Genomics & Computational Biology

Section 10: Metabolic Networks

Lan Zhang and Xiaoxia (Nina) Lin

Nov. 25th, 2003

Metabolic Networks

- ***Why study biological networks?***
  - Genomics and proteomics tell us the 'parts' of the cell, but they give limited information about their interactions and co-dependencies. Ideally, we'd like to understand the interactions well enough that we could predict the system's behavior computationally – "in silico biology".
- ***Applications:***
  - Phenotype prediction
  - Bioengineering
  - Synthetic media design
- ***"In Silico Biology"***
  - Uses computational representations of a cell or organism to model the organism's behavior.
  - Kinetic models (e.g., the red blood cell model discussed in class)
  - **Flux Balance Analysis:** based on the stoichiometry of biochemical reactions associated with different cellular components and mass conservation

Kinetics of Enzymatic Reactions

- **Reaction rate (flux)**: the increase/decrease in molar concentration of product/reactant of a reaction per unit time

$$v = \frac{1}{S_i}\frac{d[X_i]}{dt}$$

where $[X_i]$: the molar concentration of metabolite X_i

S_i : the stoichiometric coefficient of X_i,

negative for reactants and positive for products

e.g.

$$\mathrm{A} \rightarrow 2\mathrm{B}$$

$$v = -\frac{d[A]}{dt} = \frac{1}{2}\frac{d[B]}{dt}$$

- Reaction kinetics: generally non-linear functions of
  - Temperature
  - Concentrations of substrates and enzymes

e.g. Michaelis-Menten equation

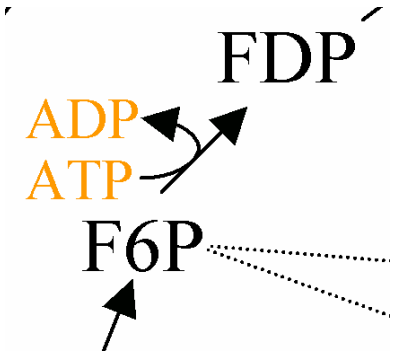
$$S \xrightarrow{E} P$$

$$v = \frac{V_{\max}[S]}{[S] + K_M}$$

where $V_{\max}$ = Maximum rate

K_M = Michaelis - Menten constant (measures enzyme/substrate affinity)

e.g. Phosphofructokinase



$$v_{PFK} = \frac{v_{mx}^{PFK}}{N_{PFK}} \left(\frac{F6P / K_{F6P}^{PFK}}{1 + F6P / K_{F6P}^{PFK}} \right) \left(\frac{Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}}{1 + Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}} \right)$$

$$N_{PFK} = 1 + L_0^{PFK} \frac{\left(1 + ATP_{free} / K_{ATP}^{PFK}\right)^4 \left(1 + Mg / K_{Mg}^{PFK}\right)^4}{\left(1 + AMP / K_{AMP}^{PFK}\right)^4 \left(1 + F6P / K_{F6P}^{PFK}\right)^4}$$

Mass Action Kinetics

- The rate of a reaction is proportional to the product of the concentrations of the reactants

e.g.

$$A + B \xrightarrow{k} C$$

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

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

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

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

$$A + 2B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C$$

$$v_1 = k_1[A][B]^2; \qquad v_{-1} = k_{-1}[C]$$

$$\frac{d[A]}{dt} = -k_1[A][B]^2 + k_{-1}[C]$$

$$\frac{d[B]}{dt} = -2k_1[A][B]^2 + 2k_{-1}[C]$$

$$\frac{d[C]}{dt} = k_1[A][B]^2 - k_{-1}[C]$$

Michaelis-Menten Kinetics

Assumptions of the mechanism:

- substrate and enzyme form intermediate complex
- the intermediate complex is in steady state with respect to the overall reaction rate
- the reverse reaction of the final step is neglectable

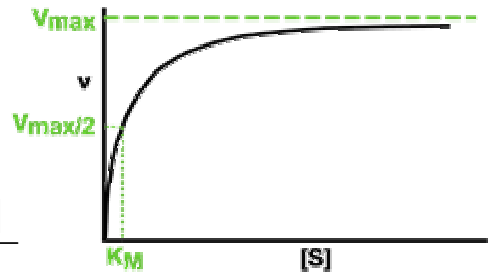
$$\frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] = 0$$

$$E_t = [E] + [ES]$$

$$\Rightarrow [ES] = \frac{k_1[E_t][S]}{k_1[S] + k_{-1} + k_2}$$

$$\Rightarrow v = \frac{d[P]}{dt} = k_2[ES] = \frac{k_1 k_2[E_t][S]}{k_1[S] + k_{-1} + k_2} = \frac{k_2[E_t][S]}{[S] + (k_{-1} + k_2)/k_1} \equiv \frac{V_{\max}[S]}{[S] + K_M}$$

where $V_{\max} \equiv k_2[E_t]; K_M = (k_{-1} + k_2)/k_1$



Kinetic Models of Metabolic Networks

- Simulation of dynamic processes
- Use of ordinary differential equations (ODEs) correlating the metabolite concentrations and their change rates

