Integration of high-throughput datasets through dynamical modeling of regulatory networks

Hidde de Jong
IBIS
INRIA Grenoble – Rhône-Alpes
Hidde.de-Jong@inria.fr

Grenoble, November 30, 2012

INRIA Grenoble - Rhône-Alpes and IBIS

- IBIS: systems biology group at INRIA/Université Joseph Fourier/CNRS
  - Analysis of bacterial regulatory networks by means of models and experiments
  - Biologists, computer scientists, mathematicians, physicists, …

http://ibis.inrialpes.fr

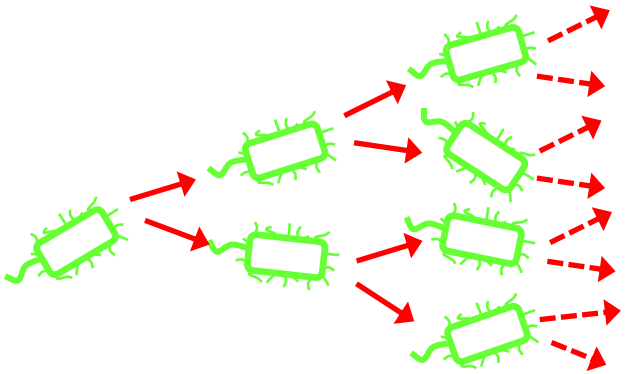
Tutorial overview

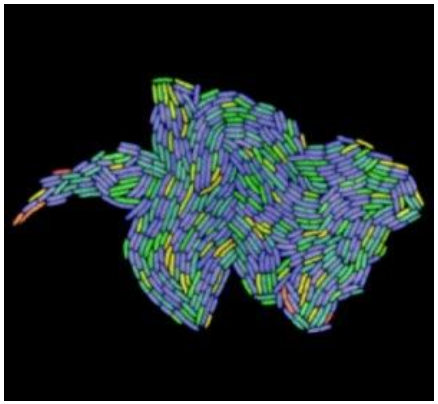
- Bacterial growth and metabolism: high-throughput data sources
- Modeling of regulatory networks in bacteria
  - Flux balance models
  - Kinetic models of metabolic networks
  - Towards models of integrated networks
- Conclusions

Bacterial growth and metabolism

- **Bacteria** are unicellular organisms geared towards growth and division

  Escherichia coli cells have doubling times up to 20 min





Stewart *et al.* (2005), *PLoS Biol.*, 3(2): e45

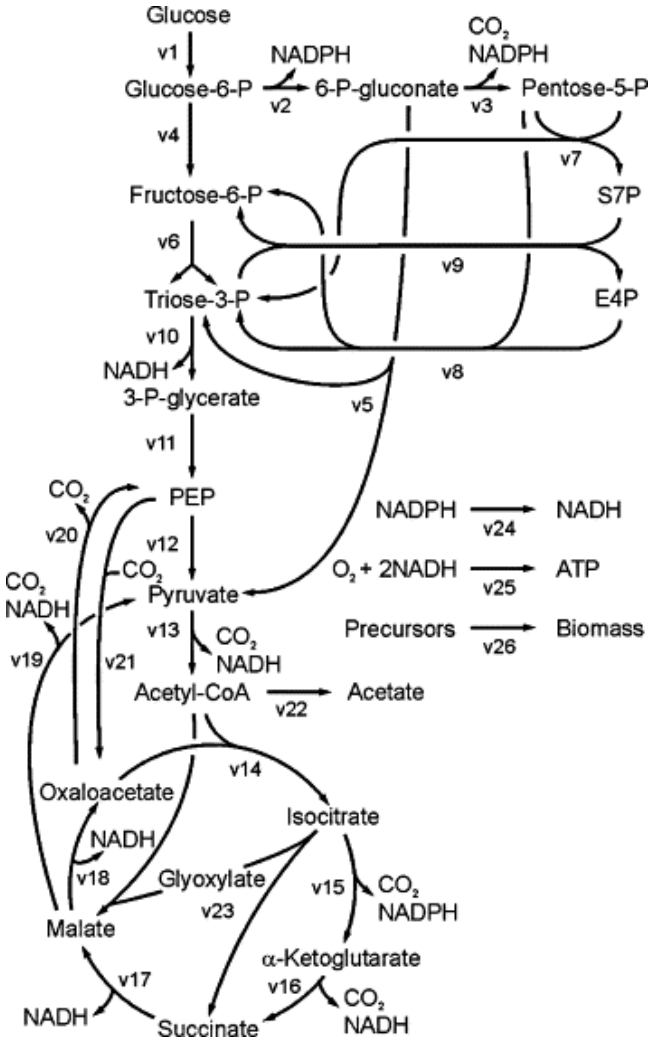
- **Metabolism** fuels growth by production of energy and building blocks for macromolecules, using nutriments from environment

  ATP, amino acids, nucleotides, …

Bacterial growth and metabolism

- Central **carbon metabolism** breaks down carbon sources for energy production and macromolecular synthesis

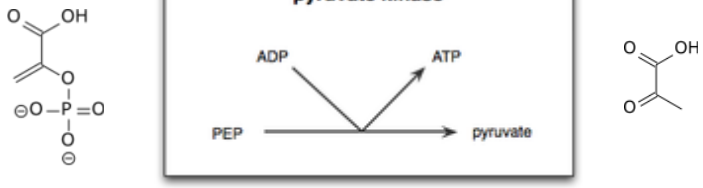
  Glucose, acetate, lactose, …

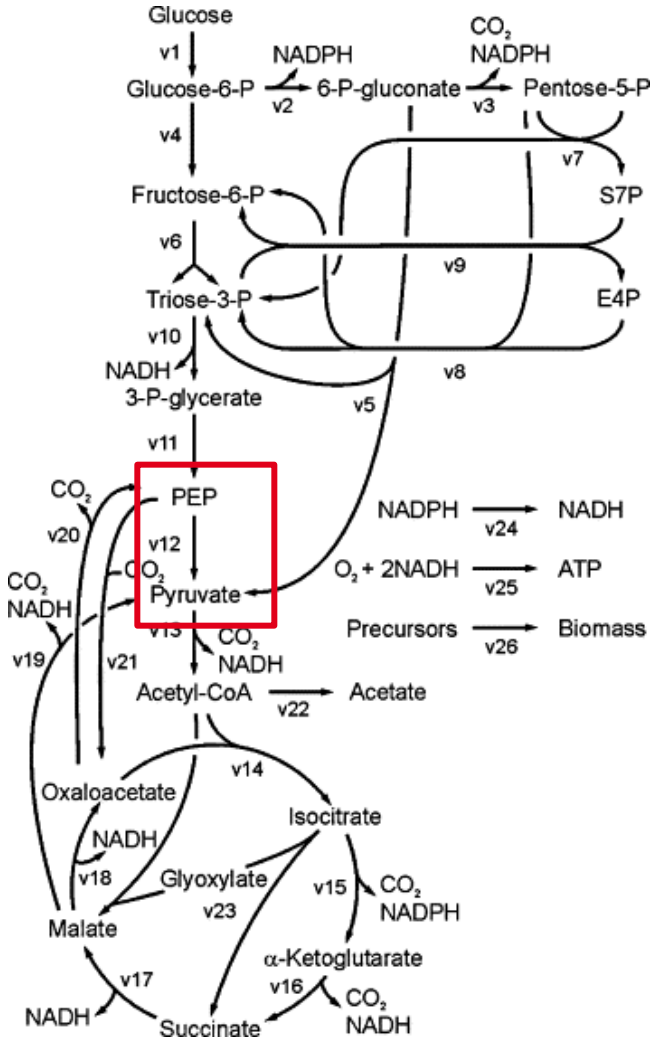


Fischer *et al.* (2004), *Anal. Biochem.*, 325(2):308–16

Bacterial growth and metabolism

- Central **carbon metabolism** breaks down carbon sources for energy production and macromolecular synthesis
  - Glucose, acetate, lactose, …
- **Enzymes** catalyze individual steps in metabolic network
  - Pyruvate kinase transforms phosphoenolpyruvate (PEP) into pyruvate



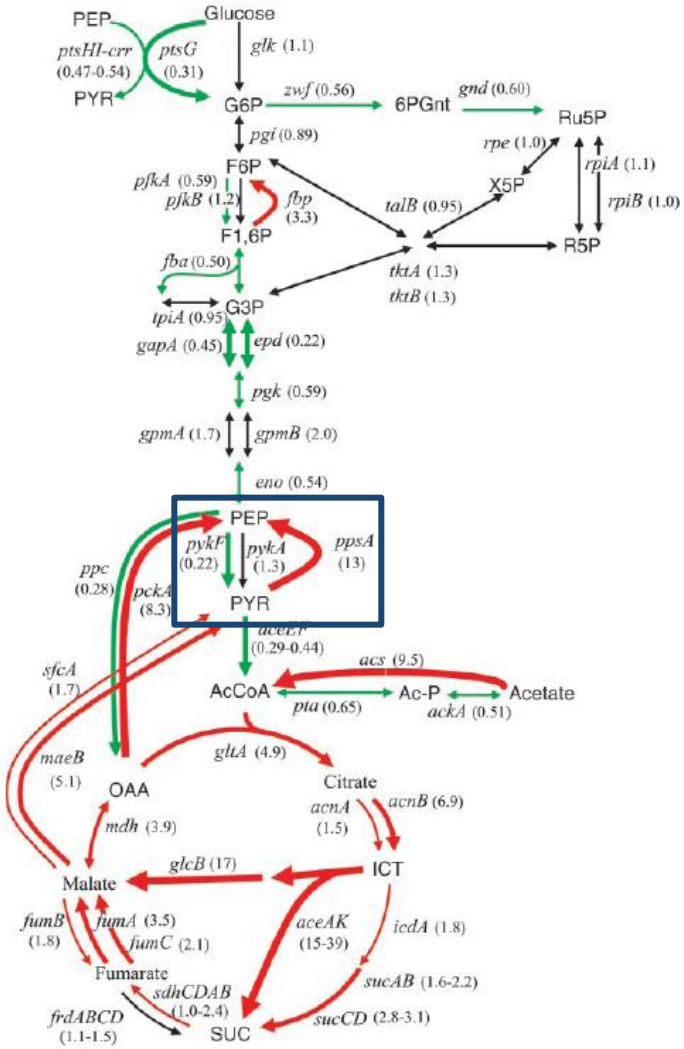


Bacterial growth and metabolism

- Central **carbon metabolism** breaks down carbon sources for energy production and macromolecular synthesis

