系统生物学

(Systems Biology)

马彬广

代谢系统建模

(流平衡分析简介)

（第九讲）

流平衡分析
(Flux Balance Analysis)

流平衡分析的含义

- 什么是流（Flux）？
- 流就是代谢网络中化学反应的反应速率。
- 在稳态下，各代谢物浓度不变，反应速率满足一定的分布（流分布）。
- 什么流（量）？分子类型转化流，单位时间内的转化量。

流平衡分析的含义

- 何谓平衡（balance）？

➢ 平衡就是限制（约束）。

➢ 约束来自两个方面：

基本物理规律的约束，物质不灭，能量守恒等。

边界条件的约束，底物的供应量（上限），产物的生成量（下限）等。

➢ 因此，流平衡分析又称为基于约束的模型（constraint-based model）.

流平衡分析的含义

- 优化目标（Phenotype）

➢生长最快（生物量Biomass最大）
➢某代谢物（ATP、次级代谢物）产量最大或最小（致死）

- 确定优化目标下的流分布，即，线性约束下的优化问题（Linear Programming问题）

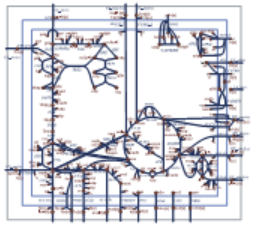
流平衡分析的用途

- 表型与代谢相关的问题，如：
- 生长速率模拟预测
- 次生代谢产物最大化
- 微生物致死的研究（抗生素研发）
- 动植物特定组织或发育过程中特定阶段的代谢分析等。

流平衡分析的流程

- 构建代谢网络；
- 把代谢网络表示成数学形式（计量学矩阵）；
- （基于稳态假设），根据物料平衡，得到线性方程组（约束条件）；
- 定义优化目标（函数）；
- 求解约束下的优化问题，得到流分布和相应的目标值。

a Genome-scale metabolic reconstruction



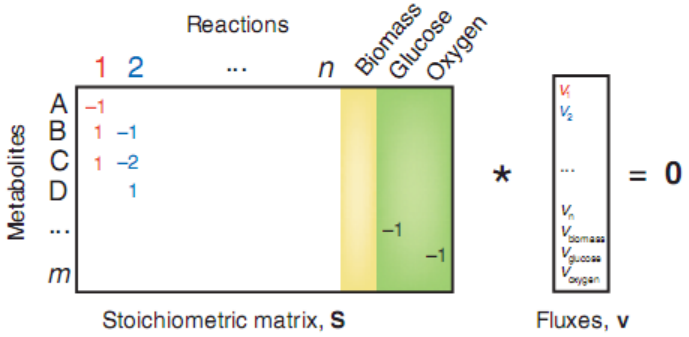
$A \leftrightarrow B + C$ Reaction 1

$B + 2C \rightarrow D$ Reaction 2

...

Reaction *n*

b Mathematically represent metabolic reactions and constraints



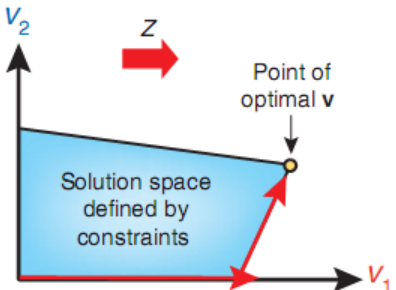
c Mass balance defines a system of linear equations

$$-v_1 + \ldots = 0$$
$$v_1 - v_2 + \ldots = 0$$
$$v_1 - 2v_2 + \ldots = 0$$
$$v_2 + \ldots = 0$$
etc.

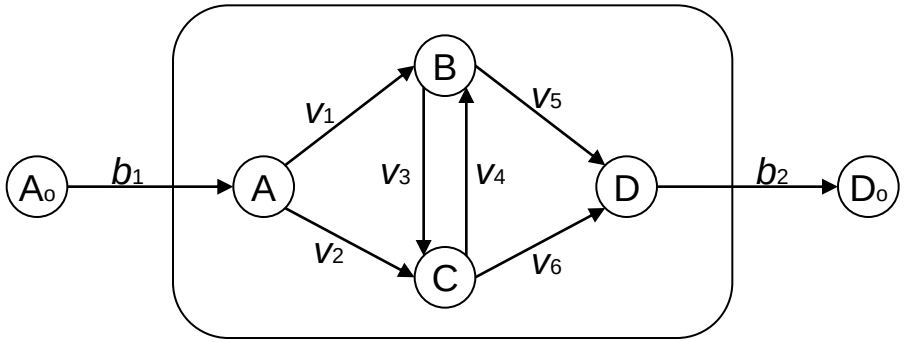
d Define objective function ($Z = c_1 * v_1 + c_2 * v_2 \ldots$)

To predict growth, $Z = v_{biomass}$

e Calculate fluxes that maximize *Z*



流平衡分析的举例



$$S = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \begin{array}{cccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 \end{array} \\ \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix} \end{array}$$

