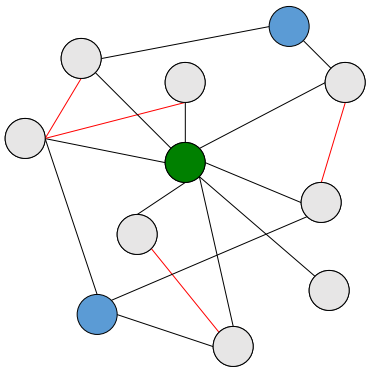
Network Analysis/Measures



Longest Shortest-Path: diameter

- Degree k_i 度
- Degree distribution $P(k)$ 度分布
- Mean path length 平均路长
- Network Diameter 网络直径
- Clustering Coefficient 集聚系数
- Betweenness介数
- Closeness紧密度
- Topology Coefficient拓扑系数

Clustering Coefficient



The density of the network surrounding node i, characterized as the number of triangles through i. Related to network modularity

$$C_i = \frac{n_i}{\binom{k}{2}} = \frac{2n_i}{k \cdot (k-1)}$$

The center node has 8 (grey) neighbors

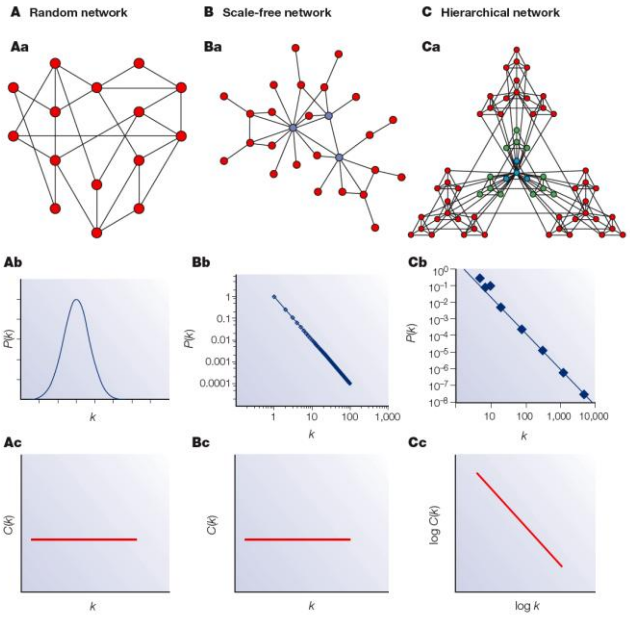
There are 4 edges between the neighbors

C = 2*4 /(8*(8-1)) = 8/56 = 1/7

k: neighbors of *i*

n_i: edges between node *i*'s neighbors

Hierarchical Networks



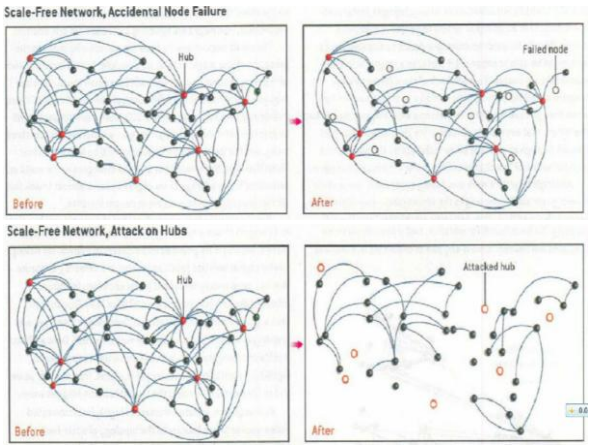
Scale-Free Networks are Robust

- Complex systems (cell, internet, social networks), are resilient to component failure
- Network topology plays an important role in this robustness
  - Even if ~80% of nodes fail, the remaining ~20% still maintain network connectivity
- *Attack vulnerability* if hubs are selectively targeted
- In yeast, only ~20% of proteins are lethal when deleted, and are 5 times more likely to have degree $k>15$ than $k<5$.

Attack Tolerance

Complex systems maintain their basic functions even under errors and failures

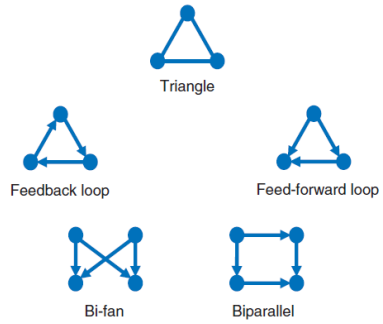
(cell → mutations; Internet → router breakdowns)



Biological Networks Properties

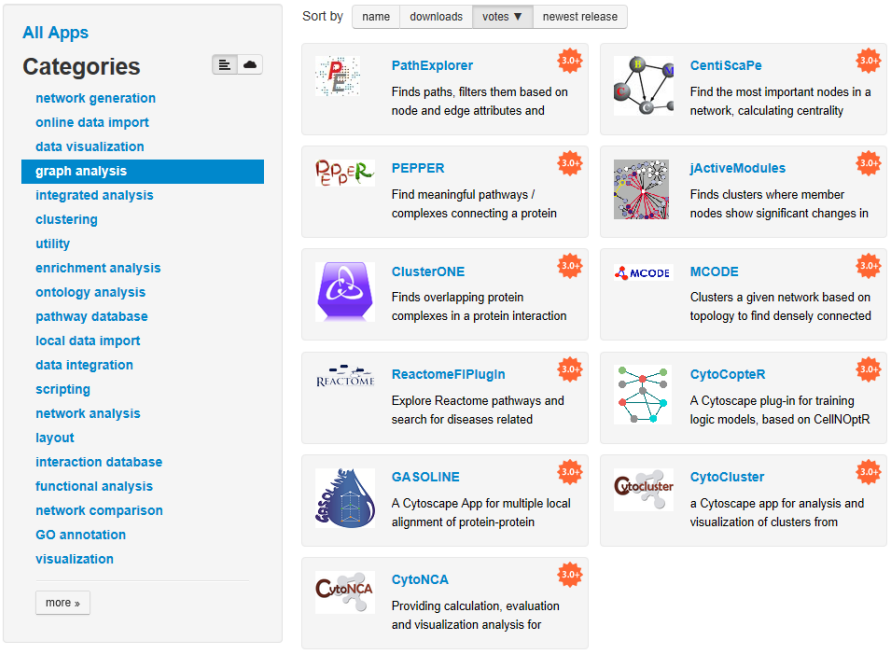
- ✓ Small World: A small average path length
  - Mean shortest node-to-node path
- ✓ Power law degree distribution: Rich get richer
- ✓ Hierarchical Modularity: A large clustering coefficient
  - How many of a node's neighbors are connected to each other
- ✓ Robustness: Resilient and have strong resistance to failure on random attacks and vulnerable to targeted attacks
- Assortative: hubs tend not to interact directly with other hubs.
  - Hubs tend to be "older" proteins (so far claimed for protein-protein interaction networks only)
- Hubs also seem to have more evolutionary pressure—their protein sequences are more conserved than average between species (shown in yeast vs. worm)
- .....

Network motifs



The feed-forward loop, Bi-fan, Biparallel are over-presented. *Science* 298, 2002
Triangles in human 194 times that in S. cerevisiae. *Nature Comm*. 2013
MotifNet: a web-server for network motif analysis. *Bioinformatics*, 2017

Cytoscape APPs



FBA application

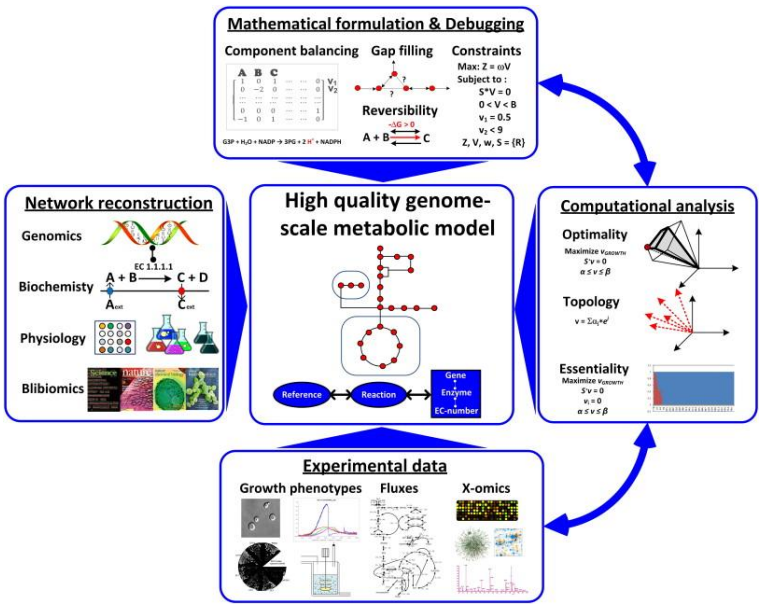
Problems

- Network Inference
  - Microarray, Protein Chips, other high throughput assay methods, NGS…
- Function prediction
- Functional module detection
- Topological Analysis
- Dynamics Analysis
  - Signal Flow Analysis
  - Metabolic Flux Analysis
  - Steady State, Response, Fluctuation Analysis
- Evolution Analysis

- With very limited, noisy, and incomplete information, discovering underlying network principles is very challenging.

Systems Biology Modeling & Discussion

- Pathway inference from gene regulation networks
- Pathway inference by integrating dynamic data with protein interaction networks
- Inferring network structure from dynamic data
- Uncovering organizational principles underlying network dynamics
- Relating the effect of gene knockouts to network topology
- Relating double mutant perturbations with the physical interactome
- Interpersonal genetic variation as perturbations of cellular systems



Biotechnology Advances, 2011

- Directed graphs 有向图
- Ruled based formalism 基于规则的形式方法
- Boolean networks 布尔网络
- Bayesian networks 贝叶斯网络
- Stochastic equations 随机方程
- Differential equations 微分方程
- Petri nets （派翠）Petri网
- 。。。

A typical iteration of biological modelling

1. data gathering
2. formalizing natural description
3. applying analysis methods
4. interpreting analysis results

5 ODE & FBA

The simplest chemical reaction

A → B

- irreversible, one-molecule reaction
- any metabolic pathway can be described by a combination of processes of this type (including reversible reactions and, in some respects, multi-molecule reactions)

various levels of description:

- homogeneous system, large numbers of molecules = ordinary differential equations, **kinetics**
- small numbers of molecules = probabilistic equations, **stochastics**
- spatial heterogeneity = partial differential equations, **diffusion**
- small number of heterogeneously distributed molecules = single-molecule tracking (e.g. cytoskeleton modelling)

Some (Bio)Chemical Conventions

Concentration of Molecule A = [A], usually in units mol/litre (molar)

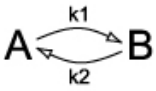
Rate constant = k, with indices indicating constants for various reactions (k_1, k_2...)