  Glucose, acetate, lactose, …

- Enzymes produced from information encoded in **genes**
  - *pykF* is gene encoding pyruvate kinase



Oh *et al.* (2002), *J. Biol. Chem.*, 277(15):13175–83

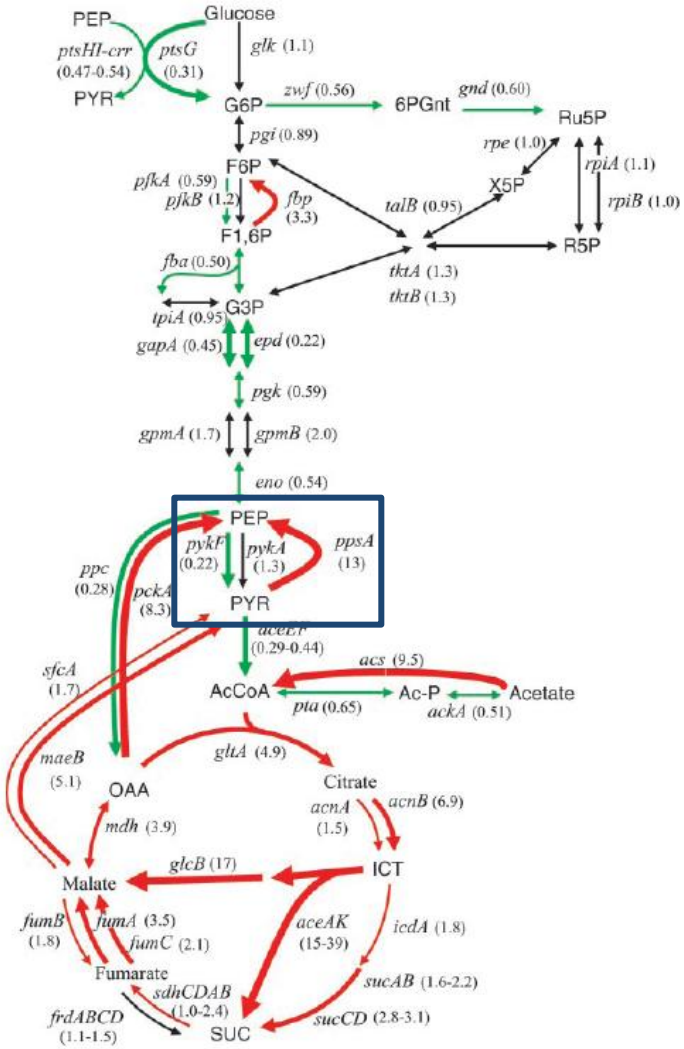
Bacterial growth and metabolism

- Central **carbon metabolism** breaks down carbon sources for energy production and macromolecular synthesis

  Glucose, acetate, lactose, …

- Enzymes produced from information encoded in **genes**
  - *pykF* is gene encoding pyruvate kinase
  - Expression of *pykF* regulated by transcription factor Cra

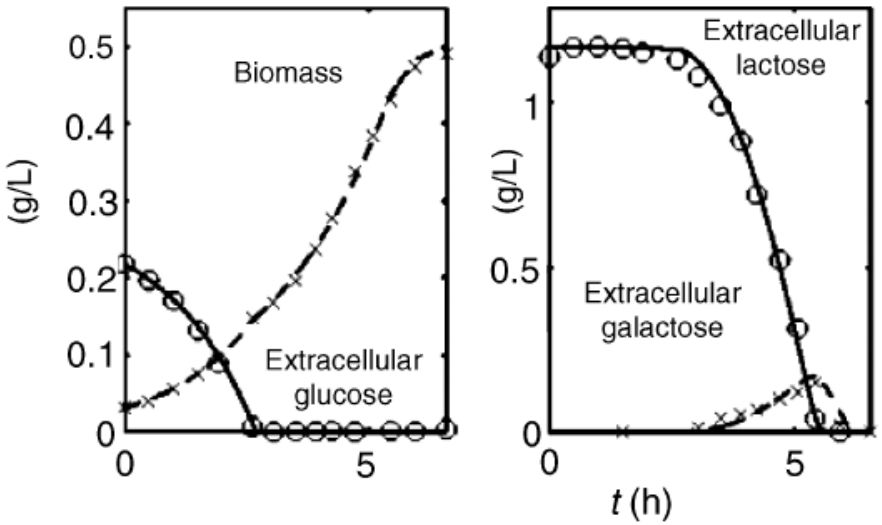
Oh *et al.* (2002), *J. Biol. Chem.*, 277(15):13175–83



Bacterial growth and metabolism

- Bacterial metabolism is **flexible**, allowing cell to grow on different carbon sources

  Preferential utilization: **diauxic growth** on glucose and lactose



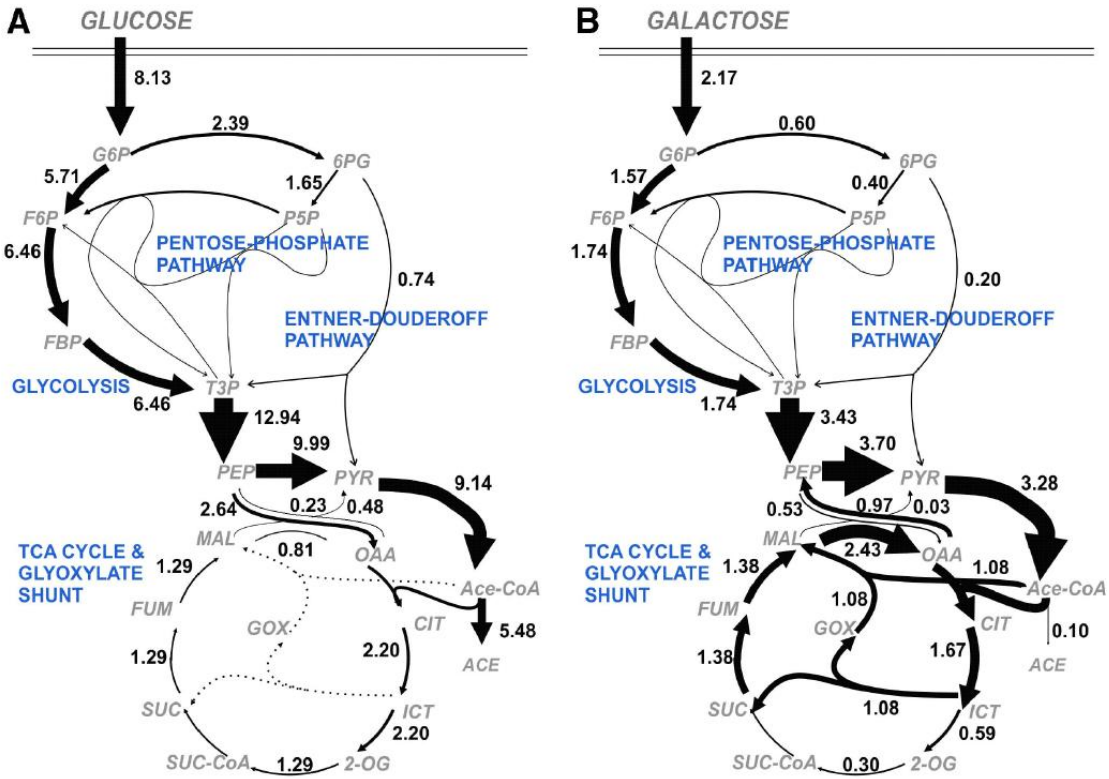
Bettenbrock *et al.* (2006), *J. Biol. Chem.*, 281(5):2578-84

- Adaptation of bacterial physiology to different carbon sources

Growth transition and metabolism

- Adaptation of bacteria to different carbon source involves changes in **metabolic fluxes**

  Different flux distribution in central metabolism of *E. coli* during growth on glucose and galactose



Haverkorn van Rijsewijk *et al.* (2011), *Mol. Syst. Biol.*, 7:477

Growth transition and metabolism

- Adaptation of bacteria to different carbon source involves adjustment of **metabolite levels**

  Different metabolite concentrations in *E. coli* cells growing on glucose and acetate

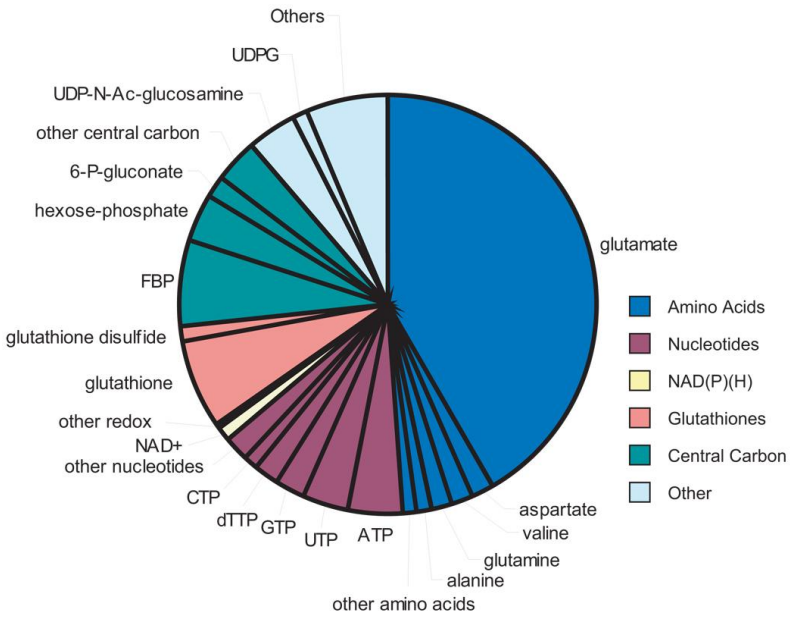


Table 1 Intracellular metabolite concentrations in glucose-fed, exponentially growing *E. coli*