问题表述：

Maximize $Z = c^T v$

Subject to $S \cdot v = 0$

$\alpha \le v_i \le \beta$ for all reactions i

$$Z = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1) \cdot (v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad b_1 \quad b_2)^T$$

$$\alpha = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)^T$$

$$\beta = (2 \quad 10 \quad 4 \quad 6 \quad 10 \quad 8 \quad 100 \quad 100)^T$$

流平衡分析的举例

MatLab 求解：

$$[v, Z] = \text{linprog}(c,\ Aeq,\ beq,\ S,\ b,\ \alpha,\ \beta)$$

LP problem:

Minimize $Z = c^T \times v$

Subject to $Aeq \times v \le b_{eq}$

$S \cdot v = b$

$\alpha \le v \le \beta$

令Aeq为单位阵，b = β，则：

$$A_{eq} \cdot v \le b_{eq} \quad \Leftrightarrow \quad v \le b$$

Code:

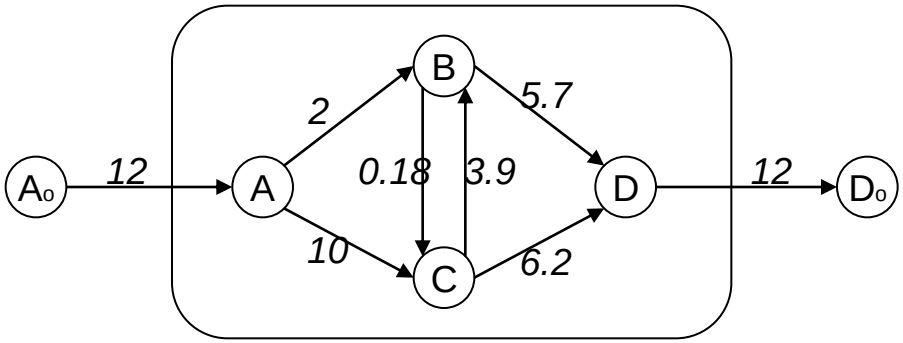
```
>> S = [-1, -1, 0, 0, 0, 0, 1, 0;
        1, 0, -1, 1, -1, 0, 0, 0;
        0, 1, 1, -1, 0, -1, 0, 0;
        0, 0, 0, 0, 1, 1, 0, -1]

>> b = [0; 0; 0; 0]
>> Alpha = [0; 0; 0; 0; 0; 0; 0; 0;]
>> beta = [2, 10, 4, 6, 10, 8, 100, 100]
>> c = [0, 0, 0, 0, 0, 0, 0, 1]
>> Aeq = eye(8)
>> [v, Z] = linprog(-c, Aeq, beta, S, b, Alpha, beta)
```

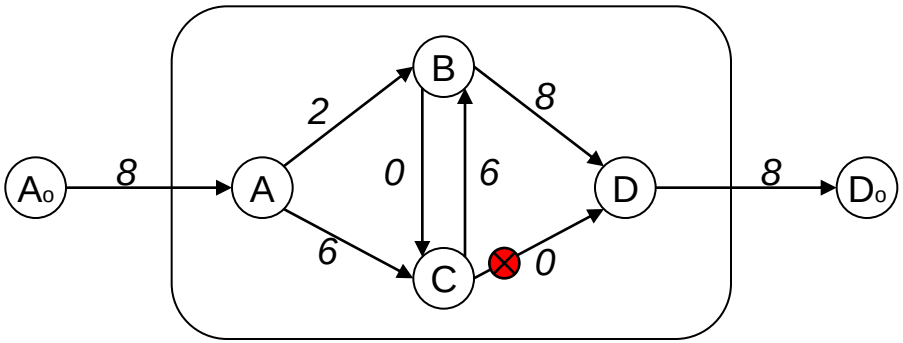
v = 2.0000 10.0000 0.1822 3.9137 5.7315 6.2685 12.0000 12.0000

Z = -12.0000

流平衡分析的举例

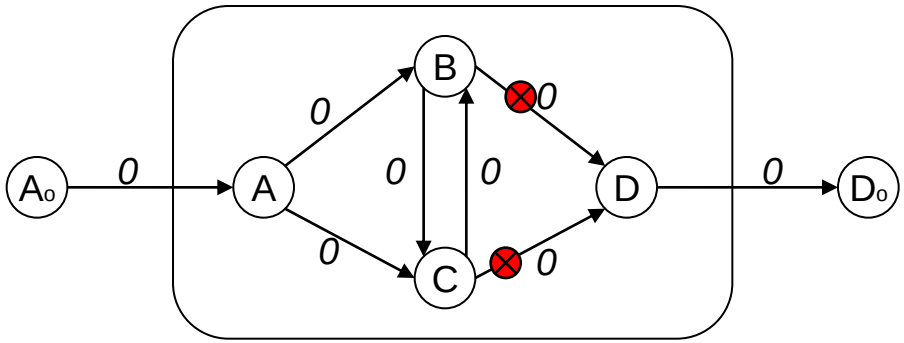
Wild Type：





V_6 Knockout

V_5 and V_6 Knockout



流平衡分析的进展

- 经典文献（里程碑）
- FBA模型的提出（AEM-Varma1994）
- DFBA模型的提出（BioPhysiJ-Mahadevan2002）
- MOMA模型的提出（PNAS-Segre2002）
- M_DFBA模型的提出（MSB-Luo2006）
- idFBA模型的提出（PlosCB-Lee2008）