Therefore:

A→B

$$\frac{d[A]}{dt} = -\frac{d[B]}{dt} = -k_1[A]$$

Reversible, Single-Molecule Reaction

A ←→ B, or A → B || B → A, or 

Main principle: Partial reactions are **independent**!

- Differential equations:

forward reverse

$$\frac{d[A]}{dt} = -k_1[A] + k_2[B]$$

$$\frac{d[B]}{dt} = k_1[A] - k_2[B]$$

- Equilibrium (=steady-state):

$$\frac{d[A]_{equi}}{dt} = \frac{d[B]_{equi}}{dt} = 0$$

$$-k_1[A]_{equi} + k_2[B]_{equi} = 0$$

$$\frac{[A]_{equi}}{[B]_{equi}} = \frac{k_2}{k_1} = K_{equi}$$

Irreversible, two-molecule reaction

A+B→C

Differential equations:

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[C]}{dt}$$

$$\frac{d[A]}{dt} = -k[A][B]$$

Non-linear!

Underlying idea: Reaction probability = Combined probability that both [A] and [B] are in a "reactive mood":

$$p(AB) = p(A)\,p(B) = k_1^*[A]k_2^*[B] = k[A][B]$$

A simple metabolic pathway

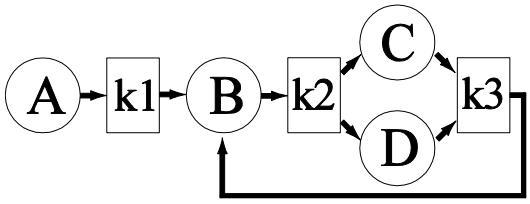
A→B←→C+D

Differential equations:

d/dt	decay	forward	reverse
[A]=	`-k1[A]`		
[B]=	`+k1[A]`	`-k2[B]`	`+k3[C][D]`
[C]=		`+k2[B]`	`-k3[C][D]`
[D]=		`+k2[B]`	`-k3[C][D]`

Metabolic Networks as Digraphs

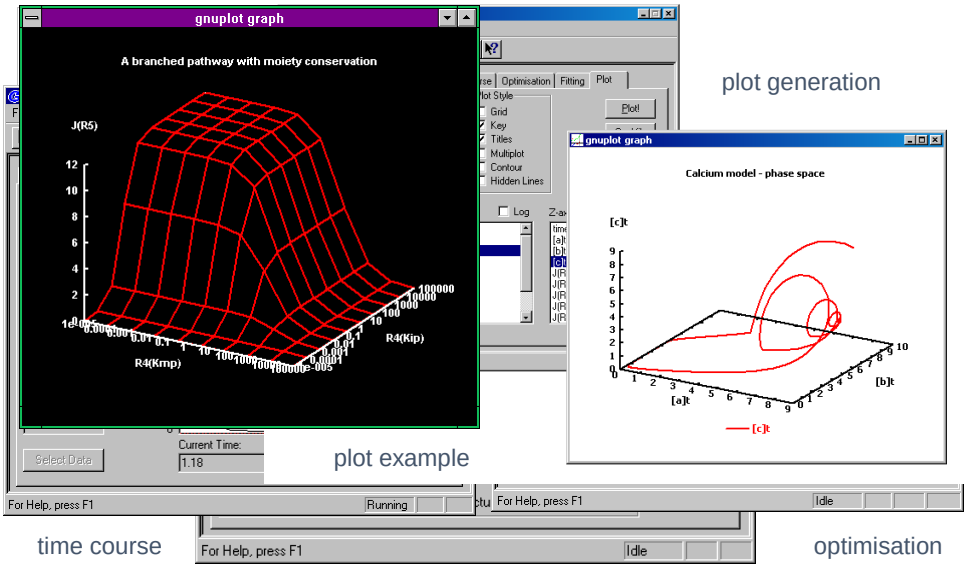
A→B←→C+D



	k1	k2	k3
A	**-1**	**0**	**0**
B	**1**	**-1**	**1**
C	**0**	**1**	**-1**
D	**0**	**1**	**-1**

d/dt	decay	forward	reverse
[A]	`-k1[A]`		
[B]	`+k1[A]`	`-k2[B]`	`+k3[C][D]`
[C]		`+k2[B]`	`-k3[C][D]`
[D]		`+k2[B]`	`-k3[C][D]`

Gepasi/Copasi simulation

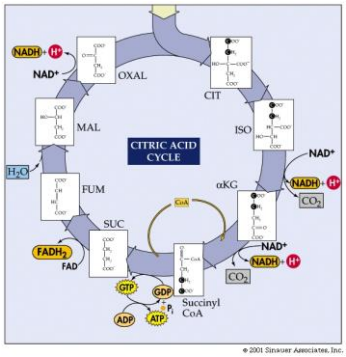


Various modeling methods

- Metabolic control analysis (MCA) (Kinetic approaches)
  - can yield greater dynamic insights, but are limited by the availability of the kinetic rate expression for various reactions in the metabolic network of a microorganism.
    - kinetic coefficients difficult to obtain
    - loss of overview with large models
- Metabolic flux analysis (MFA) (stoichiometric balance method)
  - commonly used, as it requires only the stoichiometry of the reactions present in the metabolic network
- Flux balance analysis (FBA)
  - used to determine the fluxes in a metabolic network through linear programming to evaluate the fluxes. In this method, the intracellular metabolites are included as variables and the pseudo steady state approximation is used in evaluating the overall fluxes. The solution inherently does not ensure directionality of the reactions.

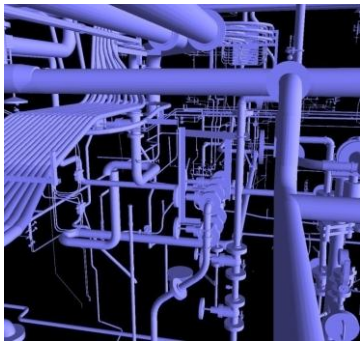
Flux balance analysis in metabolic networks

1. Metabolic networks



http://bill.srnr.arizona.edu/classes/182/CitricAcidCycle-LowRes.jpeg

2. Flux Balance Analysis



http://images.google.com/imgres?imgurl=http://www.cs.unc.edu/~walk/models/double_eagle/pieces/pipes.jpg&imgrefurl=http://www.cs.unc.edu/~walk/models/double_eagle/pieces/&h=1095&w=1156&sz=307&tbnid=1C0069tnx901:&tbnh=142&tbnw=150&hl=en&start=1&prev=/images%3Fq%3D%2Bpipes%2B%26hl%3Den%26lr%3D

Dynamic mass balance

Solution !

In order to identify invariant characteristics of the network we assume the network is at **steady state**.

Problem!

V=V(k1, k2,k3…) is actually a function of concentration as well as several kinetic parameters.

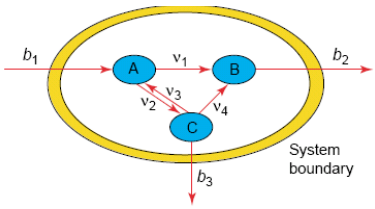
it is very difficult determine kinetic parameters experimentally.

Consequently there is not enough kinetic information in the literature to construct the model.

Concentration vector | Stoichiometry Matrix | Flux vector

$$\frac{dx}{dt} = S \cdot v$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{V}$$



Kauffman K J, Prakash P, Edwards J S. Advances in flux balance analysis. Current Opinion in Biotechnology, 2003, 14(5):491–496.

Steady state assumption

$$\frac{dx}{dt} = S \cdot v \xrightarrow{\text{Steady state assumption}} 0 = S \cdot v$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S V

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S V

Flux cone and metabolic capabilities

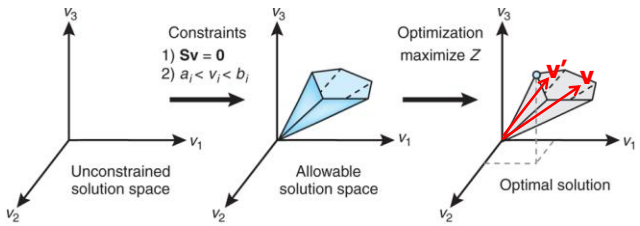
$$0 = S \cdot v$$

Observation: the number of reactions considerably exceeds the number of metabolites (The S matrix will have more columns than rows)变量数大于方程数

无唯一解！

The null space of viable solutions to our linear set of equations contains an infinite number of solutions.

What about the constraints?



- It defines what the network can do and cannot do!
- Each point in this cone represents a flux distribution in which the system can operate at steady state.