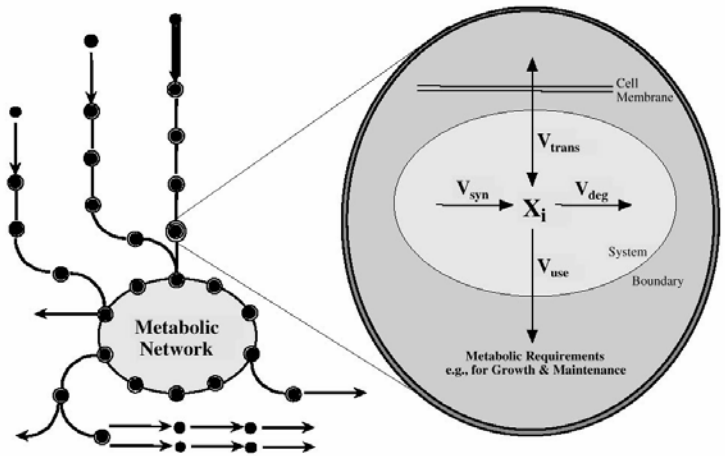
  e.g. Human Red Blood Cell ODE model:

  44 rate expressions with ~ 5 constants each → ~ 200 parameters
- Mass balance equations are linear functions of the reaction rates, BUT are non-linear functions of the actual metabolite concentrations.
- Limitations
  - reaction mechanisms or exact reaction rate expressions are generally unavailable
  - kinetic parameters are hard to measure/obtain
  - *in vitro* values of kinetic parameters may not represent the *in vivo* values
  - difficult to treat mathematically

Flux Balance Analysis (FBA)

Four basic types of metabolic flux rates affect the concentration of any given metabolite pool:

- v_{syn} = synthesis rate
- v_{deg} = degradation rate
- v_{trans} = transport rate (import and export)
- v_{use} = consumption rate (use rate required for growth and maintenance)



The **net flux rate** for a metabolite X_i is:

$$\frac{d[X_i]}{dt} = (V_{syn} - V_{deg} - V_{use}) - V_{trans} = (S_{ij} v_j) - b_i$$

where v_j is the jth reaction rate, b_i is the transport rate, S_{ij} is the "stoichiometric matrix" = moles of metabolite i produced in reaction j

Steady State Analysis

$$\frac{d\mathbf{X}}{dt} = \mathbf{S} \cdot \mathbf{v} - \mathbf{b} = 0$$

- The steady state assumption
  - **Time constants for metabolic reactions are very fast (sec - min) compared to cell growth and culture fermentations (hrs)**
  - **There is not a net accumulation of metabolites in the cell over time.**
- Removes **X** and *t* as variables; helps to get around the problem of non-linearity
- Additional constraints
  - specific measurement – typically for uptake rxns
  - the flux level through certain reactions is known
  - maximal internal flux

FBA

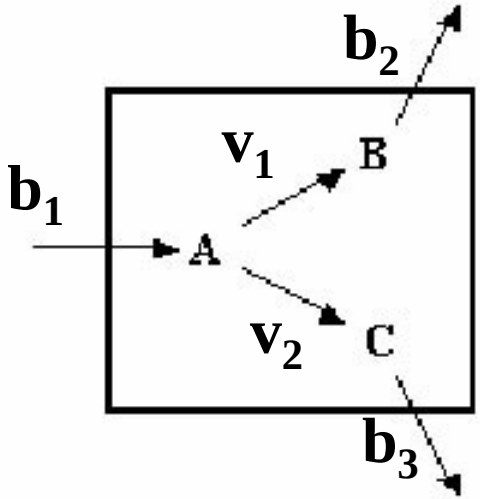
- When enough measurements are available, we can solve the steady state mass balance equations (**S** is available from genomic analysis and **b** can be determined experimentally)
- Underdetermined system
  - In general, not enough measurements of the metabolic fluxes can be made
  - More fluxes to determine than mass balance constraints on the system
- Identify a specific point within the feasible set under a given condition
  - **Assumption:** The cell has found the optimal solution by adjusting the system specific constraints (enzyme kinetics and gene regulation) through evolution and natural selection
  - Use of linear programming (LP): Determine the optimal utilization of the metabolic network, subject to the physicochemical constraints, to maximize, for example, the growth of the cell

How to Analyze a Network Using FBA

1. Label Fluxes

- External fluxes: b_1, b_2, b_3, …
- Internal fluxes: v_1, v_2, v_3, …

Note: 1 flux per enzymatic reaction. E.g.: A $\xrightarrow{v_1}$ B



2. Derive Equations

- One equation per biological component
- Make sure you include **all** arrows/fluxes
- Keep your signs consistent
  - E.g., arrows towards a component positive, arrows away negative

Here we have three components, so we have three equations.