流平衡分析的进展

- FBA模型（AEM-Varma1994）

$$S \cdot v = b \qquad (1)$$

$$\text{Minimize } Z = -V_{gro} \qquad (2)$$

$$\sum_{\text{all } M} d_M \cdot M \xrightarrow{V_{gro}} \text{biomass} \qquad (3)$$

流平衡分析的进展

- DFBA模型

DOA form:

$$\underset{\mathbf{z}(t),\mathbf{v}(t),X(t)}{\text{Max}} \quad \hat{w}_{\text{end}}\Phi(\mathbf{z},\mathbf{v},X)\big|_{t=t_f} + \hat{w}_{\text{ins}}\sum_{j=0}^{M}\int_{t_0}^{t_f} L(\mathbf{z},\mathbf{v},X(t))\delta(t-t_j)\,dt$$

$$A\cdot v = 0, \qquad (1)$$

$$\text{s.t.} \quad \frac{d\mathbf{z}}{dt} = A\mathbf{v}X$$

$$\frac{dz}{dt} = A\mathbf{v}X, \quad \frac{dX}{dt} = \mu X, \quad \mu = \Sigma w_i v_i, \qquad (2)$$

$$\frac{dX}{dt} = \mu X$$

$$\mu = \Sigma w_i v_i$$

$$t_j = t_0 + j\frac{t_f - t_0}{M} \quad j = 0\cdots M$$

$$c(\mathbf{v},\mathbf{z}) \le 0 \quad |\dot{\mathbf{v}}| \le \dot{\mathbf{v}}_{\max} \quad \forall t \in [t_0, t_f]$$

$$\mathbf{z} \ge 0 \quad X \ge 0 \quad \forall t \in [t_0, t_f]$$

$$\mathbf{z}(t_0) = \mathbf{z}_0 \quad X(t_0) = X_0, \qquad (3)$$

（BioPhysiJ-Mahadevan2002）

流平衡分析的进展

- DFBA模型

$$A \cdot v = 0, \qquad (1)$$

$$\frac{dz}{dt} = A\mathbf{v}X, \quad \frac{dX}{dt} = \mu X, \quad \mu = \Sigma w_i v_i, \qquad (2)$$

SOA form:

$$\begin{aligned}
&\underset{\mathbf{v}(t)}{\text{Max}}\ \Sigma w_i v_i(t) \\
\text{s.t.}\quad &\mathbf{z}(t+\Delta T) \geq 0 \quad \mathbf{v}(t) \geq 0 \\
&\hat{c}(\mathbf{z}(t))\mathbf{v}(t) \leq 0 \quad \forall t \in [t_0, t_f] \\
&|\mathbf{v}(t) - v(t-\Delta T)| \leq \dot{\mathbf{v}}_{max}\Delta T \quad \forall t \in [t_0, t_f] \\
&\mathbf{z}(t+\Delta T) = \mathbf{z}(t) + A\mathbf{v}\Delta T \\
&X(t+\Delta T) = X(t) + \mu X(t)\Delta T, \qquad (4)
\end{aligned}$$