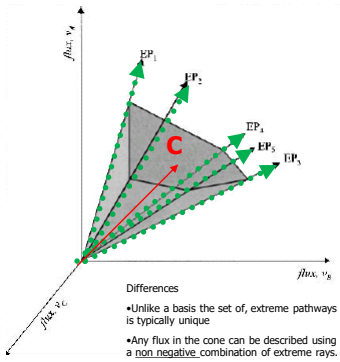
Nature Biotechnology 28: 245–248 (2010)

Flux cone – "Extreme pathways"

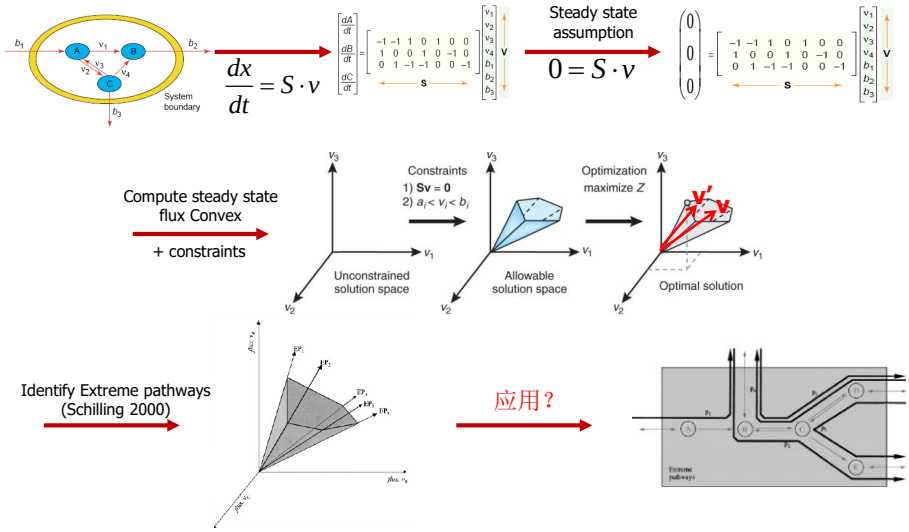
Extreme rays - "extreme rays correspond to edges of the cone. They are said to generate the cone and cannot be decomposed into non-trivial combinations of any other vector in the cone."- schilling 2000

We use the term **Extreme Pathways** when referring to Extreme rays of a convex polyhedral cone that represents metabolic fluxes.

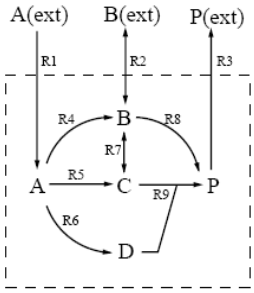
Any steady state flux distribution can be represented by a non-negative linear combination of extreme pathways.



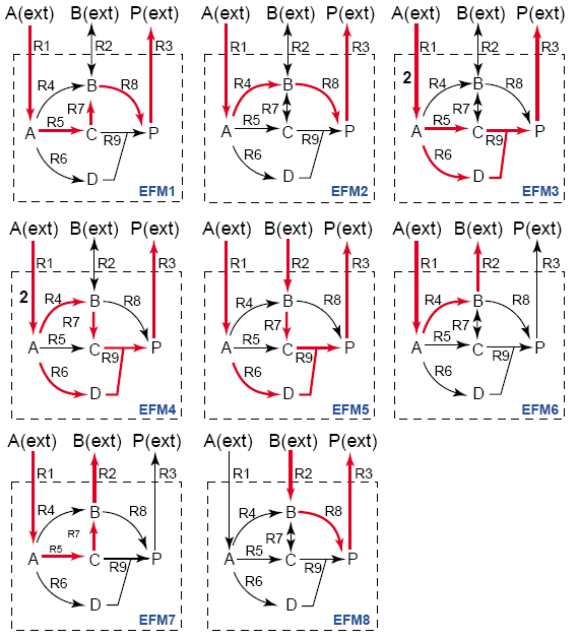
FBA: The entire process



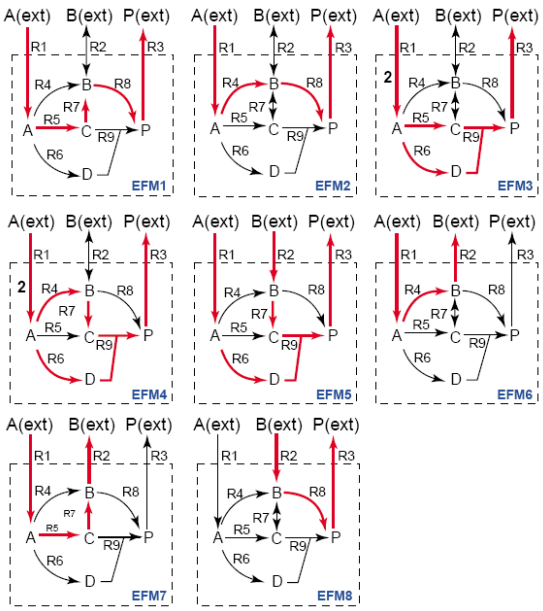
FBA application



EFM1	R1-R5-R7-R8-R3
EFM2	R1-R4-R8-R3
EFM3	2R1-R5-R6-R9-R3
EFM4	2R1-R4-R7-R6-R9-R3
EFM5	R1-R6-R2-R7-R9-R3
EFM6	R1-R2-R4
EFM7	R1-R5-R7-R2
EMF8	R2-R8-R3



FBA application



1. 识别可运行模式
   如：A→P，EFM1、EFM2、EFM3、EFM4。
2. 找出最优途径
   如：找出产生P的最优途径　EFM1　EFM2
3. 识别重要反应
   从底物B产生P的途径,R8是必需的。而且仅仅以A为底物转化为P的优化途径（EFM1,EFM2）中，R8也是必需的。这说明R8对于用A或者用B作为底物来产生P都是很重要的。
4. 分析途径长度
   由A产生P的最短的途径是EFM2，而最长的途径是EFM4
5. 识别没有被途径利用的反应
6. 识别酶子集
7. 计算网络的冗余性
8. 预测优化途径和优化生长速率
9. 预测基因修饰后结果
10. 计算最小切割子集
11. 标注未知孤独基因功能
12. 预测转录调控
13. 预测最小底物组成
14. 。。。

FBA algorithms

方法名称	功能分类	描述
EMILiO	工程菌设计	证实能促进与生长相偶联的产物生产的反应集合和预测反应流量范围来实现理论最大产量
FBAME	流量平衡分析	研究细菌细胞膜组成如何影响其生理和代谢
FBAwMC	流量平衡分析	利用细胞质空间大小作为额外的限制模拟胞内大分子对细胞生理的影响
FBA	流量平衡分析	预测目标函数最优化时的流量分布
FCA	流量平衡分析	通过发现配对反应集合阐明基因组规模代谢系统的拓扑学属性和流量连通性
FSA	流量平衡分析	检查目标函数流量值对其他流量值变化的响应
FVA	流量平衡分析	预测在设定的目标函数条件下代谢反应流量的上下限值
Flux-sum	流量平衡分析	通过预测代谢物转化率模拟细胞生理变化
GapFind/GapFill	代谢漏洞填补	发现代谢漏洞并通过设置反应为可逆或添加新的反应填补漏洞
GeneForce	代谢调控	寻找最经济的方式改变调控规则使模型预测与实验观测相符
GDLS	工程菌设计	使用 MILP 快速计算经过复杂的基因修饰的代谢网络
GrowMatch	代谢漏洞填补	修正反应或生物量组成使模型预测与细胞表型一致
IOMA)	组学信息限制	整合蛋白组学和代谢组学数据预测流量分布
★ MOMA	基因/反应扰动	预测基因删除后新的流量分布

刘立明, 刘婷, 邹伟. 基因组规模代谢网络模型的约束算法及其应用. 生物加工过程, 2012, 6: 70-77.

FBA algorithms

代谢调节最小化分析(**MOMA:** minimization of metabolic adjustment)

为了更好地了解发生扰动时的系统流分布情况,2002 年哈佛大学系统生物学系的Segre 等人提出了代谢调节最小化MOMA思想. 该方法的假设是, 代谢网络在有扰动的影响下会对系统做出调整来维持生存, 而这种调整使系统在调整后, 其流分布状态与野生型情况下相比发生的变化最小. MOMA 利用二次规划的方法寻求与用 FBA 分析得到的正常野生型最优流分布之间欧几里德距离最小的流状态, 并将其作为扰动后的网络状态. 事实上MOMA 方法也就是计算与基因敲除前的最佳流分布状况最接近的可行的解.

MOMA 方法被成功地用于预测大肠杆菌基因敲除后的网络可行性并定量分析流分布状况, 这个实验在使用FBA方法计算时并不能得到理想的结果.除此之外, MOMA也是十分实用型的工具, 可以用来约束代谢产量等, 基于MOMA 方法的基因敲除的模拟在提高番茄红素的生产量、有效抗氧化剂化合物、大肠杆菌的试验中都被证实具有良好的改善外界扰动的能力. 大量结果表明MOMA 的结果优于FBA.

FBA algorithms

OptForce	工程菌设计	证实实现目的产品过量生产必需的最经济的代谢流量操作
OptGene	工程菌设计	利用遗传算法预测基因删除靶点
OMNI	代谢漏洞填补	证实实现模型预测流量与实验检测流量相一致时需修正的代谢反应
★ OptKnock	工程菌设计	结合某一产物的优化生产预测反应删除靶点
OptReg	工程菌设计	预测基因删除靶点和反应调控靶点
★ OptStrain	工程菌设计	通过添加外源代谢途径设计生产与生长偶联的代谢产物
pFBA	流量平衡分析	在最优生长条件下用双水平线性规划求解有酶关联的反应的最小流量
PhPP	流量平衡分析	以 2 个代谢反应流量为变量通过线性规划计算目标函数最优解空间
★ PROM	代谢调控	介绍描述基因状态和基因-转录因子相互作用关系的可能性
rFBA	代谢调控	整合转录调控信息的流量平衡分析
★ ROOM	基因/反应扰动	预测基因删除后新的流量分布
SR-FBA	代谢调控	用 MILP 问题描述布尔逻辑构建的调控机制
TMFA	热力学	添加线性热力学限制预测热力学可行的流量范围

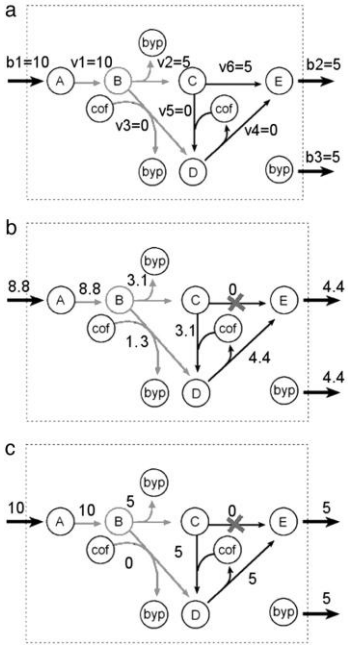
FBA algorithms

调节开关最小化分析ROOM(regulatory on/off minimization)

ROOM是一种确定扰动后代谢流状况的分析方法， 该方法寻找与正常野生型流分布状况相比具有显著流值改变的数目最小的解, 用它来作为扰动后的代谢流分布, 这里“显著”是指用户在考虑可行的计算时间的前提下所指定的一系列阈值. 实验证实ROOM准确地预测了大肠杆菌在基因敲除等扰动后, 经过一段时间所计算出的调整的流分布.