A: $b_1 - v_1 - v_2 = 0$

B: $v_1 - b_2 = 0$

C: $v_2 - b_3 = 0$

Notice that we have five unknowns (v_1, v_2, b_1, b_2, b_3) and three equations. This is an example of an **underdetermined system**, which means there is no unique solution to it. Instead we get a range of allowable solutions, where any set of values that falls within that range is a solution to the set of equations. This range is called the **null space**. We can find the null space of a set of equations using linear algebra, but to do this we need to convert our equations to matrix form.

How to Analyze a Network Using FBA

3. Convert the system of equations to a matrix equation: S•v=0

- **S** is a matrix containing the numerical coefficients in the equations
- **v** is a vector whose components are the fluxes
- Each row of **S** corresponds to one equation, so our matrix has as many rows as the number of metabolites in our system
- Each column of **S** corresponds to a flux, so our matrix has as many columns as the number of fluxes in our system
- The equation **S•v=0** means that the **dot product** of matrix **S** and vector **v** equals 0
- The matrix equation for our system is:

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

- Note that you can perform the product to confirm that your matrix is correct.
  - A: $(-1)(v_1) + (-1)(v_2) + (1)(b_1) + (0)(b_2) + (0)(b_3) = -v_1 - v_2 + b_1$
  - .....

How to Analyze a Network Using FBA

4. Solve for the "null space" of S (if interested)

- There are linear algebra techniques that will allow you to do this by hand, but it's much easier to just do it with Mathematica which has a built-in function for it:

  MatrixForm[NullSpace[m]]

- Here '**m**' is the matrix whose null space we want, "NullSpace" is the Mathematica command to do it, and nesting it within "MatrixForm" turns the result into matrix format.
- For our example, the solution is: $\begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$
- Each row is a linearly- and genetically-independent pathway in our system. Because they are linearly independent, they form a **basis**, meaning any pathway in our system can be constructed from linear combinations (additions and subtractions) of these vectors.
- All linear combinations of the basis vectors are solutions to **S•v=0**

How to Analyze a Network Using FBA

5. Predict system-level behavior based on optimality assumptions

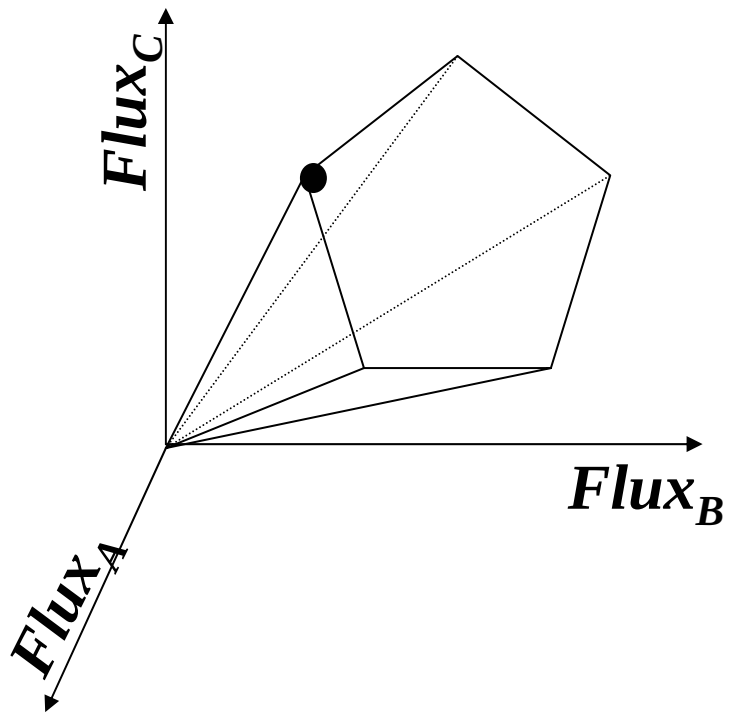
- We can predict the behavior of the system under a variety of different conditions/constraints. We do this by using **linear programming (LP)**, which is a type of algorithm that optimizes with respect to a set of **decision variables** for a given **objective function** and under a set of **constraints**.
- The **objective function** is the function or component that you would like to optimize. For example, if we want to minimize the export of C, the cost function is b_3.

$$\begin{aligned} &\underset{v}{Max} && Z = f(v) \\ &subject\ to && \mathbf{S} \cdot \mathbf{v} = \mathbf{b} \\ & && \alpha_i \leq v_i \leq \beta_i \end{aligned}$$

• The feasible space for the flux distribution (the null space of S) is a **convex polyhedral cone** and represents the **metabolic capabilities** of the network.