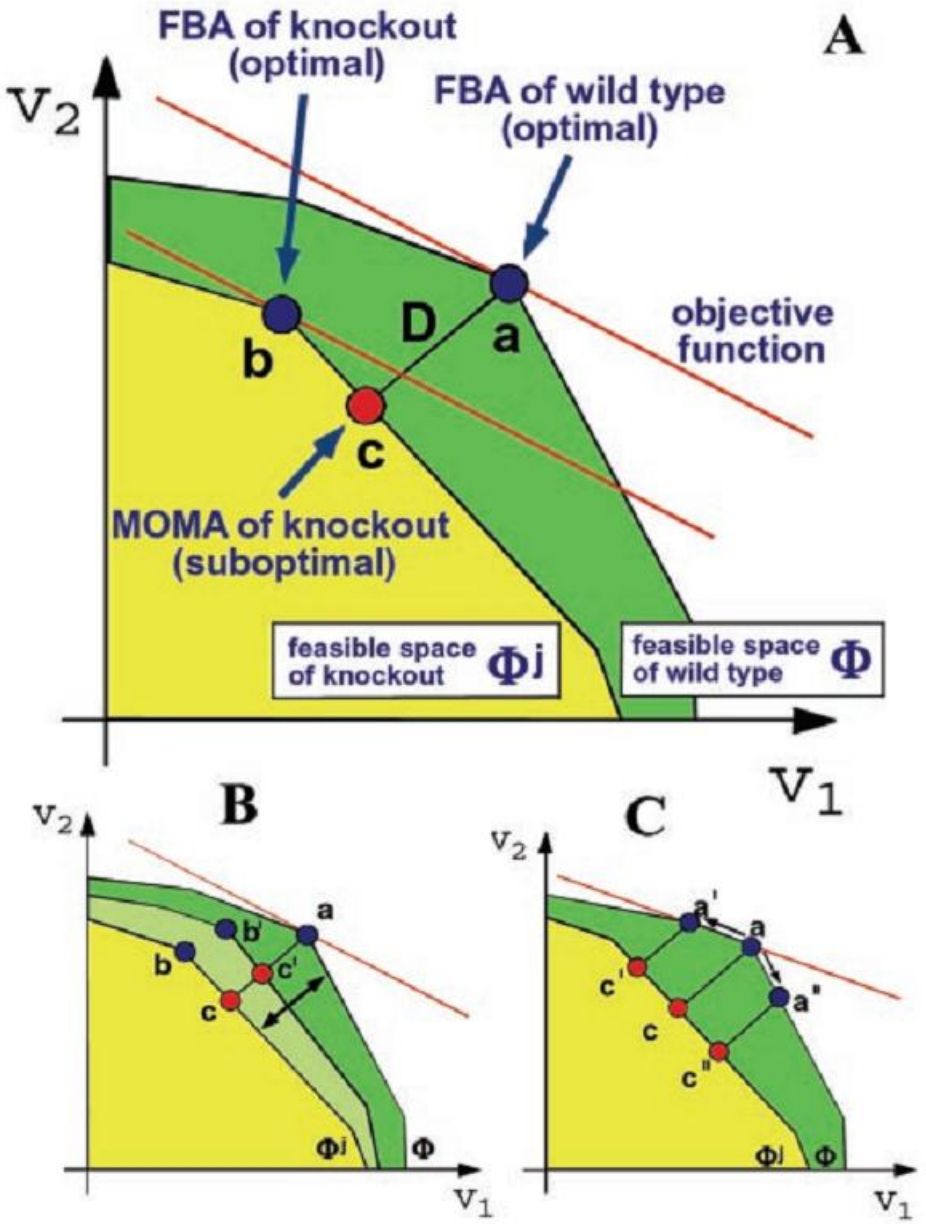
（BioPhysiJ-Mahadevan2002）

流平衡分析的进展

- MOMA模型

$$D(\mathbf{w}, \mathbf{x}) = \sqrt{\sum_{i=1}^{N} (w_i - x_i)^2} \qquad [4]$$

（PNAS-Segre2002）



流平衡分析的进展

- M_DFBA模型

$$\frac{dX}{dt} = S \cdot V + F(\mathrm{Out}_X - \&_X \mathrm{In}_X)$$

(MSB-Luo2006)

$$\text{Objective} = \mathrm{Min} \sum_{j=1}^{M} \int_{t_0}^{t_f} \sqrt{\sum_{i=1}^{N} (x_{i,j} - x_{i,j-1})^2 \delta(t - t_j)} dt$$

s.t.

$$\frac{dx^{\mathrm{Gluc}}}{dt} = S_{\mathrm{Gluc}} \cdot V^{\mathrm{Gluc}} + F(\mathrm{Out}_{\mathrm{Gluc}} - \&_{\mathrm{Gluc}} \mathrm{In}_{\mathrm{Gluc}})$$

$$\frac{dx^{\mathrm{Gly}}}{dt} = S_{\mathrm{Gly}} \cdot V^{\mathrm{Gly}}$$

$$\frac{dx^{\mathrm{FA}}}{dt} = S_{\mathrm{FA}} \cdot V^{\mathrm{FA}} + F(\mathrm{Out}_{\mathrm{FA}} - \&_{\mathrm{FA}} \mathrm{In}_{\mathrm{FA}})$$

$$\frac{dx^{\mathrm{Lac}}}{dt} = S_{\mathrm{Lac}} \cdot V^{\mathrm{Lac}} + F(\mathrm{Out}_{\mathrm{Lac}} - \&_{\mathrm{Lac}} \mathrm{In}_{\mathrm{Lac}})$$

$$\frac{dx^{\mathrm{O_2}}}{dt} = S_{\mathrm{O_2}} \cdot V^{\mathrm{O_2}} + F(\mathrm{Out}_{\mathrm{O_2}} - \&_{\mathrm{O_2}} \mathrm{In}_{\mathrm{O_2}})$$

$$\frac{dx^{\mathrm{PC}}}{dt} = S_{\mathrm{PC}} \cdot V^{\mathrm{PC}}$$

$$\frac{dx^{\mathrm{ATP}}}{dt} = S_{\mathrm{ATP}} \cdot V^{\mathrm{ATP}}$$

$$X_0 = [x_0^{\mathrm{Gluc}}, x_0^{\mathrm{Gly}}, x_0^{\mathrm{FA}}, x_0^{\mathrm{Lac}}, x_0^{\mathrm{O_2}}, x_0^{\mathrm{PC}}, x_0^{\mathrm{ATP}}]^{\mathrm{T}}$$

$$\sqrt{F} X_0 \leqslant [x^{\mathrm{Gluc}}, x^{\mathrm{Gly}}, x^{\mathrm{FA}}, x^{\mathrm{Lac}}, x^{\mathrm{O_2}}, x^{\mathrm{PC}}, x^{\mathrm{ATP}}]^{\mathrm{T}}$$

$$S_{\mathrm{Gluc}} \cdot V^{\mathrm{Gluc}} < V_{\max}^{\mathrm{Gluc} \to \mathrm{GP}}$$