PNAS 2005, 102(21): 7695-7000

FBA algorithms



(a) A given flux distribution for the wild-type intact network that can be obtained by FBA and experimental flux data.
(b) MOMA's prediction for the knocked-out network following the knockout of reaction v6.
(c) ROOM's prediction for the knocked-out network. ROOM finds changes in flux only along a short alternative pathway through v5 and v4, preserving the optimal growth rate of the wild-type strain

Regulatory on/off minimization of metabolic flux changes after genetic perturbations, PNAS,2005: 7695-7700

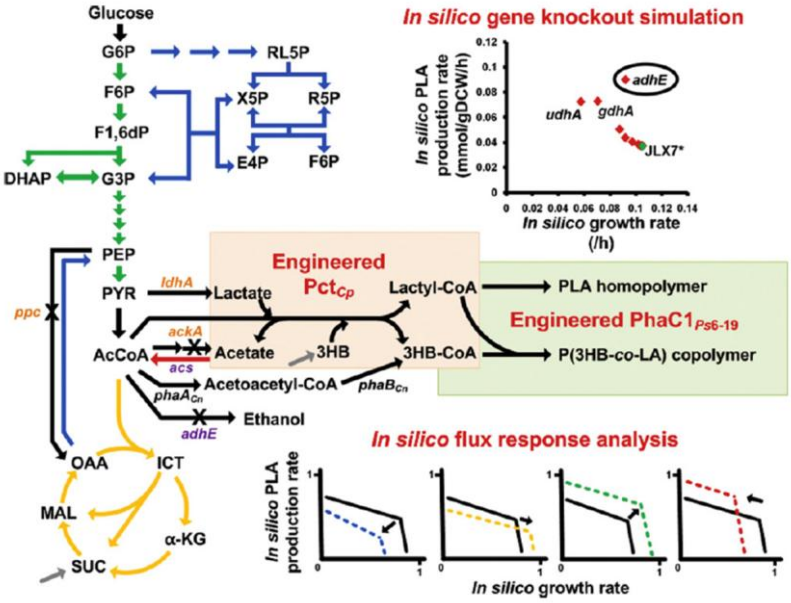
FBA algorithms

基于二次线性规划调节器的流平衡分析方法**DFBA-LQR**
(dynamic flux balance analysis with linearquadratic regulator)

由于FBA 是一种静态分析方法, Mahadevan 等人提出了动态流平衡分析方法(DFBA), 这种方法利用代谢反应的计量信息来计算某一时间的流分布情况, 但是这种方法具有一定的局限性, 如在构建模型时需要部分重要反应的动力学参数等. 因此Uygun 和 Matthew 等人在DFBA 的基础上提出将最优控制理论和改进的二次线性规划方法相结合, 用来估计给定的代谢模型经过调整后的代谢网络流分布情况, 这种方法被称为DFBA-LQR. 该方法能实现对于动态流平衡模型的构建、分析扰动后代谢流的反应, 也能用来计算流控制系数和系统的线性化反应动力学. 通过老鼠肝细胞环境下油脂生成的网络构建和模拟, 该方法被证实也具有很好的应用价值.

Ind. Eng. Chem. Res., **2006**, *45* (25): 8554–8564

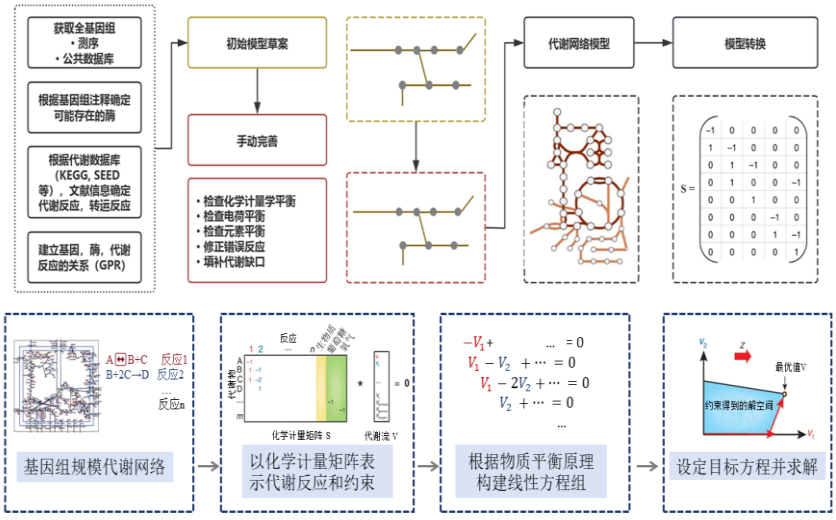
FBA application



Jung et al. Biotech & Bioeng, 2010

FBA Tools

- CellNetAnalyzer/
- BioSPICE MOMA
- COBRA Toolbox
- FluxExplorer
- MetaFluxNet
- PathwayAnalyser
- Fbatool



6 Synthetic Biology

Synthetic Biology（合成生物学）

is an emerging area of research that can broadly be described as **the design and construction of novel artificial biological pathways**, organisms or devices, or the **redesign of existing natural biological systems**.

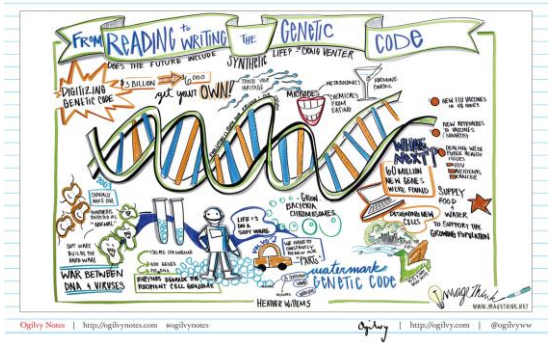
(1) 新的生物零件（part)、组件(device) 和系统的设计与构建；

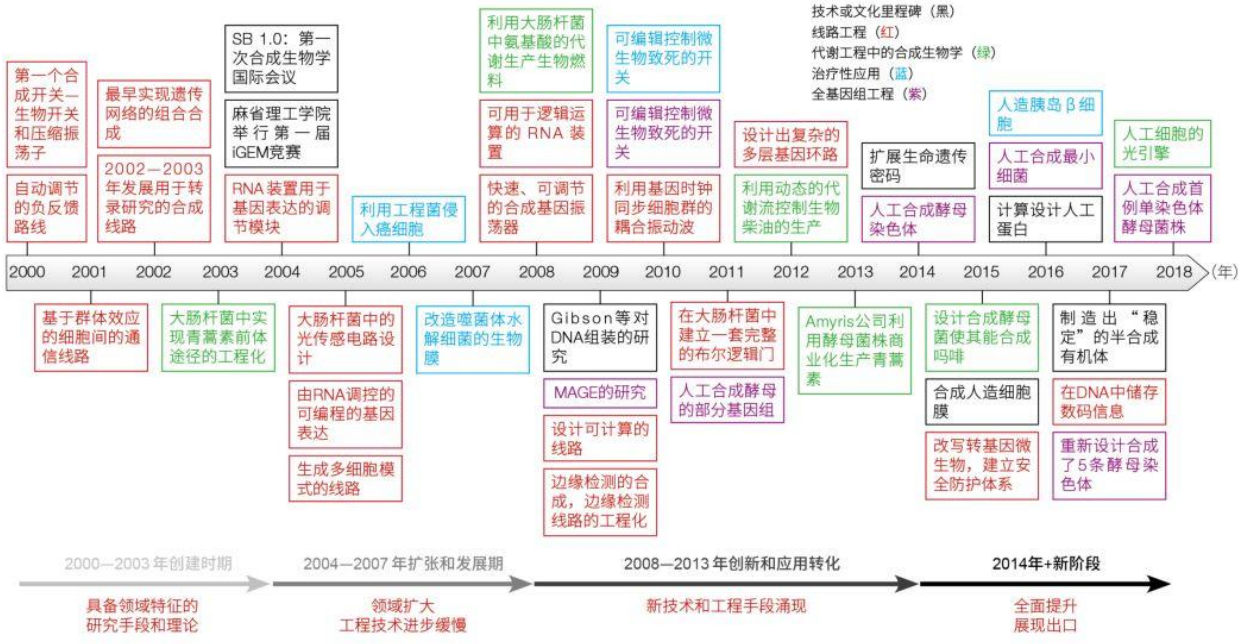
(2) 对现有的、天然存在的生物系统的重新设计,以造福人类社会。

1.传统生物学**Reverse engineering**:
Taking apart sequenced whole genomes through cross-genome comparison to see how they work.
2.合成生物学**Forward engineering**:
Design and build engineered biological systems to test biological hypothesis.