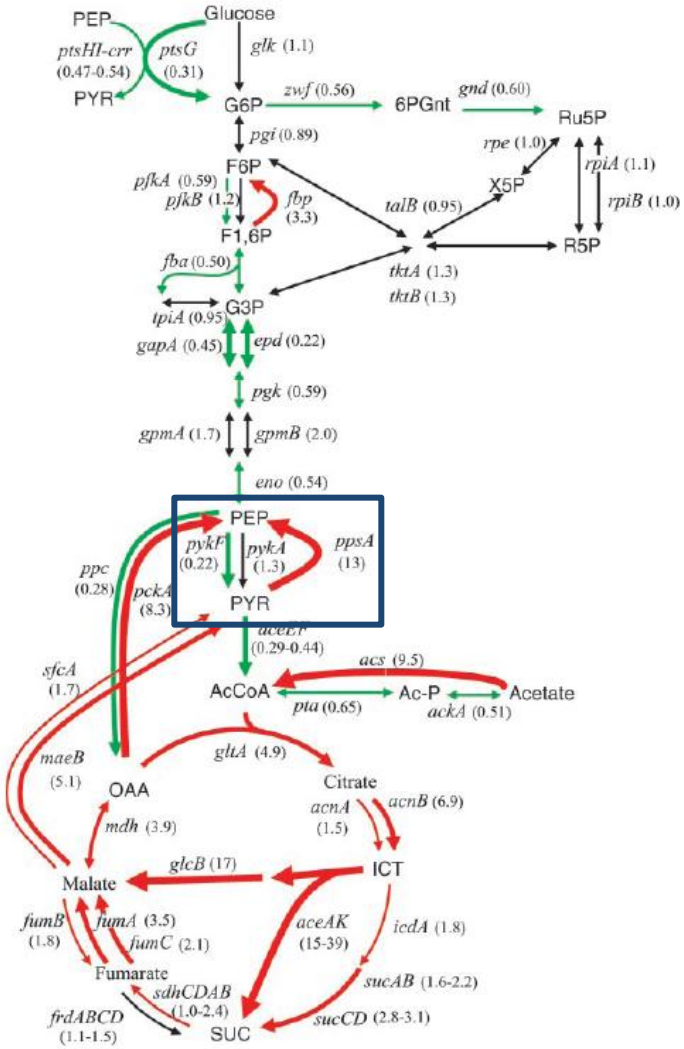
Metabolite	mol l^{-1}	Metabolite	mol l^{-1}
Glutamate	9.6×10^{-2}	UDP-glucuronate (51)	5.7×10^{-4}
Glutathione	1.7×10^{-2}	ADP	5.6×10^{-4}
Fructose-1,6-bisphosphate	1.5×10^{-2}	Asparagine (52)	5.1×10^{-4}
ATP	9.6×10^{-3}	α-Ketoglutarate	4.4×10^{-4}
UDP-*N*-acetylglucosamine (29)	9.2×10^{-3}	Lysine (53)	4.1×10^{-4}
Hexose-P[a]	8.8×10^{-3}	Proline (54)	3.9×10^{-4}
UTP (30)	8.3×10^{-3}	dTDP (55)	3.8×10^{-4}
GTP (31)	4.9×10^{-3}	Dihydroxyacetone phosphate	3.7×10^{-4}
dTTP	4.6×10^{-3}	Homocysteine (56)	3.7×10^{-4}
Aspartate	4.2×10^{-3}	CMP (57)	3.6×10^{-4}
Valine (32)	4.0×10^{-3}	Deoxyribose-5-P (58)	3.0×10^{-4}
Glutamine	3.8×10^{-3}	Isoleucine (59) + leucine (60)	3.0×10^{-4}
6-Phosphogluconate	3.8×10^{-3}	AMP	2.8×10^{-4}

Bennett *et al.* (2009), *Nat. Chem. Biol.*, 5(8):593-9

Growth transition and gene expression

- Adaptation of bacteria to different carbon source involves adjustment of **expression levels of enzymatic genes**

  Difference in expression levels of genes encoding enzymes in central metabolism of *E. coli* during growth on glucose and acetate

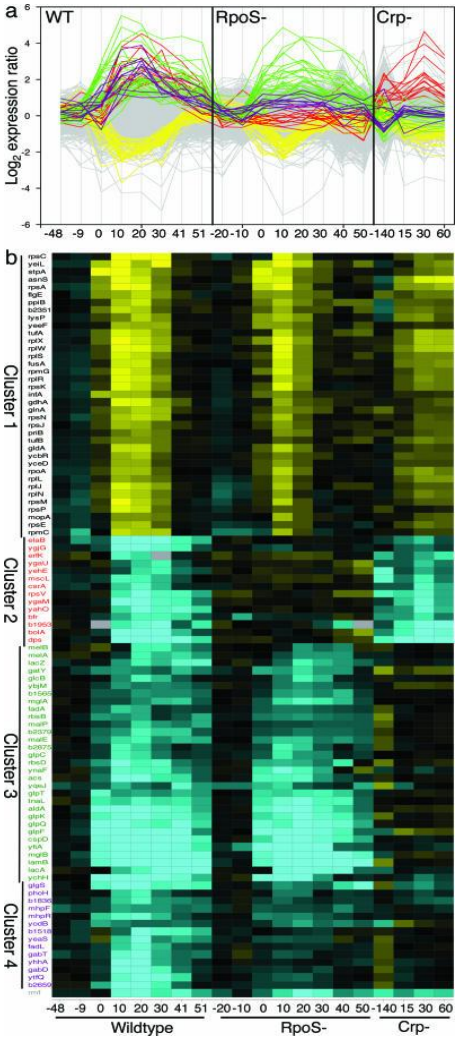


Oh *et al.* (2002), *J. Biol. Chem.*, 277(15):13175–83

Growth transition and gene expression

- Adaptation of bacteria to different carbon source involves genome-wide **reorganization of gene expression**

  Gene expression during glucose-acetate shift in *E. coli*

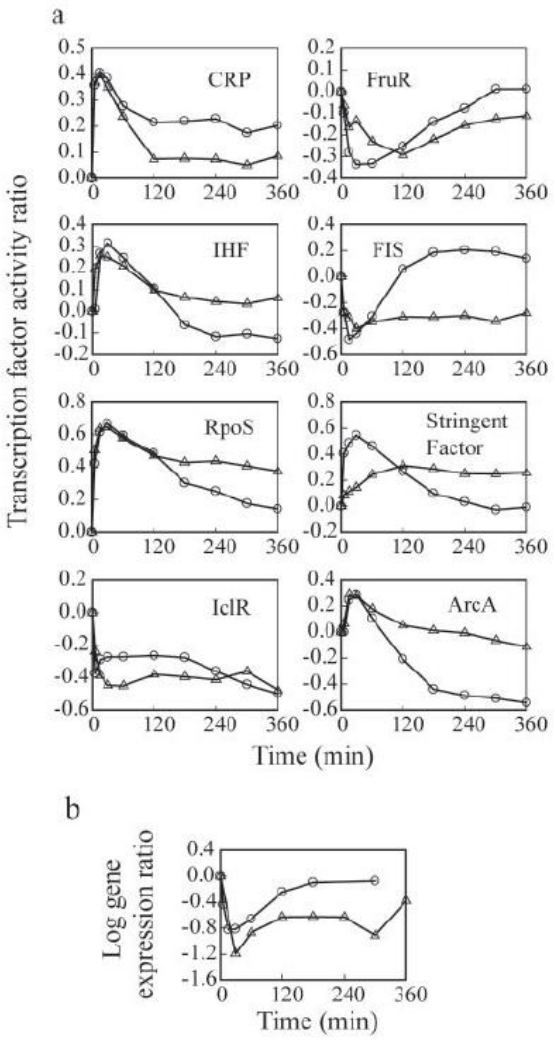


Traxler *et al.* (2006), *Proc. Natl. Acad. Sci. USA*, 103(7):2374–9

Growth transition and gene expression

- Adaptation of bacteria to different carbon source involves adjustment of **transcription factor activity**