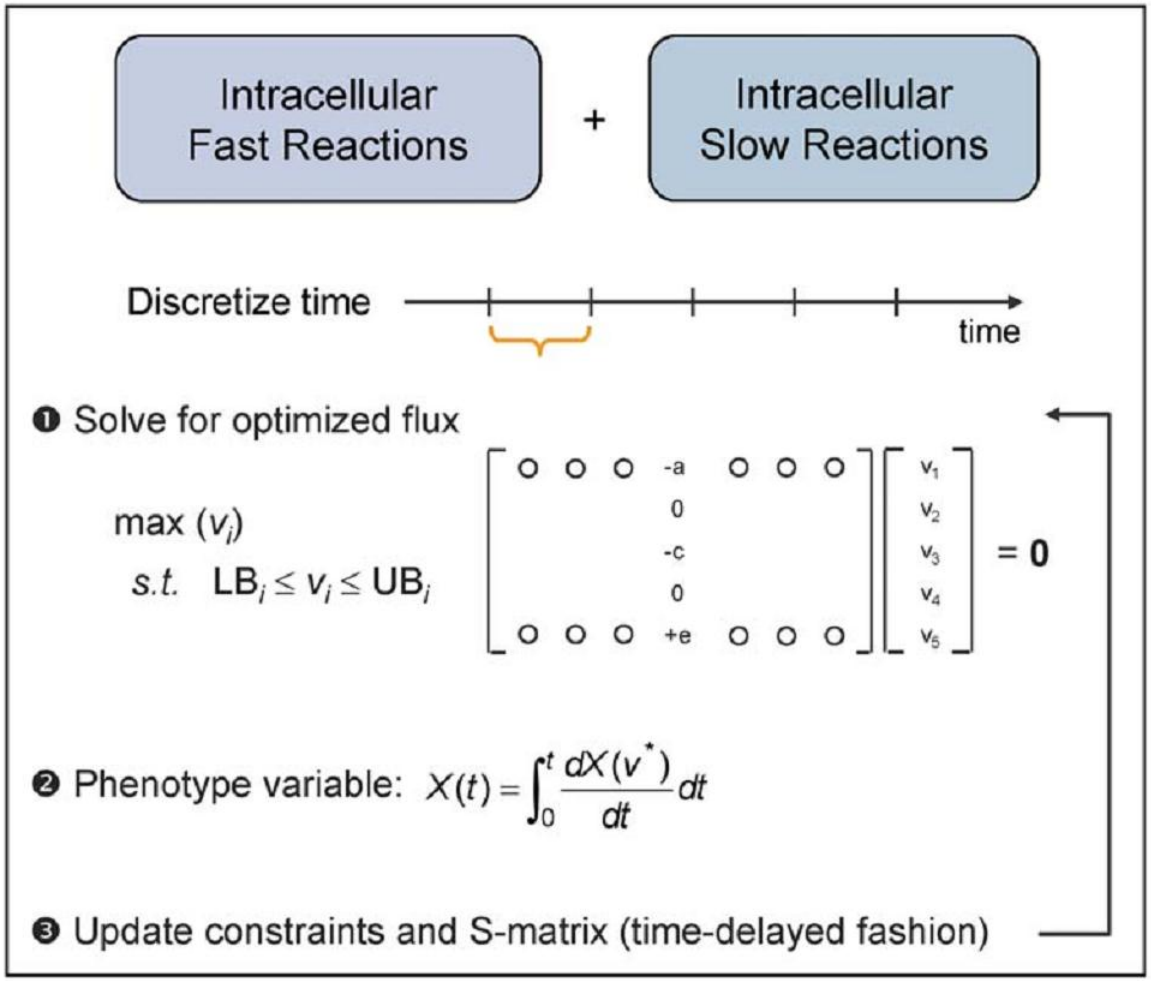
$$S_{\mathrm{O_2}} \cdot V^{\mathrm{O_2}} < V_{\max}^{\mathrm{O_2} \to \mathrm{H_2O}}$$

$$\frac{v_1 + v_4 + v_6}{v_1 + v_3 + v_4 + v_6} < 0.7175$$

$$S_{\mathrm{ATP}} \cdot V^{\mathrm{ATP}} \geqslant \alpha \times (\sqrt{F} + F)/2$$