认知

改造





Biological synthetic circuits（回路、线路）

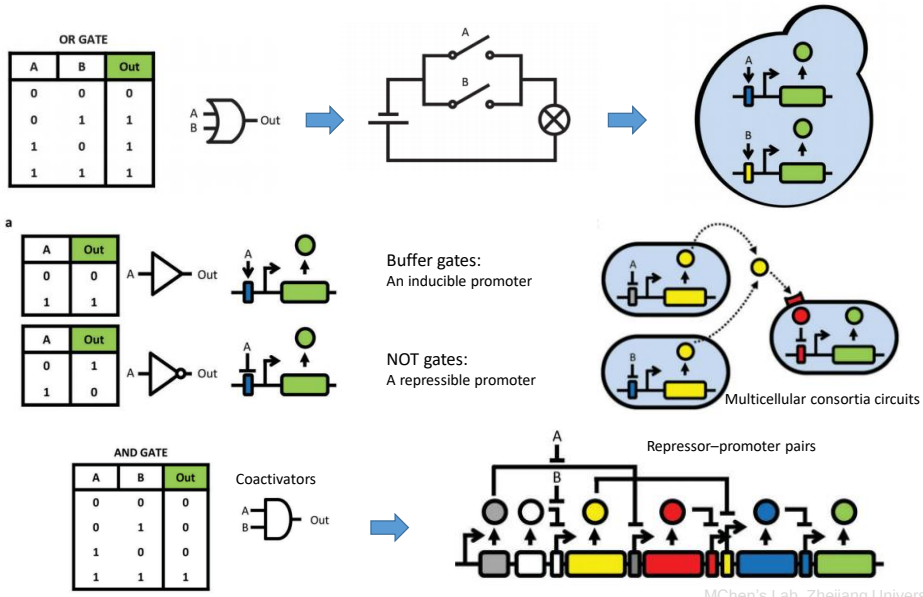
e.g,

fluorescent proteins荧光蛋白
light bulbs or LEDs灯

transcription factors转录因子
transistors or logic gates逻辑门
repressors抑制剂
NOT gates
activators激活剂
OR/AND gates

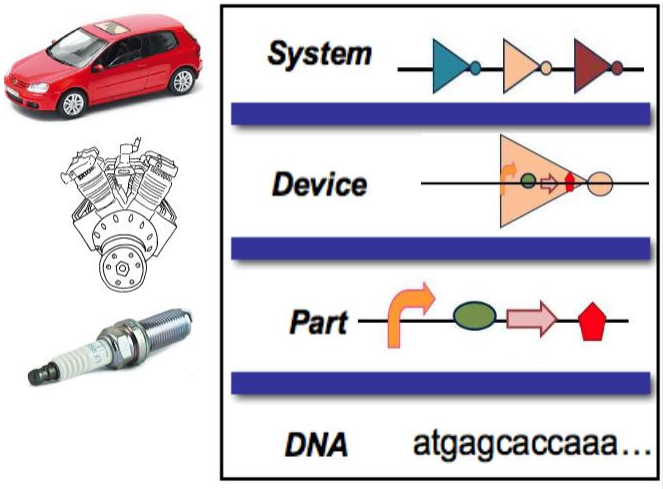
Polymerases聚合酶
(transcriptional machinery)
batteries?电池

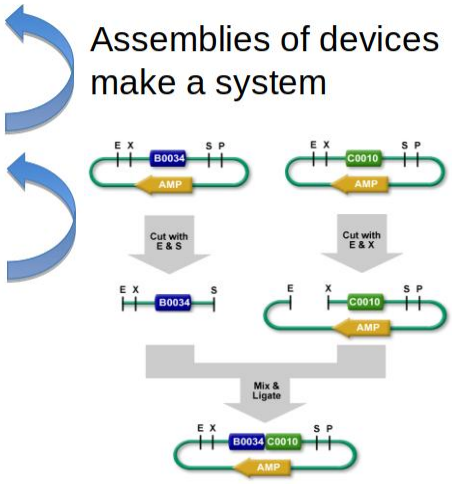
and so on...



合成生物学部件、积木、砖块（Biobricks）

生物部件（Biobricks）和标准化：开发标准化的生物组件，使其可以在不同的生物系统中通用。





A synthetic oscillatory network of transcriptional regulators

转录调节器的合成振荡网络

Michael B. Elowitz & Stanislas Leibler

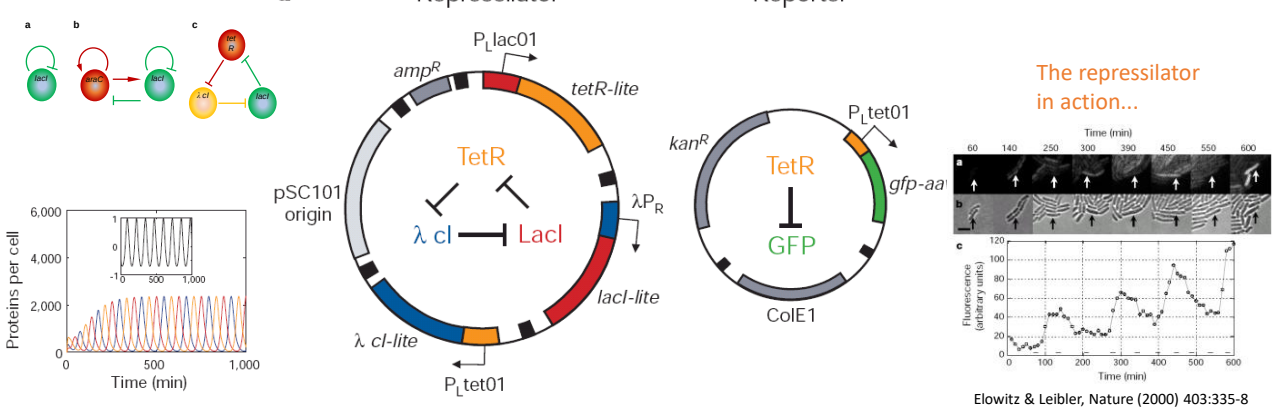
Departments of Molecular Biology and Physics, Princeton University, Princeton, New Jersey 08544, USA

The Repressilator – an engineered genetic circuit designed to make bacteria glow in a periodic fashion

= "repressor" + "oscillator"

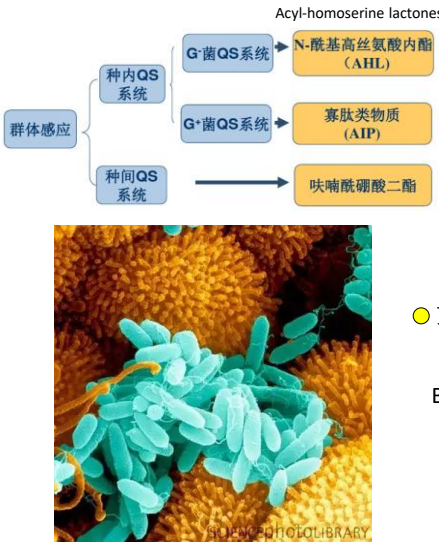
repressors (genes that turn off other genes)

green fluorescent protein (glows!)

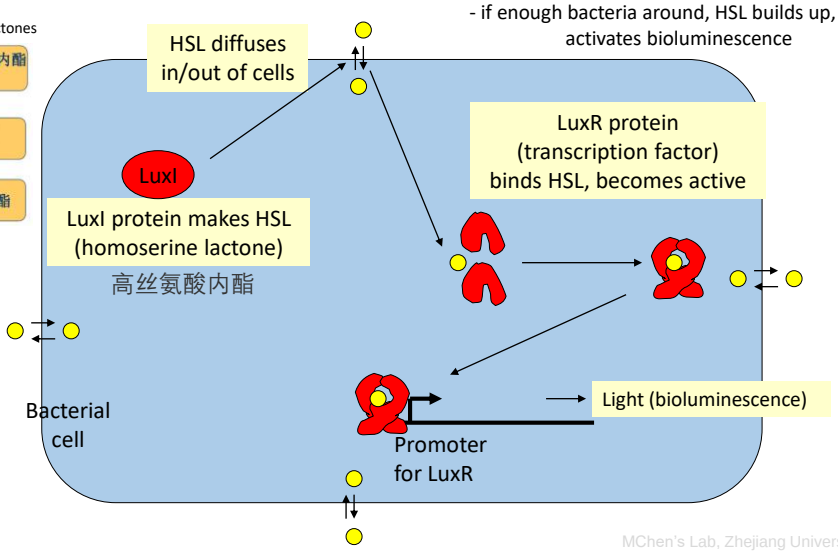


Elowitz & Leibler, Nature (2000) 403:335-8

群体效应 Quorum sensing: chemical-based bacterial communication

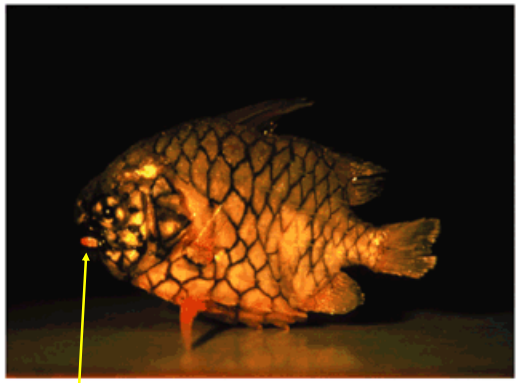


Neighboring bacteria produce HSL also
- if enough bacteria around, HSL builds up, activates bioluminescence



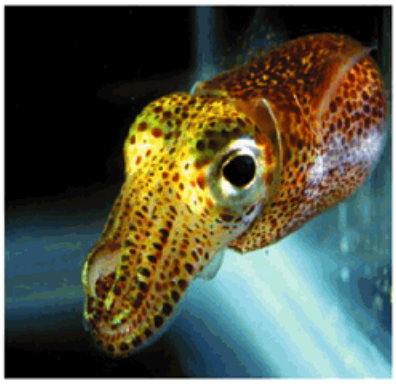
Bioluminescence – only occurs when bacteria are present at high density
==> bacteria communicate in order to establish their density

Australian pinecone fish



~10^{10} Vibrio bacteria / ml fluid
- fish uses to hunt for prey

Hawaiian bobtail squid



海洋细菌费氏弧菌

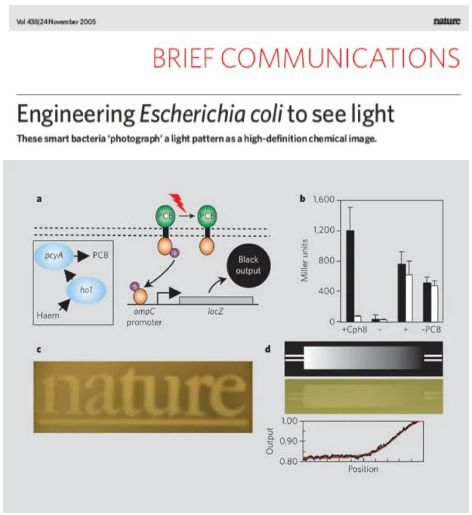
~10^{11} Vibrio bacteria / ml fluid
in light organ in squid mantle
- squid uses for disguise
(light shines downward, looks like moonlight)