• The actual flux distribution is determined by the cell's regulatory mechanisms.

• In absence of kinetic information, we can estimate the metabolic flux distribution by postulating **objective functions(Z)** that underlie the cell's behavior.



A schematic representation of the feasible space in a hypothetical 3-flux system

- The **constraints** are:
  - Steady state: **S•v=0**
  - Biochemically feasible: non-negativeness
  - The range of values each flux can take
- We can use **Excel** to solve linear programs
- Here we are telling Excel the following:
  - Minimize the export flux b_3
  - The fluxes must satisfy matrix **S•v=0**
  - Internal fluxes v_1 and v_2 can take any positive value
  - Export fluxes b_1, b_2, and b_3 can be between 0 and 100, 5 and 10, and 0 and 10, respectively
- This gives the following result: **{5,0,5,5,0}**
- This means that b_3 reaches its lowest value given the above constraints when the fluxes are: $v_1 = 5$, $v_2 = 0$, $b_1 = 5$, $b_2 = 5$, $b_3 = 0$.

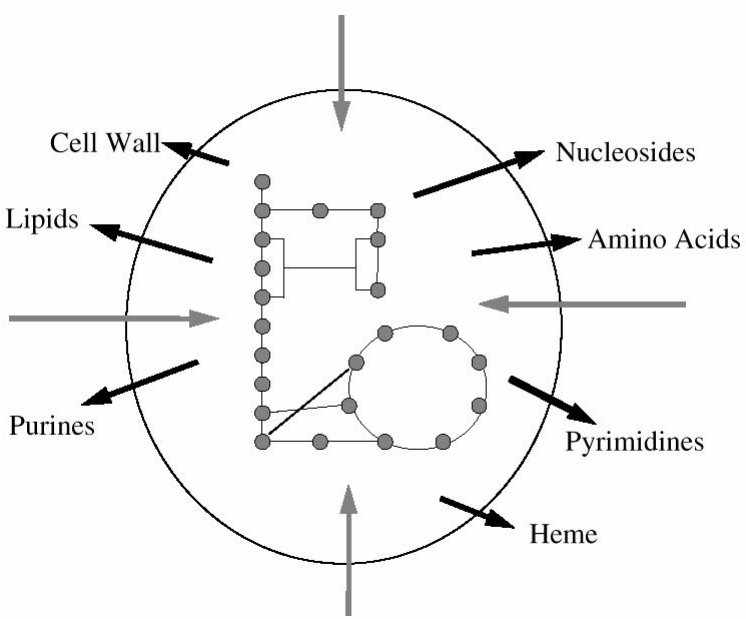
Excel

Flux Balance Example		
Objective		
Z	0	
Decision variables		
v1	5	
v2	0	
b1	5	
b2	4.999999	
b3	0	
Constraints		
Export limit 1: b1 <= 100	5	100
Export limit 2: b2 >= 5	4.999999	5
Export limit 3: b2 <= 10	4.999999	10
Export limit 4: b3 <= 10	0	10
Non-negativity 1: v1 >= 0	5	0
Non-negativity 2: v2 >= 0	0	0
Non-negativity 3: b1 >= 0	5	0
Non-negativity 4: b2 >= 0	4.999999	0
Non-negativity 5: b3 >= 0	0	0
Steady state A: b1-v1-v2 = 0	0	0
Steady state B: v1-b2 = 0	1E-06	0
Steady state C: v2-b3 = 0	0	0

- One widely used biologically meaningful objective function, at least for prokaryotic metabolic networks such as *E. coli*, is the biomass growth function.

$$\sum_{allM} d_M \cdot M \xrightarrow{v_{growth}} biomass$$

where d_M represents the stoichiometric ratio in which metabolite M is required for growth.