流平衡分析的进展

- idFBA模型



（PlosCB-Lee2008）

流平衡分析的进展

- 发展趋势
- 引入更多的约束，使模型跟趋近于真实；
- 与刺激响应和信号转导相结合；
- 与转录调控和基因表达相结合；

流平衡分析的工具

- COBRA

Box 2 Tools for FBA

FBA computations, which fall into the category of constraint-based reconstruction and analysis (COBRA) methods, can be performed using several available tools[27-29]. The COBRA Toolbox[11] is a freely available Matlab toolbox (http://systemsbiology.ucsd.edu/Downloads/Cobra_Toolbox) that can be used to perform a variety of COBRA methods, including many FBA-based methods. Models for the COBRA Toolbox are saved in the Systems Biology Markup Language (SBML)[30] format and can be loaded with the function 'readCbModel'. The *E. coli* core model used in this Primer is available at http://systemsbiology.ucsd.edu/Downloads/E_coli_Core/.

In Matlab, the models are structures with fields, such as 'rxns' (a list of all reaction names), 'mets' (a list of all metabolite names) and 'S' (the stoichiometric matrix). The function 'optimizeCbModel' is used to perform FBA. To change the bounds on reactions, use the function 'changeRxnBounds'. The **Supplementary Tutorial** contains examples of COBRA toolbox code for performing FBA, as well as several additional types of constraint-based analysis.

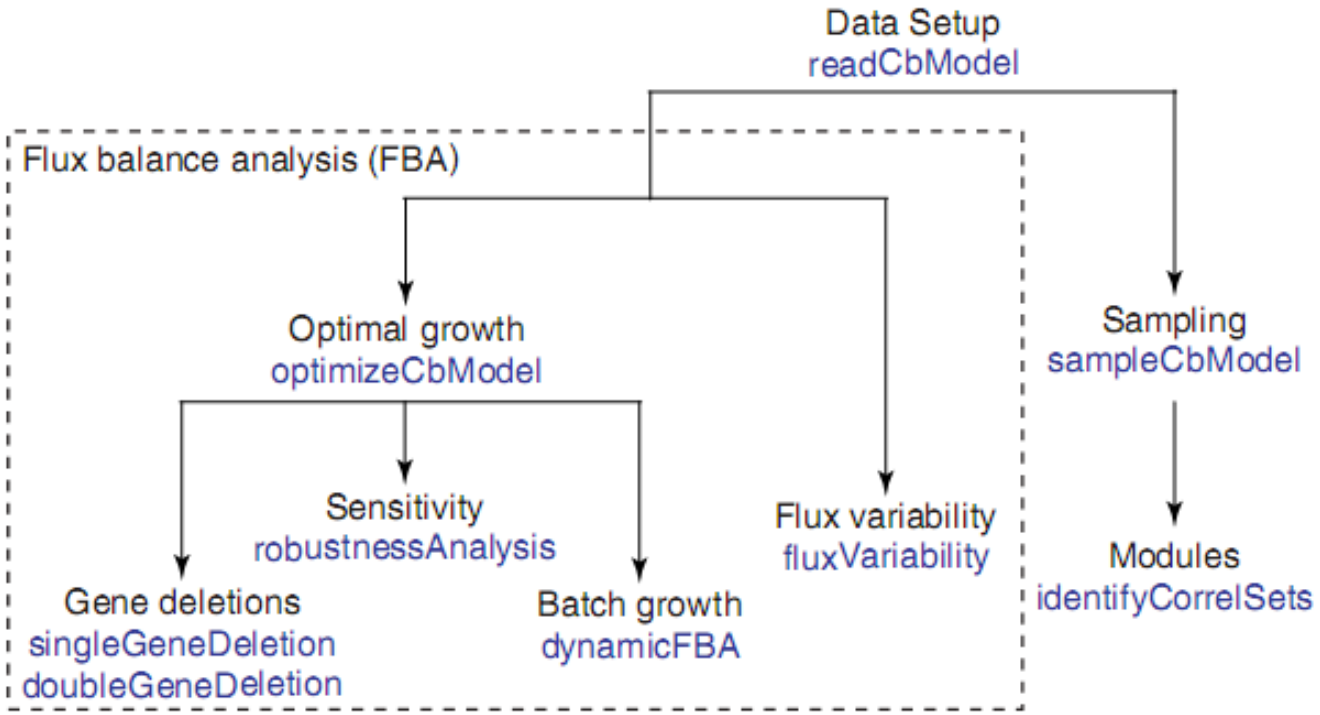
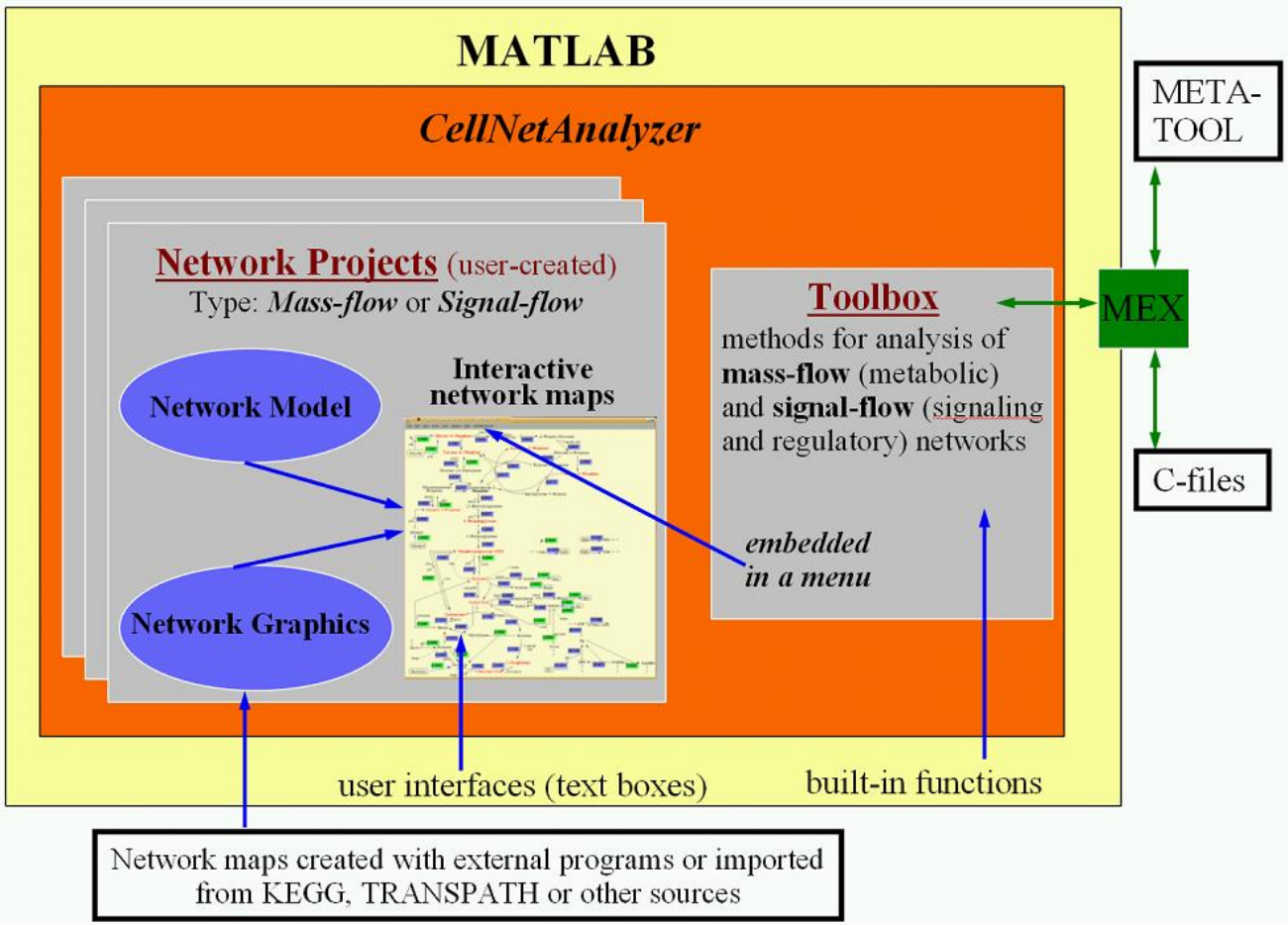