Nature Reviews Molecular Cell Biology **3**; 685-695 (2002)

浙江大学 ZHEJIANG UNIVERSITY Quorum Sensing

BiS @ZJU Bioinformatics Server

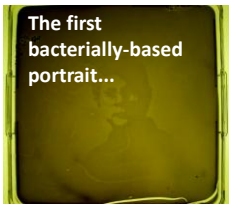
群体效应 **Quorum sensing: chemical-based bacterial communication**

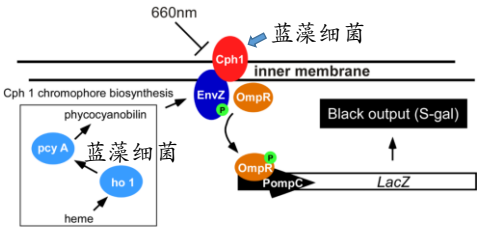

Vol 438|24 November 2005 nature

BRIEF COMMUNICATIONS

Engineering *Escherichia coli* to see light

These smart bacteria 'photograph' a light pattern as a high-definition chemical image.





How？

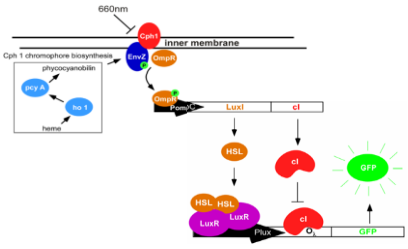
1. 蓝藻细菌HK Cph1 + 大肠杆菌EnvZ的Dhp与CA模块 = Cph8-EnvZ感光系统
2. 肠杆菌无感光蛋白，导入蓝藻细菌中的hol和pcyA基因，合成PCB藻胆青素
3. 红光下，PCB与Cph1共价二聚体化，阻止Cph1激酶活性
4. 引入lacZ基因，将OmpR调控目的基因ompC的启动子与之融合，使其表达依赖于OmpR并受光调控。

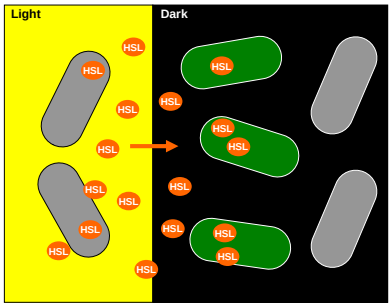
Results:

1. 当光照刺激， EnvZ的自磷酸化被抑制，从而OmpR不能被磷酸法激活，ompC启动子关闭， lacZ无法表达。
2. 相反，无光照的区域， EnvZ的自磷酸化、OmpR磷酸法顺利进行，ompC启动， lacZ表达，生成黑色沉淀。

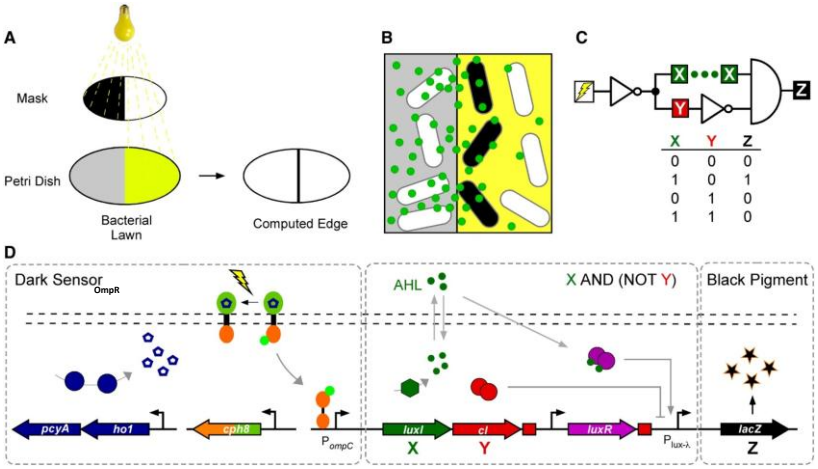
Edge detector

Now for the edge detector...





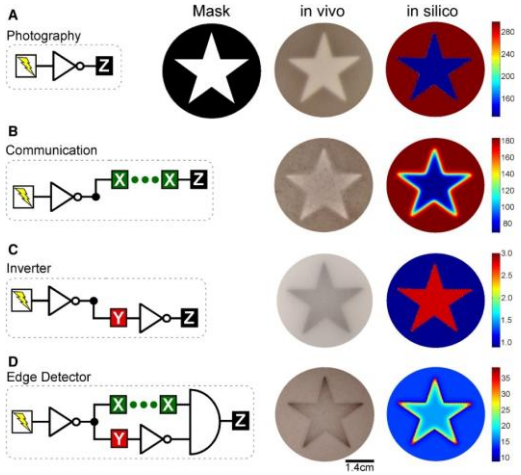
The edge detector circuit in more detail

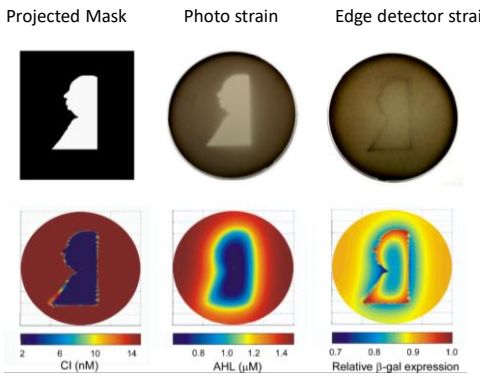


Tabor *et al. Cell 2009*

Edge detector

It works!



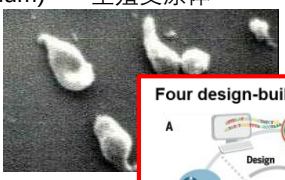


Tabor *et al. Cell 2009*

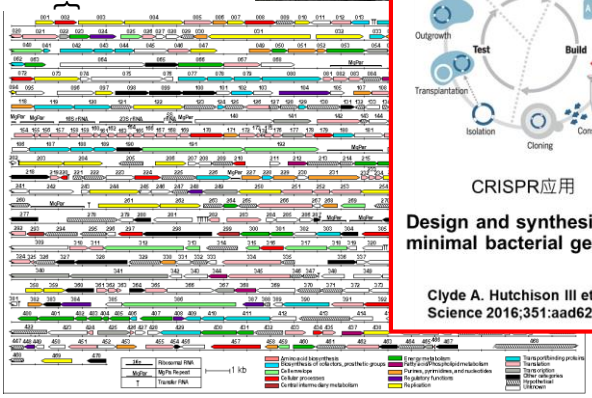
Minimal bacterium

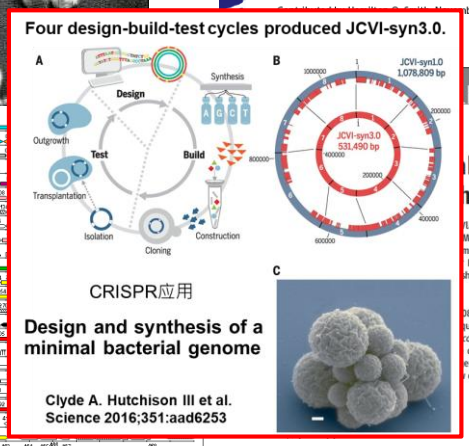
One of the smallest Genomes: *Mycoplasma genitalium*
(Small parasitic bacterium) 生殖支原体

Total genes: 521
Coding genes: 482
tRNA and rRNA: 39



Single Gene





Essential genes of a minimal bacterium

John I. Glass, Nacyra Assad-Garcia, Nina Alperovich, Shibu Yooseph, Matthew R. Lewis, Mahir Maruf, Clyde A. Hutchison III, Hamilton O. Smith*, and J. Craig Venter

Synthetic Biology Group, J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850

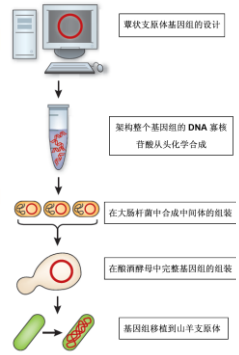
ber 18, 2005

PNAS (2006) 103: 425-430

l Cell Controlled nesized Genome

Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] mar,[1] Nacyra Assad-Garcia,[1] Lei Young,[1] Zhi-Qing Qi,[1] hanth P. Parmar,[2] Clyde A. Hutchison III,[2]

08–mega–base pair *Mycoplasma mycoides* quence information and its transplantation coides cells that are controlled only by the designed synthetic DNA sequence, ene deletions and polymorphisms, and cells have expected phenotypic properties



Gibson et al., (2010) Science, *329: 52-56*
Gibson et al., (2008) Science, *319: 1215-1220*

O2 transport protein

Vol 458|19 March 2009|doi:10.1038/nature07841 nature

ARTICLES

Design and engineering of an O_2 transport protein

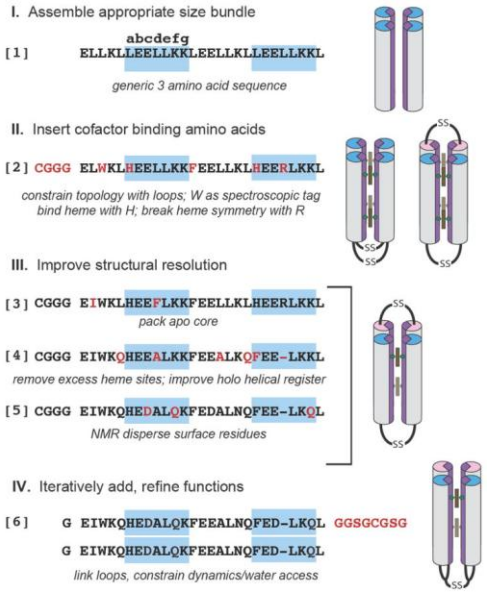
Ronald L. Koder[1]*†, J. L. Ross Anderson[1]*, Lee A. Solomon[1], Konda S. Reddy[1], Christopher C. Moser[1] & P. Leslie Dutton[1]