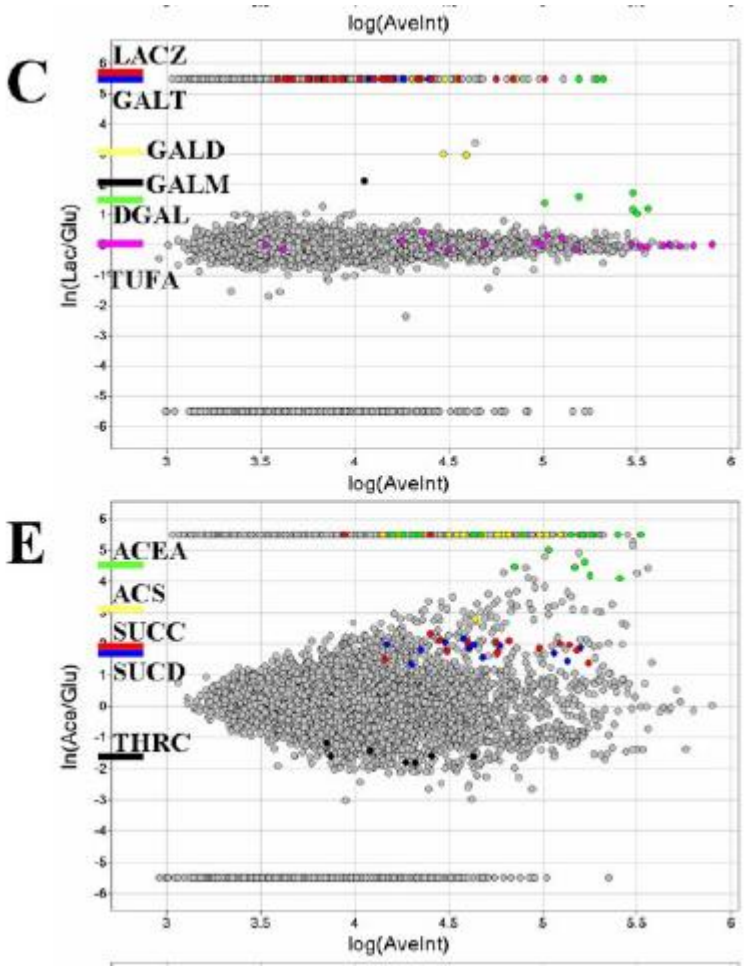
  Changes in activity and concentration of transcription factors during glucose-acetate shift in *E. coli*



Kao *et al.* (2005), *J. Biol. Chem.*, 280(43):36079–87

Growth transition and gene expression

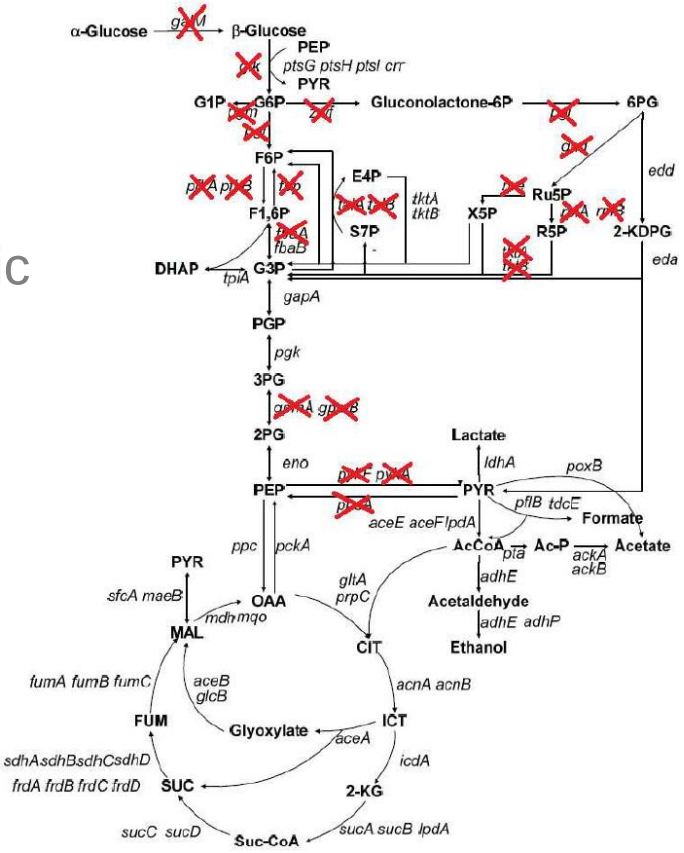
- Adaptation of bacteria to different carbon source involves adjustment of **protein levels**
  - Changes in relative protein abundance during growth on lactose vs glucose and acetate vs glucose



Silva *et al.* (2006), *Mol. Cell. Proteom.*, 5(4):589–607

Multiple high-throughput measurements

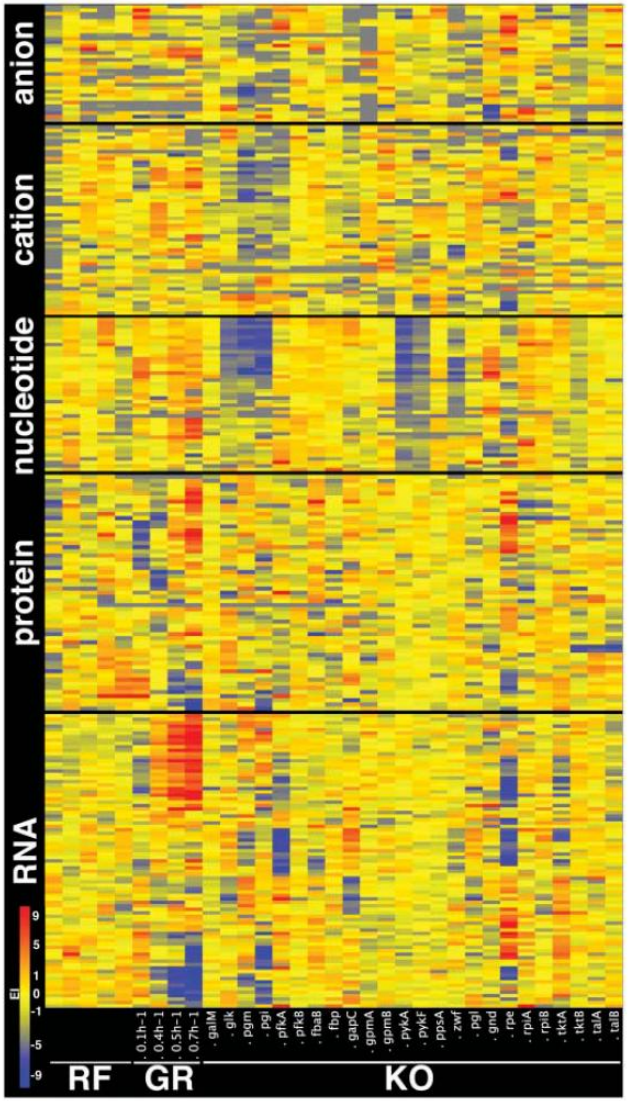
- Adaptation of bacteria to different carbon source involves adjustments on all levels at the same time!
  - Parallel measurement of enzyme and metabolite concentrations, and metabolic fluxes
- **Experimental conditions:**
  - Steady-state measurements in chemostat
  - 24 single-gene disruptants from Keio collection
  - 5 different dilution rates from 0.1 h^{-1} to 0.7 h^{-1}



Ishii *et al.* (2007), *Science*, 316(5284):593-7

Multiple high-throughput measurements

- Adaptation of bacteria to different carbon source involves adjustments on all levels at the same time!
  - Parallel measurement of enzyme and metabolite concentrations, and metabolic fluxes
- **Experimental conditions:**
  - Steady-state measurements in chemostat
  - 24 single-gene disruptants from Keio collection
  - 5 different dilution rates from 0.1 h^{-1} to 0.7 h^{-1}



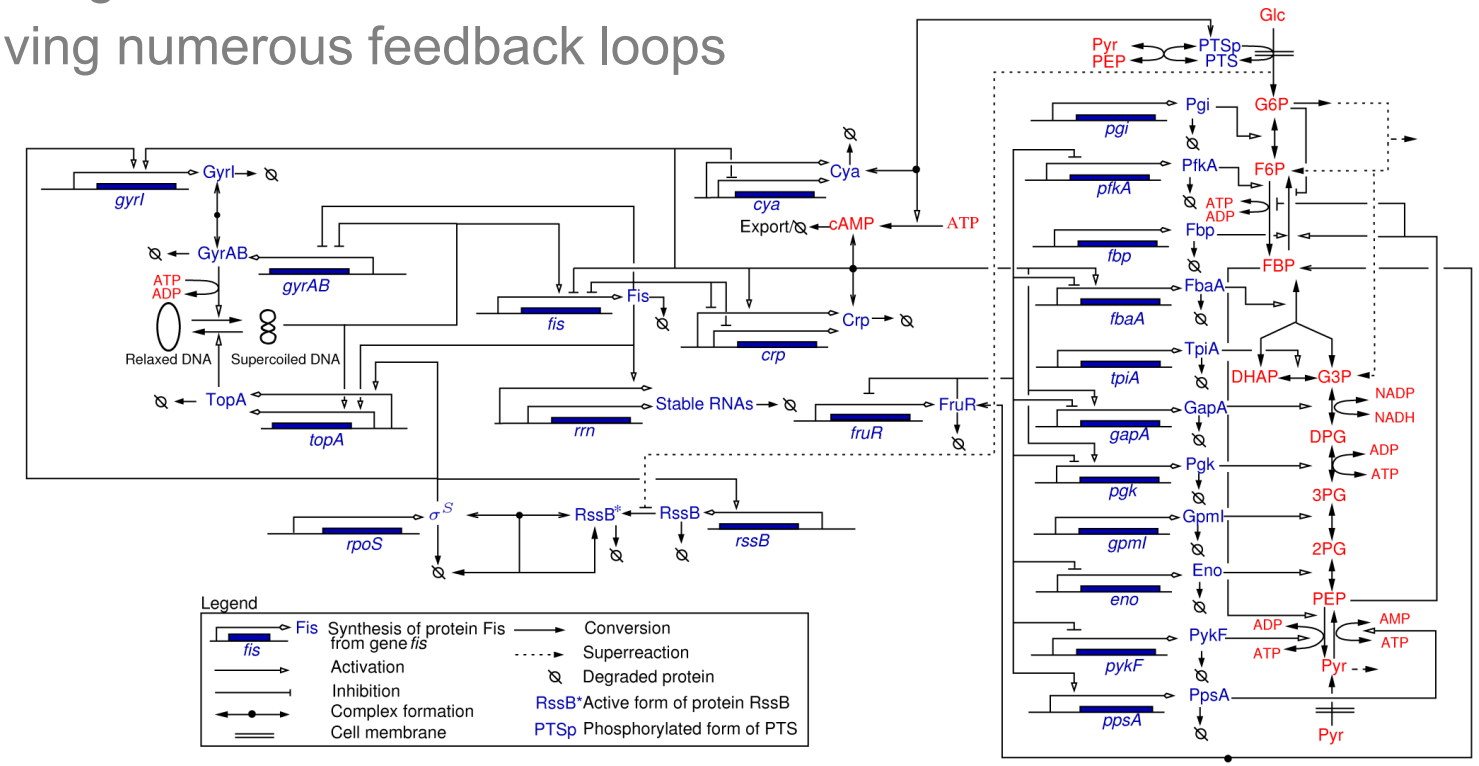
Ishii *et al.* (2007), *Science*, 316(5284):593-7