These metabolic demands for growth are based on composition analysis of cell mass which for *E. coli* are:

Cofactor	d_M (mmol)	Precursor	d_M (mmol)	Precursor	d_M (mmol)
ATP	41.2570	G6P	0.2050	3PG	1.4960
NADH	-3.5470	F6P	0.0709	PEP	0.5191
NADPH	18.2250	R5P	0.8977	Pyr	2.8328
		E4P	0.3610	AcCoA	3.7478
		T3P	0.1290	OA	1.7867
				αKG	1.0789

- **Comparison between model predicted fluxes and experimentally measured ones.**

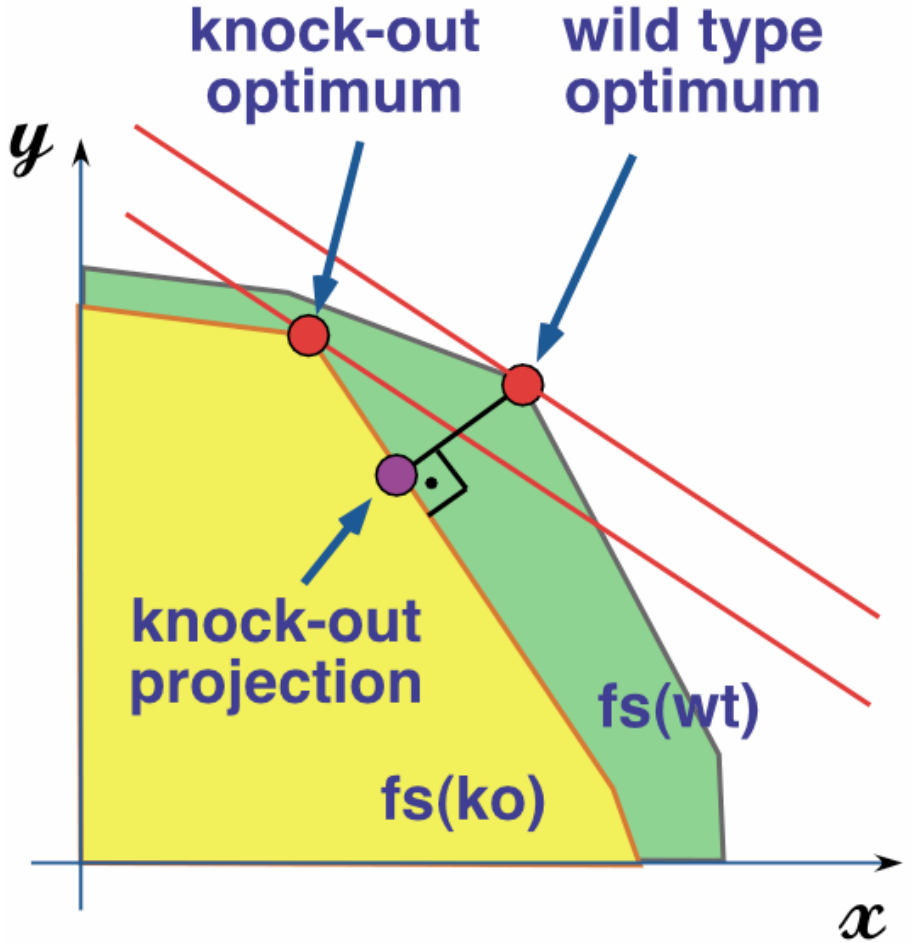
  Example E. coli:

Experimental/*in silico*

Gene	Glucose	Glycerol	Succinate	Acetate	Gene	Glucose	Glycerol	Succinate	Acetate
aceEF	-/+				*pgl*	+/+			
aceA				-/-	*pntAB*	+/+	+/+	+/+	+/+
aceB				-/-	*glk*	+/+			
ackA				+/+	*ppc*	+/+	-/+	+/+	+/+
acs				+/+	*pta*				+/+
acn	-/-	-/-	-/-	-/-	*pts*	+/+			
cyd	+/+				*pyk*	+/+			
cyo	+/+				*rpi*	-/-	-/-	-/-	-/-
eno	-/+	-/+	-/-	-/-	*sdhABCD*	+/+			
fba	-/+				*tpi*	-/+	-/-	-/-	-/-
fbp	+/+	-/-	-/-	-/-	*unc*	+/+		+/+	-/-
gap	-/-	-/-	-/-	-/-	*zwf*	+/+			
gltA	-/-	-/-	-/-	-/-	*sucAD*	+/+			
gnd	+/+				*zwf, pnt*	+/+			
idh	-/-	-/-	-/-	-/-	*pck, mez*			-/-	-/-
ndh	+/+	+/+			*pck, pps*			-/-	-/-
nuo	+/+	+/+			*pgi, zwf*	-/-			
pfk	-/+				*pgi, gnd*	-/-			
pgi	+/+	+/+			*pta,acs*				-/-
pgk	-/-	-/-	-/-	-/-	*tktA, tktB*	-/-			

Minimization of Metabolic Adjustment (MOMA)

- **Perturbed metabolic networks** (mutants, growth under conditions that seldom exist naturally)
- An **alternative optimality assumption**: the cell adjusts its flux distribution through regulations such that its metabolic state is as colse to its wild type state as possible
- The cell has not been optimized towards growth through evolution and instead, its flux distribution is close to the wild type counterpart.



x,y are two of the 100s of fluxes

MOMA- Quadratic Programming (QP)

$$\begin{aligned} \underset{v}{Min} \quad & Z = \sum_{v} (v - v_{wt})^2 \\ subject\ to \quad & \mathbf{S} \cdot \mathbf{v} = \mathbf{b} \\ & \alpha_i \le v_i \le \beta_i \\ & v_m = 0 \end{aligned}$$

where v_{wt} is the optimal flux distribution of the wild type cell; v_m are the fluxes that get eliminated due to gene knockouts.