从一个极其简单的接近七烷重复序列（1）的三个氨基酸开始，并通过一系列氨基酸变化（红色）和试验中间产物（2-5）进行，这些中间产物经过测试以揭示功能特性和阐明单个氨基酸的作用

在2-5中，所示的螺旋序列由半胱氨酸二硫环连接，并在4个螺旋束中自组装

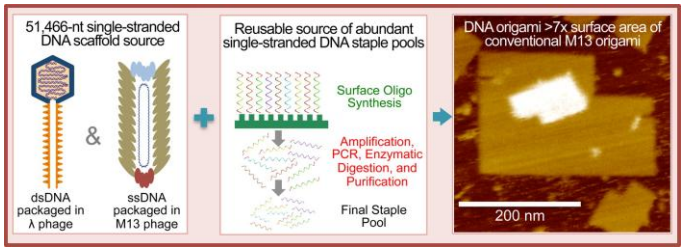
在6中，一个较长的环（红色）将两个相同的序列连接起来，环本身现在是二硫环连接的

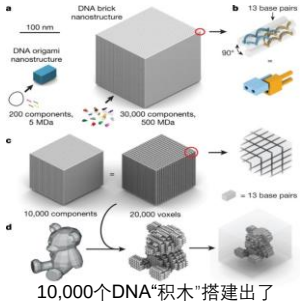
人工氧转运蛋白设计



DNA Origami

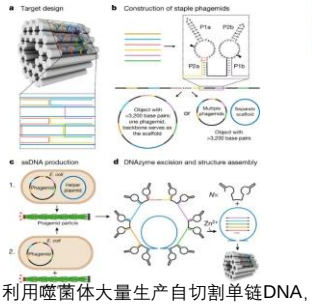
Toward Larger DNA Origami https://pubs.acs.org/doi/10.1021/nl502626s





10,000个DNA"积木"搭建出了一个泰迪熊形状的空腔

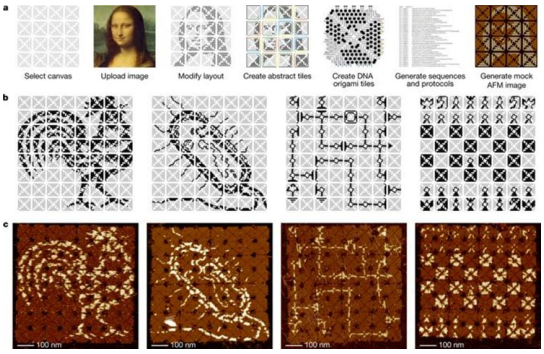
L L Ong *et al. Nature* **552,** 72–77 (2017)



利用噬菌体大量生产自切割单链DNA，将其组装成肉眼可见的纳米结构。

F Praetorius *et al. Nature* **552,** 84–87 (2017)

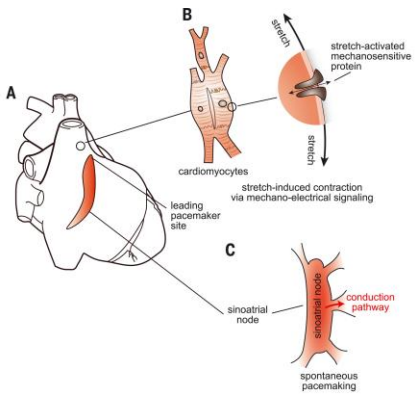
Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns

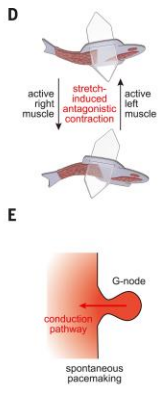


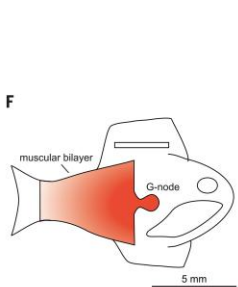
G Tikhomirov *et al. Nature* **552,** 67–71 (2017)
doi:10.1038/nature24655

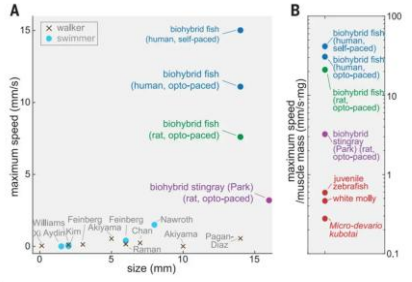
"如果要从一组特殊的DNA链开始，构建非常复杂的DNA纳米结构，就必须将组装过程分解成多个更简单的步骤。在每一步中，几个较小的结构单元都是在各自试管中组装出来，然后再将它们混合在一起，进一步自组装成更大的结构"

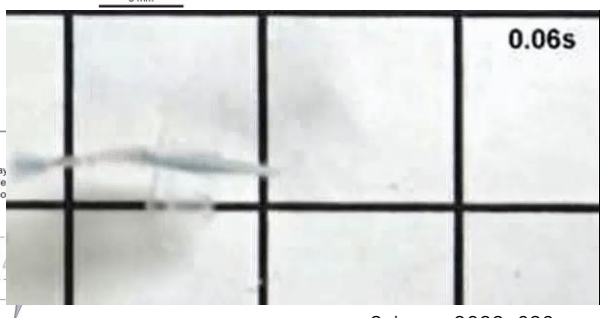
--- Lu Qian

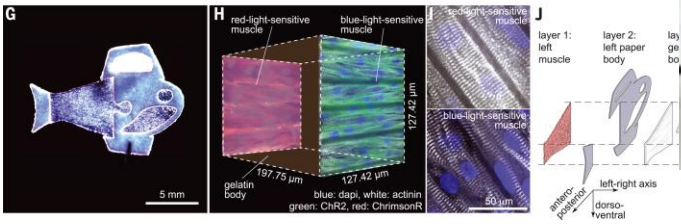








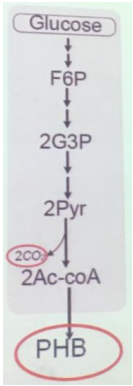


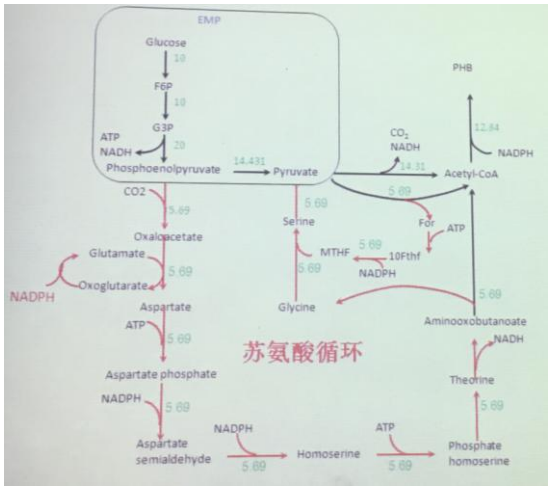


Science 2022: 639-647

基于代谢网络分析的细胞改造

- 由丙酮酸到PHB（**聚-β-羟丁酸**）前体乙酰CoA损失三分之一的碳源!