Coordination of adaptative responses

- Cells are capable of responding to a variety of changes in their environment by adapting their physiology
  - Change in carbon source, starvation, population density, …
- On the molecular level, these responses involve adjustment on different levels
  - Cellular concentrations of metabolites, enzymes, transcription factors, …
- **Coordination of adaptive responses** of bacterial cell achieved by large and complex regulatory networks

Coordination of adaptative responses

- Coordination of adaptative responses of bacterial cell achieved by **large and complex regulatory networks**
  - Variety of molecular mechanisms…
  - … operating on different time-scales…
  - … involving numerous feedback loops



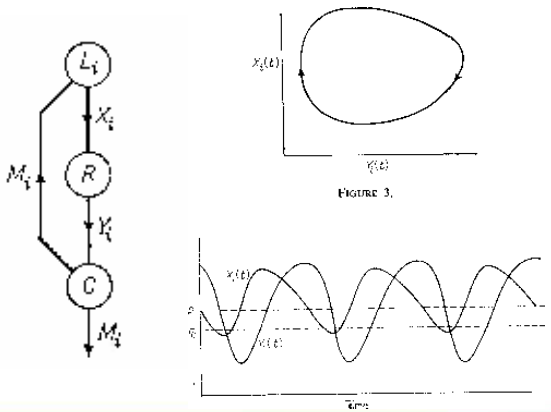
Baldazzi *et al.* (2010), *PLoS Comput. Biol.*, 6(6):e1000812

Coordination of adaptative responses

- Coordination of adaptative responses of bacterial cell achieved by **large and complex regulatory networks**
  - Variety of molecular mechanisms…
  - … operating on different time-scales…
  - … involving numerous feedback loops
- Abundant knowledge on interactions between network components
- Accumulation of data on multi-level response of network to external perturbations
- However, **global view on functioning of network** is difficult to achieve and largely absent today

Regulatory networks as dynamical systems

- Bacterial regulatory networks are **complex nonlinear dynamical systems**, evolving on different time-scales
- **Challenge:** can mathematical models and computer tools help us understand how these systems function?
  - Integration of knowledge on interaction structure and different data sources into mathematical models
  - Use of models to analyze and predict dynamical behavior of system
  - Emergence of new discipline: **systems biology**…
  - … which has a long history in biochemistry, thermodynamics, theoretical biology, …



Goodwin (1963), *Temporal Organization in Cells*, Academic Press

Regulatory networks as dynamical systems

- Bacterial regulatory networks are **complex nonlinear dynamical systems**, evolving on different time-scales
- **Challenge:** can mathematical models and computer tools help us understand how these systems function?
- Variety of **different formalisms** have been developed, describing system on different levels of detail

Stochastic master equations

Ordinary differential equations (ODEs)

Boolean networks

Modeling regulatory networks

- Kinetic models of form $\dot{x} = N\,v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \rightarrow \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$

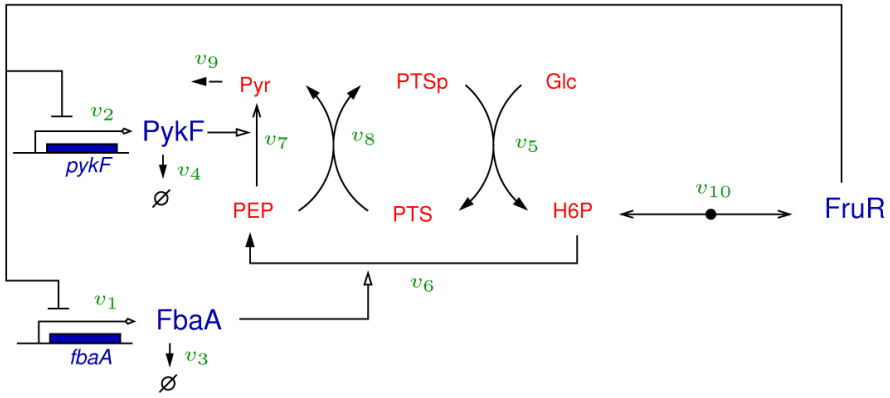
Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

- Stochiometry matrix N describes structure of reaction network
- Reaction rate v depends on concentrations of other cellular components

Modeling regulatory networks: example

- Kinetic models of form $\dot{x} = N\, v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \longrightarrow \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall



$$\begin{aligned}\dot{x}_{PEP} = \; & 2 \cdot v_6(x_{H6P}, x_{PEP}, x_{FbaA}) \\ & -1 \cdot v_7(x_{Pyr}, x_{PEP}, x_{PykF}) \\ & -1 \cdot v_8(x_{PEP}, x_{Pyr}, x_{PTSp})\end{aligned}$$

Simplified model of glycolysis pathway, with metabolic and genetic regulation

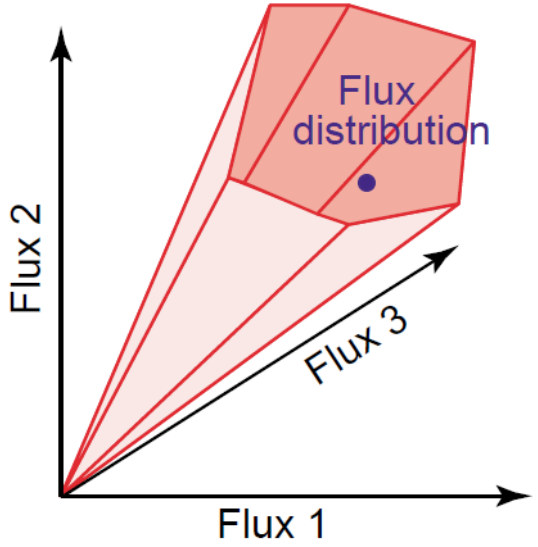
Flux balance analysis (FBA)

- Steady-state dynamics of metabolic network

$$N \cdot v = 0$$

- Stoichiometry matrix defines convex space of possible solutions (**flux cone**)

  System of steady-state equations underdetermined: more reactions than concentrations variables



Stelling (2004), *Curr. Opin. Microbiol.*, 7:513-8

Flux balance analysis (FBA)

- Steady-state dynamics of metabolic network

$$N \cdot v = 0$$

- Stoichiometry matrix defines convex space of possible solutions (**flux cone**)

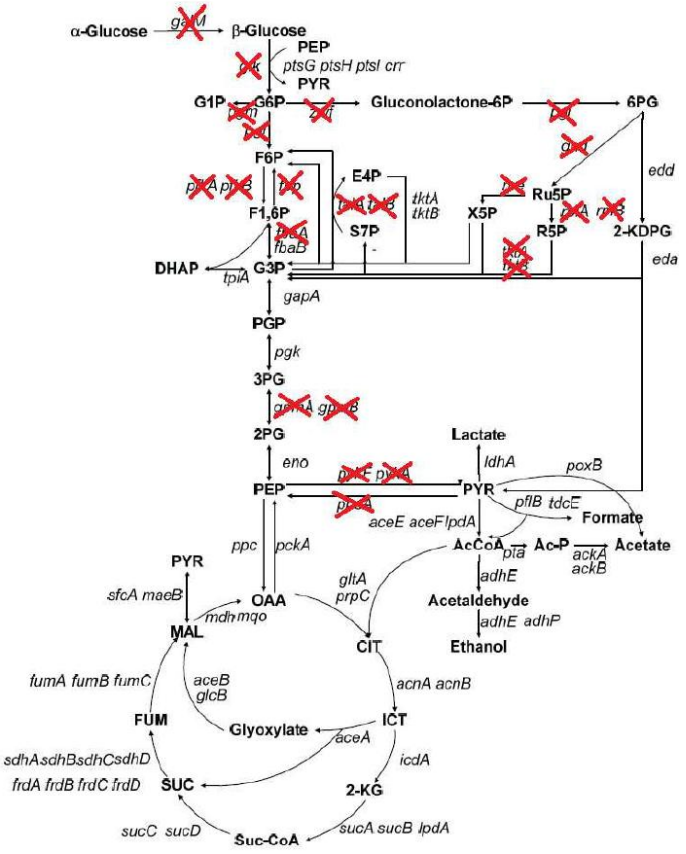
  System of steady-state equations underdetermined: more reactions than concentrations variables

- **Flux balance analysis (FBA)** predicts flux distribution enabling optimal network performance
  - Optimal network performance often interpreted as **maximal growth rate**
  - Refinement by including additional **constraints**: thermodynamics, gene expression, …
  - Constrained optimization problem

Palsson (2006), *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press

Flux balance analysis (FBA)