- Distance function
  - Normalization
  - 1-norm, **2-norm (Euclidean)**, and infinity-norm distances
  - inclusion of growth flux
- Quadratic programming
  - quadratic objective function; linear constraints
  - many solvers available (e.g. Excel)

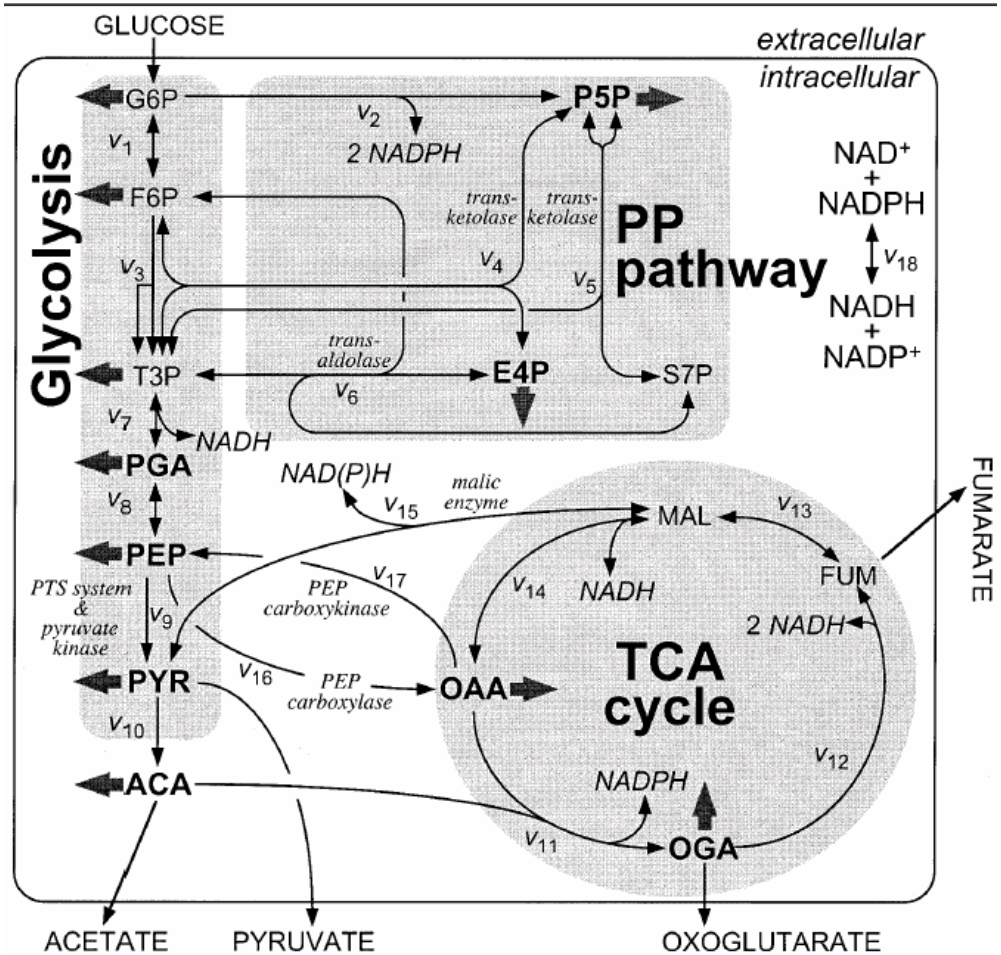
JOURNAL OF BACTERIOLOGY, Jan. 2002, p. 152–164 Vol. 184, No. 1
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.1.152–164.2002