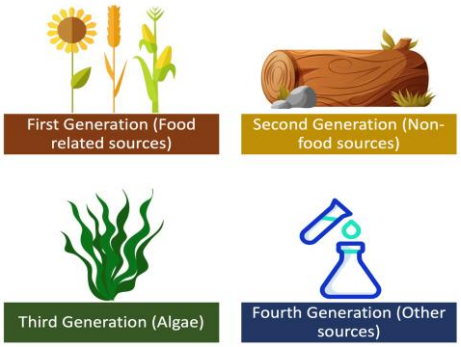




苏氨酸循环通过多余的能量和还原力回收部分AcCoA生成过程中损失的碳，使得PHB碳摩尔得率由67%提高到86%。

中科院马红武课题组

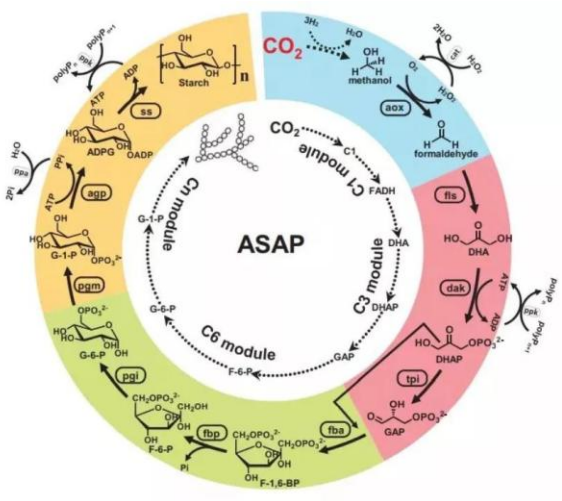
酵母-蓝藻共生体利用二氧化碳实现光合生长与高价值化合物生产



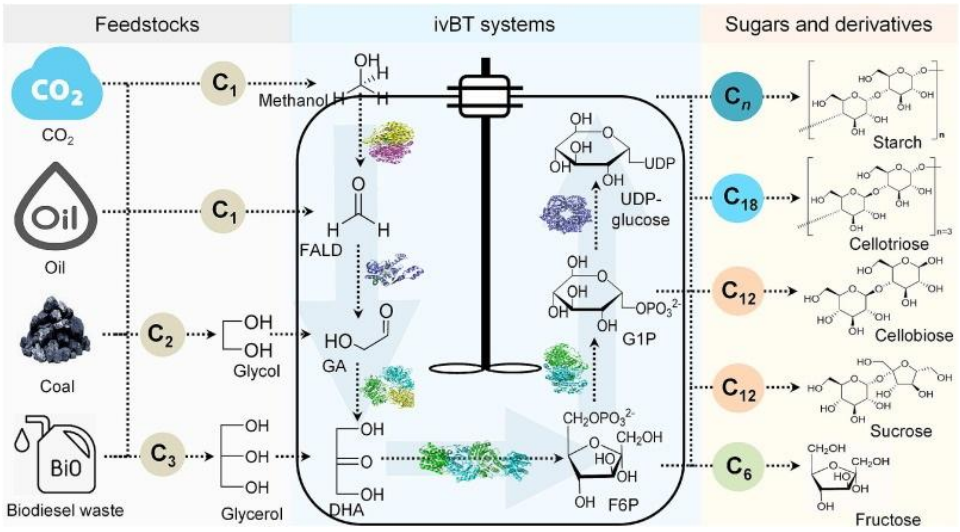
- 共生体工程化进展：通过代谢工程和基因线路设计，优化蓝藻与酵母的共生关系，使蓝藻在酵母细胞内共生，直接利用二氧化碳进行光合碳同化，生成葡萄糖等中间产物，支持酵母生长并合成烃类化合物（如生物柴油或药物前体）；这一突破避免了传统碳源依赖，显著提升二氧化碳利用率。
- 固碳途径的改进：天然固碳途径（如Calvin-Benson-Bassham循环）存在效率低和环境适应性差的问题，通过定向进化和代谢工程改造关键酶（如RuBisCO），并构建人工复合途径（如还原三羧酸循环），可增强微生物固碳能力，实现碳资源的高效回收。
- 高价值化学品生产应用：这些工程化系统可规模化生产燃料、药物前体等产品，确保碳源完全来自二氧化碳，为可持续能源和碳减排提供新方案；但商业化仍面临成本挑战，需进一步优化代谢通量和规模化工艺。

二氧化碳如何转化为淀粉

- 2019 年，美国加州大学伯克利分校杨培东教授团队在人工光合作用领域取得重要进展，成功地将二氧化碳非定向地转化为多种简单糖类化合物，引起学界广泛关注。
- 2021 年 9 月 24 日，中国科学院天津工业生物技术研究所马延和研究员团队在Science 上发表研究论文 Cell-free chemoenzymatic starch synthesis from carbon dioxide。
- 该研究巧妙地运用类似「搭房子」的方式，将淀粉转化过程拆分为 4 个模块，对这些模块逐个优化。在 31 种生物体的 62 个催化酶中筛选得到 10 种，从头设计，使其参与 11 步反应的人工固碳与淀粉合成途径。首先，在高密度氢能作用下将高浓度二氧化碳还原成碳一（C1）化合物。构建碳一聚合酶，依据化学聚糖反应原理将碳一化合物聚合为碳三（C3）化合物，进一步优化生物途径，将碳三化合物聚合为碳六（C6）化合物，最终合成直链和支链淀粉（Cn 化合物）。经过相关技术检测发现，该研究中人工合成的淀粉分子与天然淀粉分子的结构组成一致，且合成效率是传统工农业生产淀粉效率的 8.5 倍。



无细胞甲醇高效合成蔗糖



蔗糖

碳转化率达 86%
产物浓度 14 g/L

二羟基丙酮

Wang Y, et al. Sci Bull (Beijing). 2025 May 9:S2095-9273(25)00496-

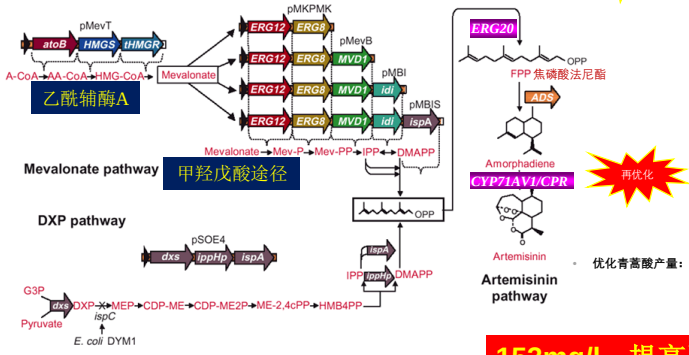
浙江大学 ZHEJIANG UNIVERSITY **Systems Biology & Synthetic Biology** BioS @ZJU Bioinformatics Server

ARTICLES nature biotechnology

Jay D Keasling小组 *Nature Biotechnology* **21**, 796 - 802 (2003)

Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids

Vincent JJ Martin[1,2,3], Douglas J Pitera[1,3], Sydnor T Withers[1], Jack D Newman[1] & Jay D Keasling[1]



idea

1）将大肠杆菌中常见的化学物质乙酰辅酶A转变成甲羟戊酸；
2）将甲羟戊酸转化成异戊烯焦磷酸酯(IPP) 或者二甲(基)丙烯焦磷酸酯 (DMAPP)；
3）将IPP或者DMAPP转化成法尼基环焦磷酸酯（FPP）；
4）利用*ADS*基因，将FPP转化成Amorphadiene。

IPP和DMAPP是所有类异戊二烯的通用前体物

How

1、净化大肠杆菌中的代谢内环境：引入*ispC*基因，利用其产物将大肠杆菌原有的DXP途径切断
2、合成和优化编码*Amorphadiene* 合成酶的基因*ADS*，克服植物基因在大肠杆菌中表达的困难。
3、将酿酒酵母中的甲羟戊酸途径转入大肠杆菌。
4、协调IPP有关的基因以平衡其合成与消耗，确保在其能够杀伤大肠杆菌以前及时转化为Amorphadiene。

再优化

1）优化FPP生物合成途径提高FPP的产量；
2）从黄花蒿中引入紫穗槐二烯合成酶(*ADS*)基因，其表达的产物将FPP转化为Amorphadiene；
3）通过比较基因组学分析得到来自于*A.annua* 的细胞色素P450氧化还原酶- CYP71AV1/CPR，克隆表达后实现Amorphadiene到青蒿酸的三步氧化还原反应而得到青蒿酸。

- 优化青蒿酸产量：

153mg/l，提高了近500倍

result

1）过量表达tHMGR，有效限制FPP向固醇的转化；2）通过甲硫氨酸可抑制启动子（PMET3）下调ERG9编码角鲨烯（squalene）合成酶，阻断FPP 向下合成固醇的支路，避免FPP在其他方面的不必要消耗；3）过表达*upc2-1*并结合ERG9的下调；4）所有被修饰基因都整合到染色体上确保稳定性；5）降低细胞密度，避免不必要的互相干扰和确保养分充足。

生物信息学 课程 Bioinformatics **教育&科研定位导向供讨论** 生物科学 101计划

大科学计划、大科学工程为大国之王道！

新突破、新机制，引领发展！ 新技术、新工具 新算法、新理论

工程化

Level 5: 重大科学计划或突破性创新研究?

Level 4: 基于数据分析的生物学发现?

Level 3: 开发方法与软件工具，技术开发

Level 2: 简单数据分析，技术应用

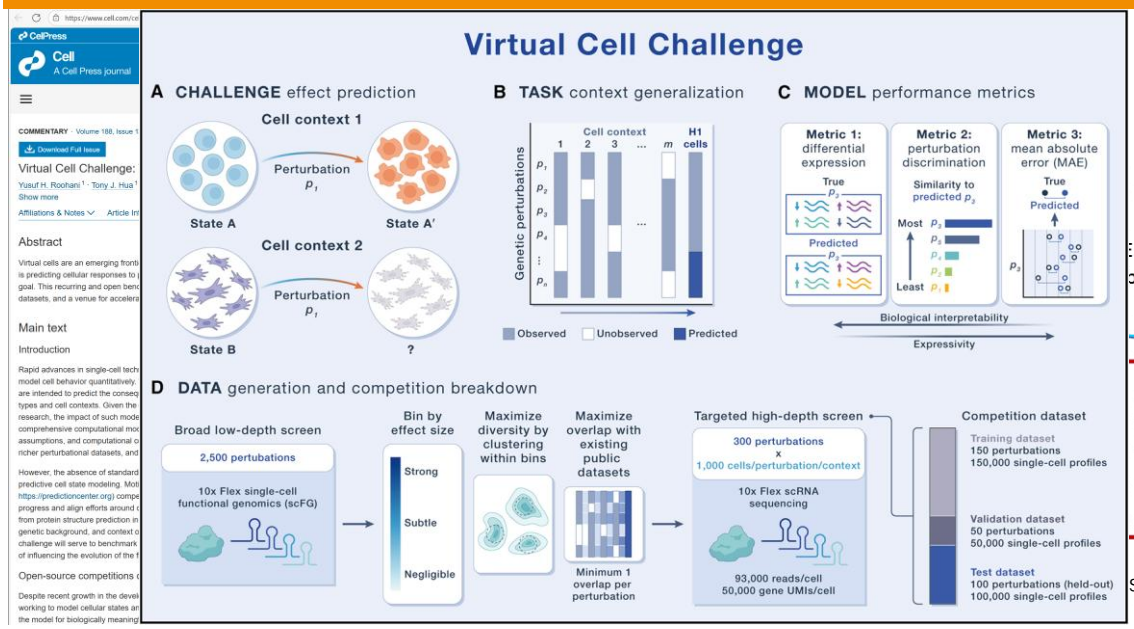
Level 1: 为建模而建模，方法探索

实验生物学 ← 生物信息学 → 计算机、数学

问题驱动 数据驱动

105

Virtual Cell Modeling



stimation
basi

s://vcell.org/

结语

“一场与工业革命和以计算机为基础的革命 有相同影响力的变化正在开始。下一个伟大时代将是基因组革命时代，它现在处于初 期阶段”。

——《第三次技术革命》

当前，多组学和生物信息学的发展已经进入又一个技术革命的时代——人工智能时代，生物信息学新方法新技术正在蓬勃发展，将对生物学、医学、药学、农业科学等领域产生巨大的影响，我们必定能够揭示各种生命现象的奥秘，并带动多个学科的跨越式发展，极有可能引发新的产业革命。