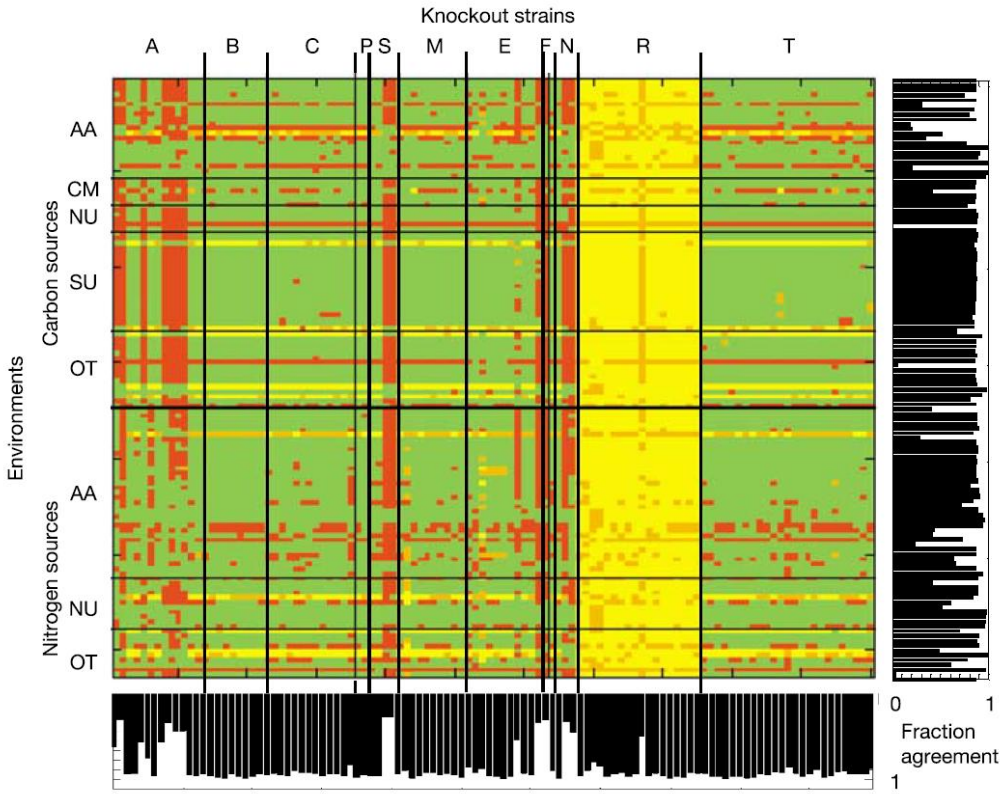
- Integration of **gene expression constraints** by deleting fluxes in strains mutated for an enzyme



Flux balance analysis (FBA)

- Integration of **gene expression constraints** by deleting fluxes in strains mutated for an enzyme
- Genome-scale model of *E. coli* metabolism, including regulation of enzymatic genes
- When including gene expression constraints, prediction of optimal growth rate in different mutants and growth conditions improved

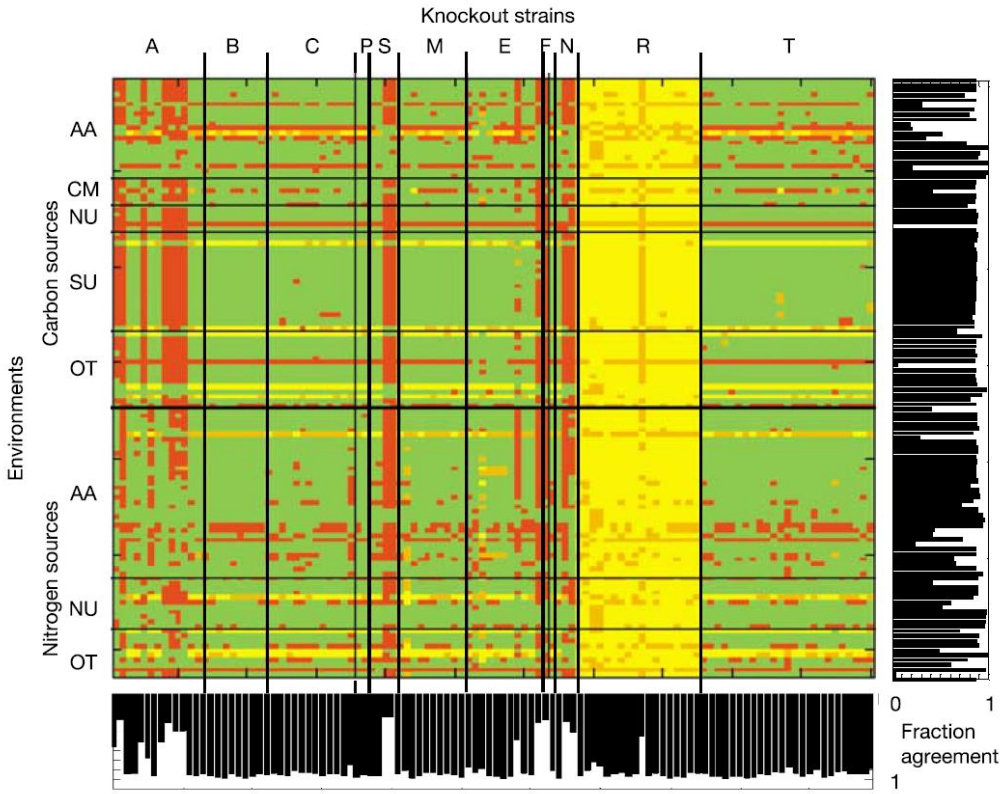
  60% vs 78%



Covert *et al.* (2004), *Nature*, 429(6987):92-6

Flux balance analysis (FBA)

- Integration of **gene expression constraints** by deleting fluxes in strains mutated for an enzyme
- Genome-scale model of *E. coli* metabolism, including regulation of enzymatic genes
- Identification of missing enzymes and regulatory interactions

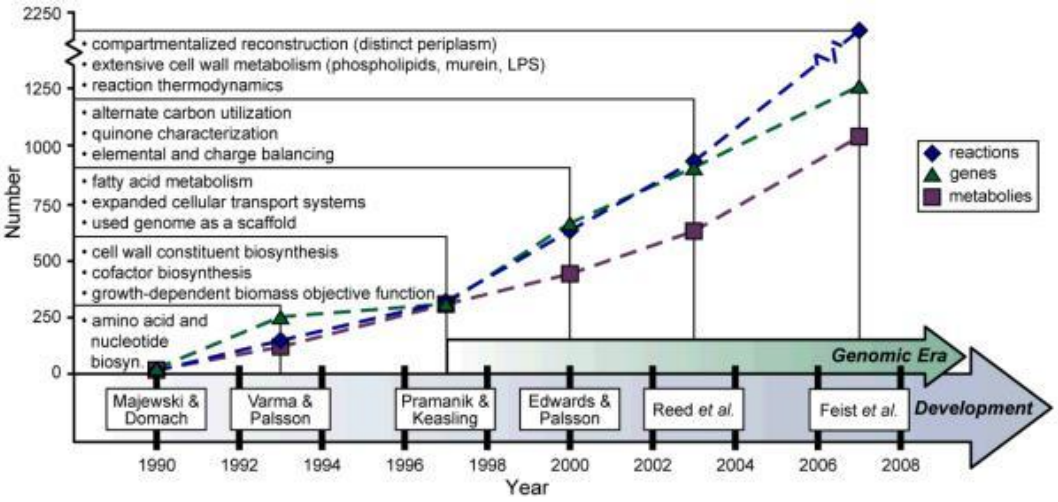


Covert *et al.* (2004), *Nature*, 429(6987):92-6

Flux balance analysis (FBA)

- Integration of **gene expression constraints** by deleting fluxes in strains mutated for an enzyme
- Genome-scale model of *E. coli* metabolism, including regulation of enzymatic genes
- FBA models provide increasingly complete, static picture of metabolism

  In *E. coli* and other (less well-characterized) organisms



Feist and Palsson (2008), *Nat. Biotechnol.*, 26(6):659-67

Modeling regulatory networks

- Kinetic models of form $\dot{x} = N\,v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \rightarrow \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

- Stochiometry matrix N describes structure of reaction network
- Reaction rate v depends on concentrations of other cellular components
  - Mass-action
  - Michaelis-Menten (reversible/irreversible)

$$v = \frac{V^+ \cdot \frac{S}{Ks} - V^- \cdot \frac{P}{Kp}}{1 + \frac{S}{Ks} + \frac{P}{Kp}}$$

Modeling regulatory networks

- Kinetic models of form $\dot{x} = N\,v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \rightarrow \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

- Stochiometry matrix N describes structure of reaction network
- Reaction rate v depends on concentrations of other cellular components
  - Mass-action
  - Michaelis-Menten (reversible/irreversible)
  - Hill
  - Monod-Wyman-Changeux

Modeling regulatory networks

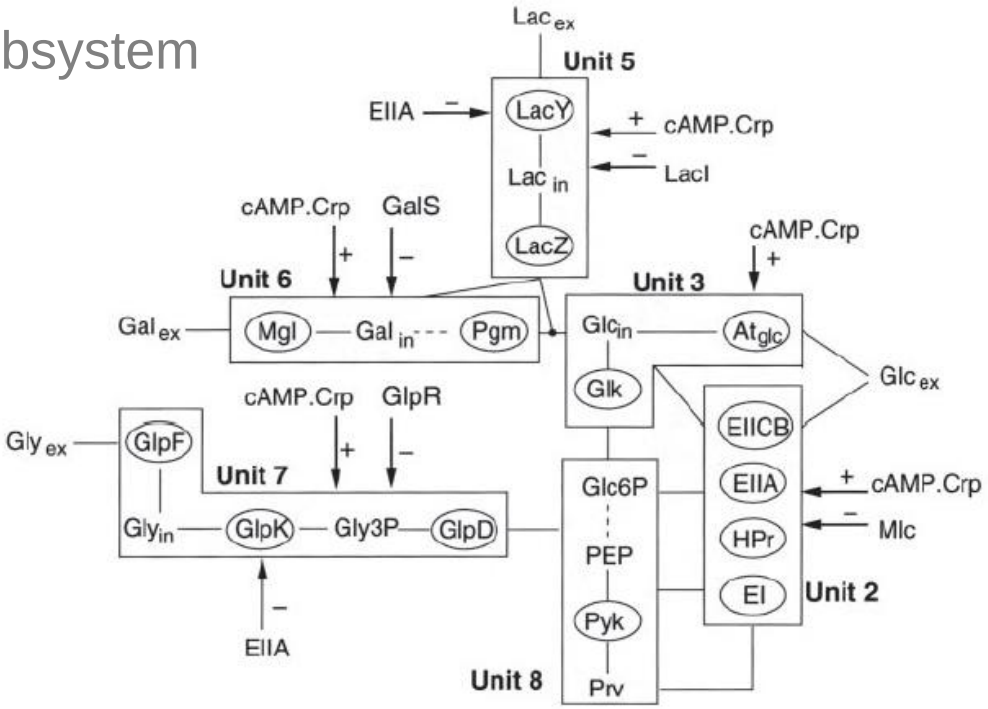
- Kinetic models of form $\dot{x} = N\,v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \rightarrow \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

- Stochiometry matrix N describes structure of reaction network
- Reaction rate v depends on concentrations of other cellular components
- In general, reaction rate functions are **nonlinear** and have **many parameters**, **difficult to measure** directly *in vivo*
- Nevertheless, some examples of well-calibrated models!