Metabolic Flux Responses to Pyruvate Kinase Knockout in *Escherichia coli*

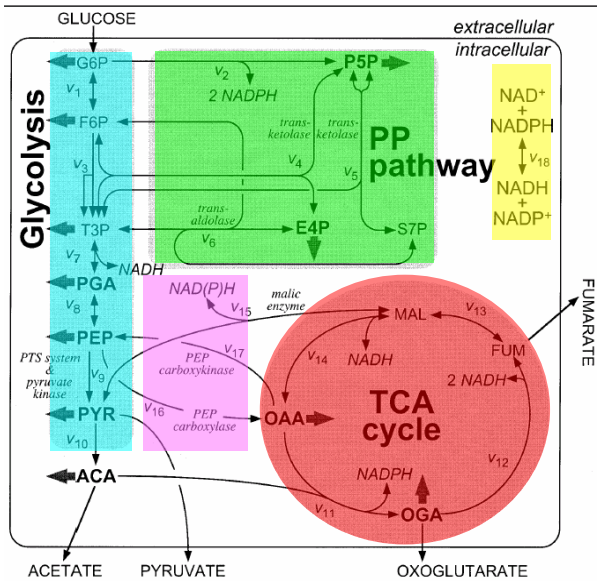
Marcel Emmerling,[1]† Michael Dauner,[1] Aaron Ponti,[1] Jocelyne Fiaux,[2] Michel Hochuli,[2] Thomas Szyperski,[3] Kurt Wüthrich,[2] J. E. Bailey,[1] and Uwe Sauer[1]*

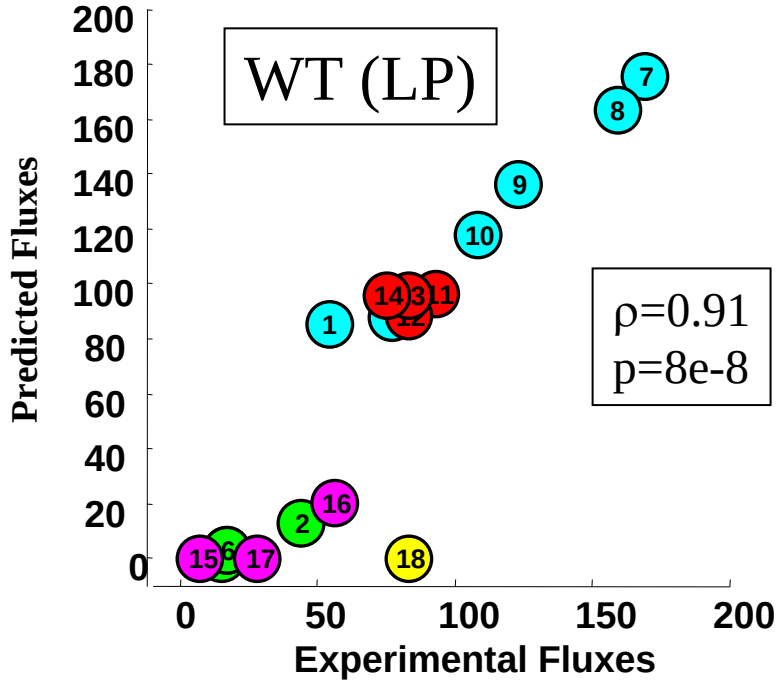
Institute of Biotechnology[1] and Institute of Molecular Biology and Biophysics,[2] ETH Zürich, CH-8093 Zürich, Switzerland, and Department of Chemistry, University at Buffalo, State University of New York, Buffalo, New York 14260[3]