Figure 3 | The workflow for using the COBRA Toolbox. Functions at lower levels of the hierarchy rely on the internal implementation of the methods at higher levels. For example, the gene deletion functions perform an optimal growth calculation for each possible metabolic network with the reaction(s) associated with one or two genes removed. The exception to this is that finding network modules requires the user to input a previously calculated set of samples from the sampling function. On the whole, most of the methods rely on the principles of FBA.

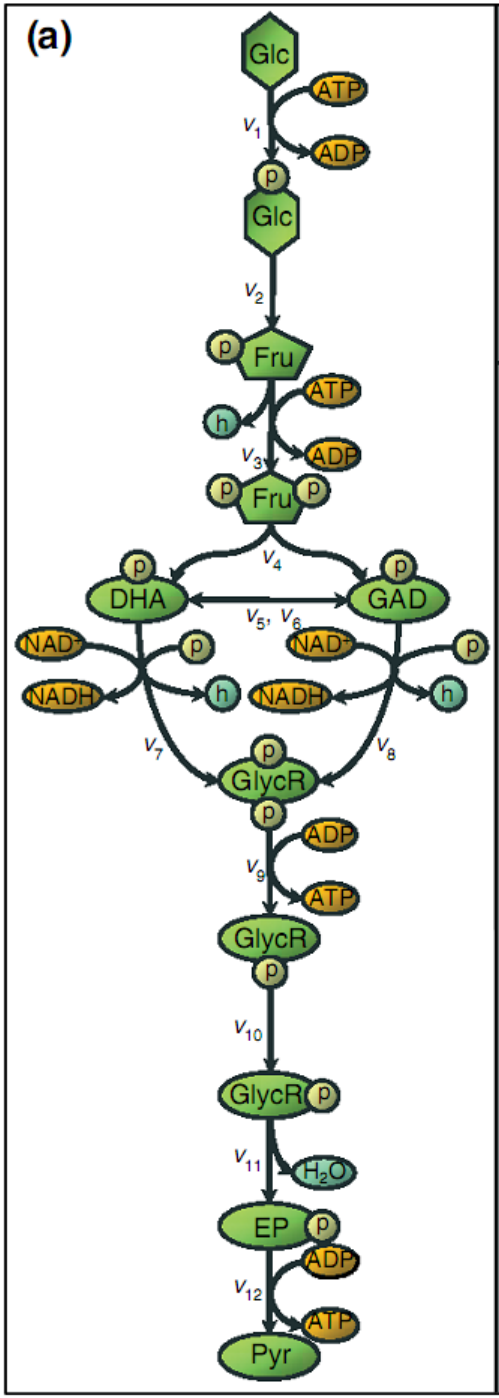
流平衡分析的工具

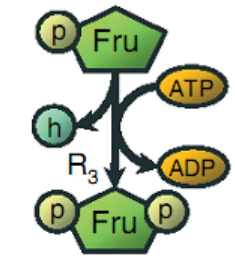
- CellNetAnalyser（前身FluxAnalyser）



流平衡分析的工具

- BioMet（Web Interface）
- FASIMU (Batch Interface)
- 其它通用代谢网络建模工具（SBW,COPASI等）
- 可视化工具





	R_3
F6P	-1
FDP	+1
ATP	-1
ADP	+1
H^+	+1

(c)

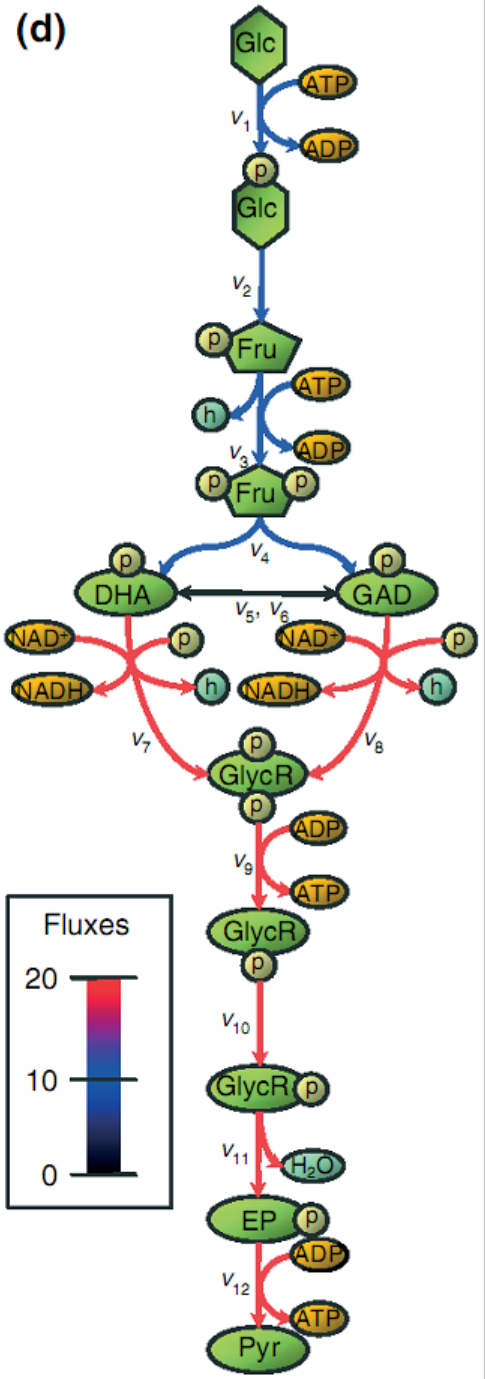
$$\max \quad Z = \mathbf{c} \cdot \mathbf{v}$$

$$\begin{cases} \dfrac{d\mathbf{C}}{dt} = \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\ \quad = \sum_{j=1}^{size(\mathbf{N})} \left(S_{i,j} \cdot v_j \right) = 0 \quad \forall i \in \mathbf{M} \\ LB_j \le v_j \le UB_j \quad \forall j \in \mathbf{N} \end{cases}$$

Vector **c** is a set of weighting coefficients on the fluxes contained in **v**; **C** denotes species concentrations; **M** is the set of components; **N** is the set of reactions; and *LB* and *UB* correspond to lower and upper flux bounds.

Objective function:

$$\max \quad Z = v_{12}$$

Flux bounds:

$$LB_1 = 10 \qquad LB_{2..12} = 0$$
$$UB_1 = 10 \qquad UB_{2..12} = 100$$