Kinetic modeling of metabolism

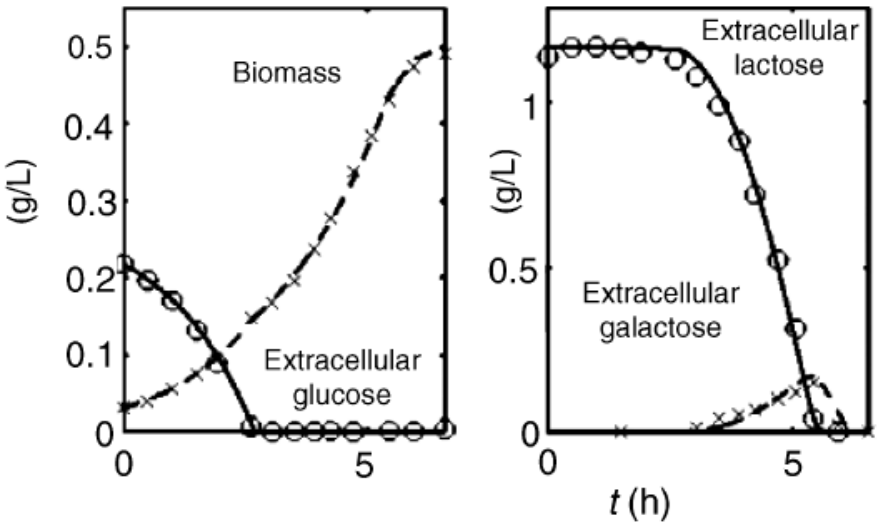
- Model of uptake of carbon sources (glucose, lactose, glycerol, …) by *E. coli*
  - Several dozens of equations and more than a hundred parameters, many of them unknown or unreliable
  - Mosty metabolic subsystem



Bettenbrock *et al.* (2005), *J. Biol. Chem.*, 281(5): 2578-2584

Kinetic modeling of metabolism

- Estimation of parameter values from time-series data on metabolite concentrations in wild-type and mutant strains
- Model has **good predictive capability**: growth kinetics well explained in variety of conditions



Bettenbrock *et al.* (2005), *J. Biol. Chem.*, 281(5): 2578-2584

Kinetic modeling of metabolism

- Estimation of parameter values from time-series data on metabolite concentrations in wild-type and mutant strains
- Parameterization of reaction rate functions is **major hurdle** towards quantitative models:
  - Noisy and partial observations in available datasets
  - Heterogeneous experimental methods and conditions
  - Large size of networks
  - Parameter estimation algorithms: heuristic and computationally costly

Ashyraliyev *et al.* (2009), *FEBS J.*, 276:886-902

van Riel (2006*), Brief. Bioinform.*, 7(4):364–74

Kinetic modeling of metabolism

- **Alternative:** model reduction and approximation
- **Linlog models** approximate classical enzymatic rate laws:

$$v(x,u,e) = \mathrm{diag}(e) \cdot \left(a + B^x \cdot \ln(x) + B^u \cdot \ln(u)\right)$$

  - Internal and external metabolite concentrations $x \in \mathbb{R}_+^n$, $u \in \mathbb{R}_+^p$
  - Enzyme concentrations $e \in \mathbb{R}_+^m$
  - Parameters $a \in \mathbb{R}^m$, $B^x \in \mathbb{R}^{m \times n}$ and $B^u \in \mathbb{R}^{m \times p}$

Heijnen (2005), *Biotechnol. Bioeng.*, 91(5):534-45

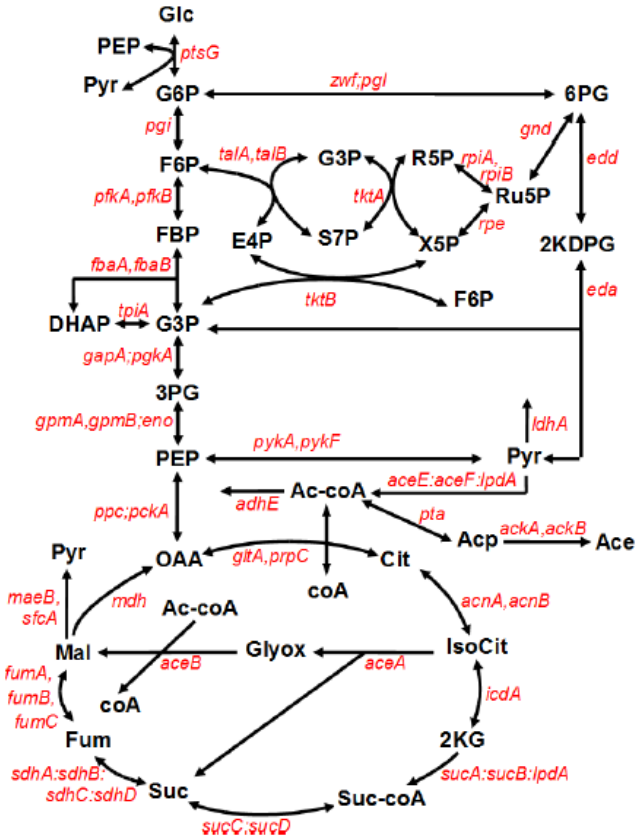
- Linlog models have several advantages for our purpose:
  - **Good local approximation** of classical rate laws
  - Parameter estimation reduced to **linear regression** problem
  - Parameters have biological interpretation (**elasticities**)