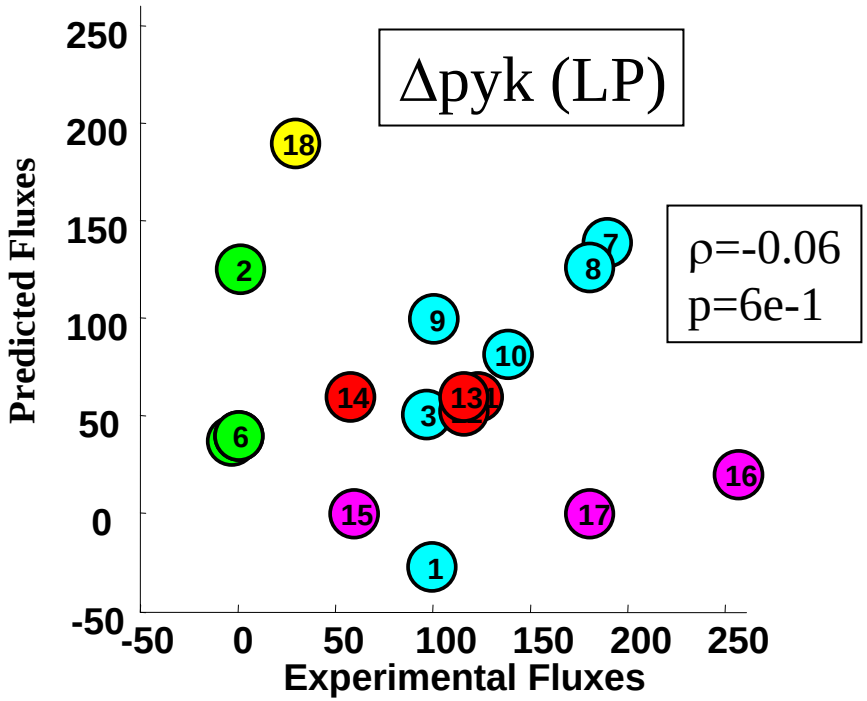
Flux Data

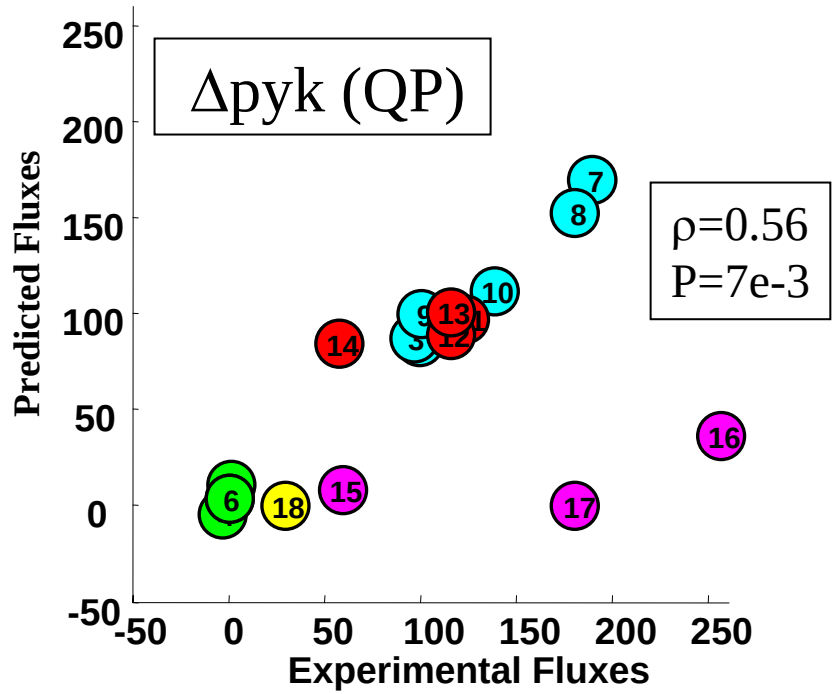


C009-limited









Flux data (MOMA & FBA)

Condition	Method	ρ_1	p-val (a)	p-val (b)	ρ_2	p-val (c)	p-val (d)
C-0.09	wt	0.91	8.2E-08				
	ko (FBA)	-0.064	**6.0E-01**	3.3E-03	-0.36	**9.0E-01**	2.4E-04
	ko MoMA	0.56	7.4E-03		0.48	2.3E-02	
C-0.4	wt	0.97	8.1E-12				
	ko (FBA)	0.77	8.1E-05	2.5E-03	0.36	**7.0E-02**	1.4E-02
	ko MoMA	0.94	2.6E-09		0.74	2.3E-04	
N-0.09	wt	0.78	7.1E-05				
	ko (FBA)	0.86	3.0E-06	**9.0E-02**	0.096	**3.5E-01**	4.6E-02
	ko MoMA	0.73	2.8E-04		0.49	2.0E-02	

Segre, et al, 2002 Analysis of optimality in natural and perturbed metabolic networks. PNAS 99: 15112-7.

References for FBA

Review:

- B. Palsson, "The Challenges of *in silico* Biology". **Nature** 18: 1147-50 (2000).

Application to E. Coli:

- J. Edwards and B. Palsson, "The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities". **PNAS** 97: 5528-33 (2000).
- J. Edwards, R. Ibarra, B. Palsson, "*In silico* predictions of E. coli metabolic capabilities are consistent with experimental data". **Nature Biotechnology** 19: 125-30 (2001).

Mathematical Methods:

- C. Schilling, B. Palsson, "The underlying pathway structure of biochemical reaction networks". **PNAS 95**: 4193-4198 (1998).
- C. Schilling et al, "Towards metabolic phenomics: Analysis of genomic data using flux balances". **Biotechnol Prog** 15: 288-295 (1999).
- C. Schilling et al, "Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era". **Biotechnol Prog** 15: 296-303 (1999).

MOMA

- Segre, et al, 2002 Analysis of optimality in natural and perturbed metabolic networks. PNAS 99: 15112-7. (Minimization Of Metabolic Adjustment) http://arep.med.harvard.edu/moma/

Acknowledgements

- This presentation contains materials written by Suzanne Komili
- Part of this presentation was taken from Prof. George Church's lection notes for Nov. 11th, 2003.
- Reference for enzymatic kinetics:

  Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Michael A. Savageau, Addion-Wesley Publishing Company, 1976.