Kinetic modeling of metabolism

- Consider metabolic system at (quasi-) steady state

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$

$$N \cdot v = 0$$



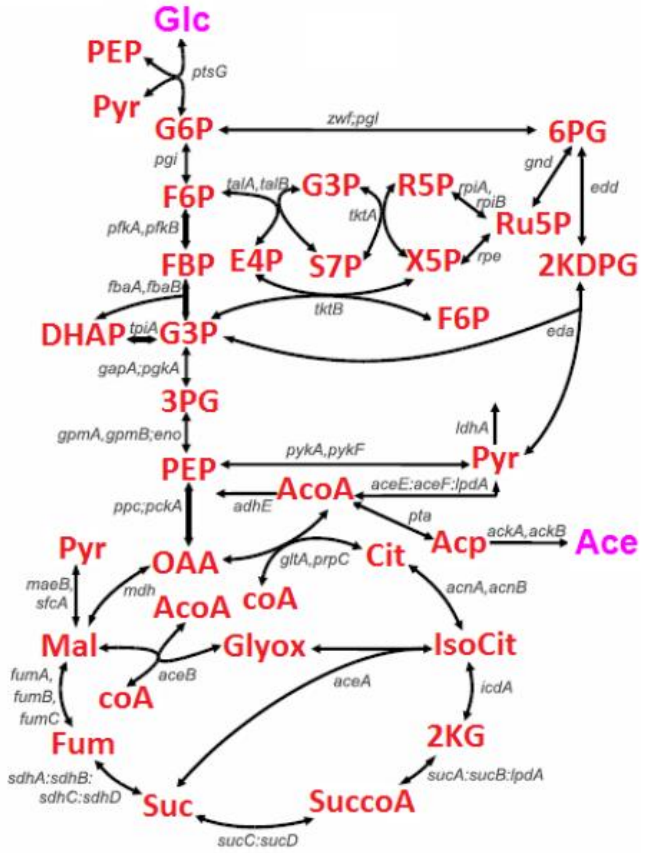
Kinetic modeling of metabolism

- Consider metabolic system at (quasi-) steady state

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$

$$N \cdot v = 0$$

- Measurements:
  - Metabolite concentrations u, x

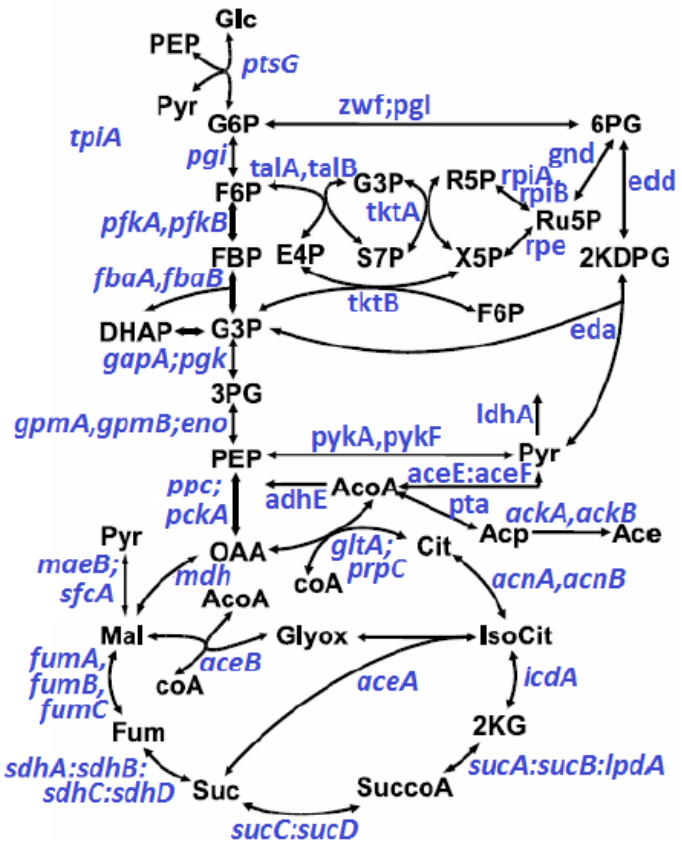


Kinetic modeling of metabolism

- Consider metabolic system at (quasi-) steady state

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$
$$N \cdot v = 0$$

- Measurements:
  - Metabolite concentrations u, x
  - Enzyme concentrations e



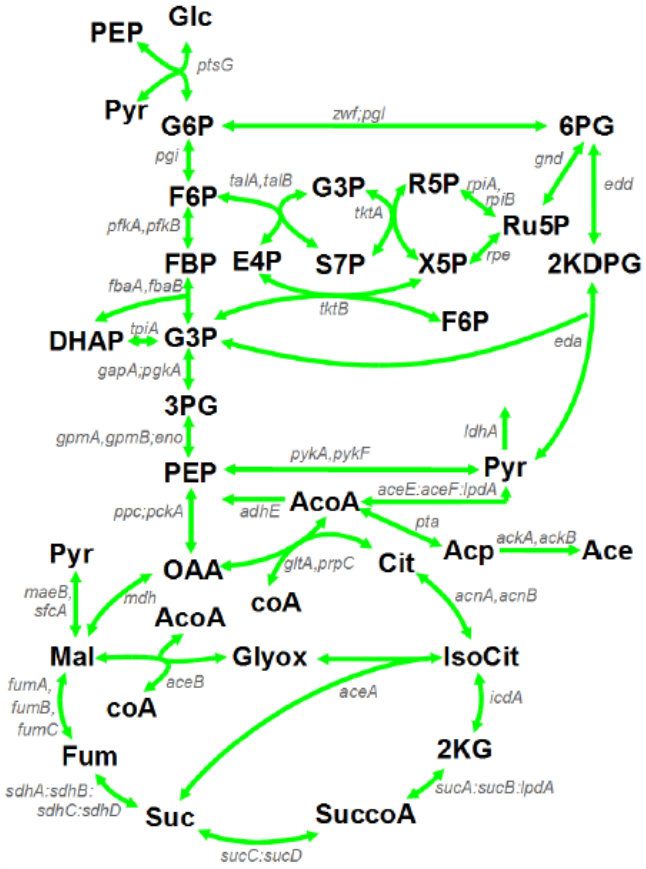
Kinetic modeling of metabolism

- Consider metabolic system at (quasi-) steady state:

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$
$$N \cdot v = 0$$

- Measurements:
  - Metabolite concentrations u, x
  - Enzyme concentrations e
  - Metabolic fluxes (reaction rates at steady state) J
- Linear regression problem:

$$\left(\frac{v}{e}\right)^T = [1 \ \ln x^T \ \ln u^T] \cdot \begin{bmatrix} a^T \\ (B^x)^T \\ (B^u)^T \end{bmatrix}$$



Kinetic modeling of metabolism

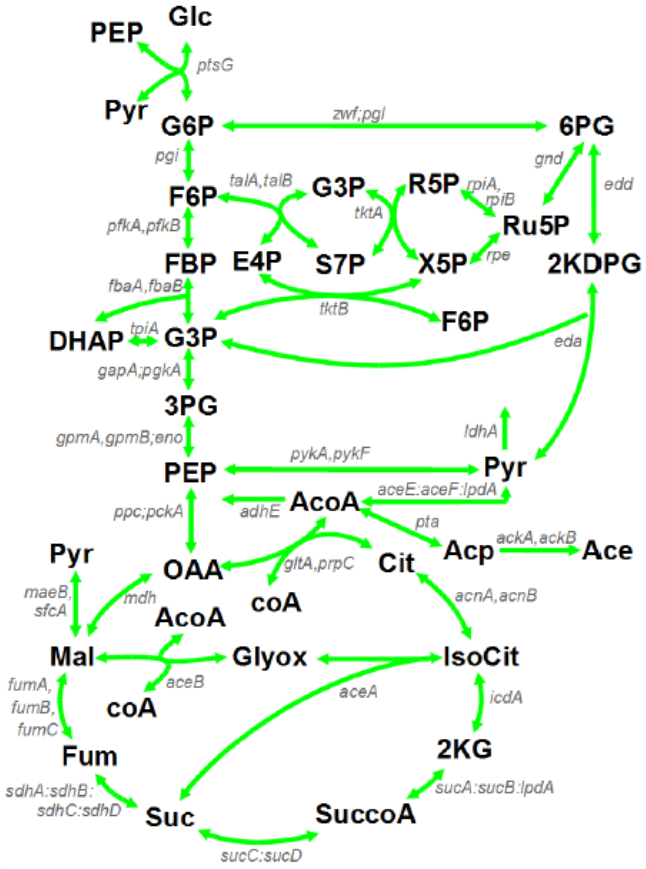
- Consider metabolic system at (quasi-) steady state:

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$

$$N \cdot v = 0$$

- Measurements:
  - Metabolite concentrations u, x
  - Enzyme concentrations e
  - Metabolic fluxes (reaction rates at steady state) J
- Parameters to estimate:

$$a \quad B^x \quad B^u$$



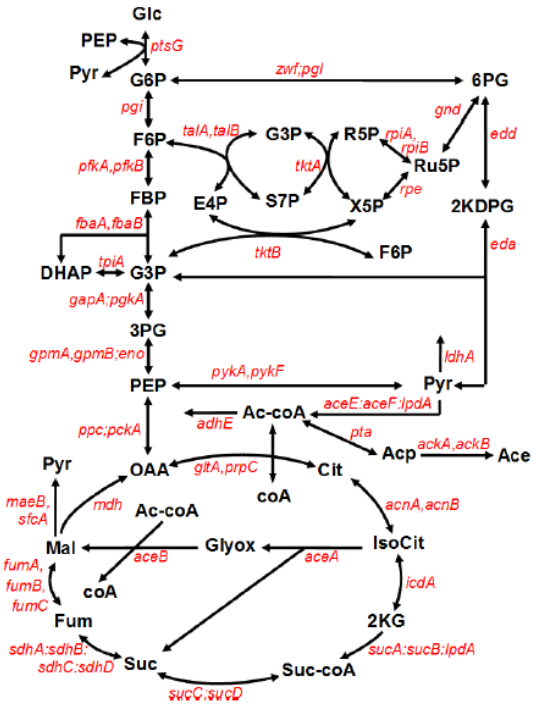
Kinetic modeling of metabolism

- High-throughput data sets are becoming available that allow estimation of parameters in linlog models

  Parallel measurement of enzyme and metabolite concentrations, and metabolic fluxes

  Ishii *et al.* (2007), *Science*, 316(5284):593-7

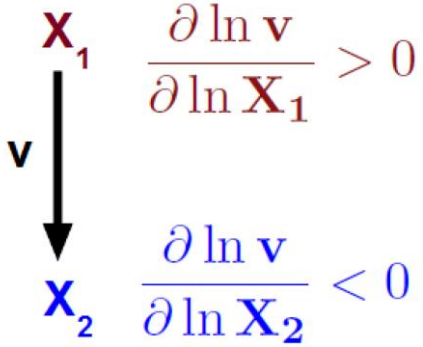
- Estimation of parameters in linlog models from experimental data
  - Technical problem: missing data
  - Expectation-Maximization (EM) approach for estimation of parameter values, tailored to linlog models



Berthoumieux *et al.* (2011), *Bioinformatics*, 27(13):i186-95

Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?
- Evaluation of results by comparing estimated and known signs of parameters
  - Parameters have biological interpretation (**elasticities**)
  - Elasticities capture normalized local response of reaction rates to changes in metabolite concentration
  - Elasticities and parameters $\left[B^x \;\; B^u\right]^T$ have same sign



Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?

Metabolites

Reactions	Glc	PEP	G6P	Pyr	F6P	FBP	DHAP	3PG	AcoA coA	6PG	Ru5P	R5P	S7P	2KG	Suc	Fum	Mal	ATP ADP	Cit	NADPH NADP	NADH NAD	FAD	Ace
	0.29	-0.89	0.79	1.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	-0.33	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	-0.16	0	0	0.04	-0.28	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0	0	0	0
	0	0	0	0	0	-0.3	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	-0.07	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	-0.18	0	-0.05	0	0	0	0	0	0	0	0	0	0.32	0	0	-0.17	0	0
	0	0.26	0	0	0	0	0	-0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0.12	0	0.49	0	-0.19	0	0	0	0	0	0	0	0	0	0	0	0.16	0	0	0	0	0
	0	0	0	0.64	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	-0.21	0	0
	0	0	-0.22	0	0	0	0	0	0	-0.24	0	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0.48	-0.09	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0	0.46	0	-0.39	0	0	0	0	0	0	0	0	0	0
	0	0	-0.74	0	0	0	0	0	0	0	0.3	-0.16	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0.01	-0.1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.32	0	0	0	0	0	0	0	0.51	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.58	0	0	0	0	0	0.35	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0.22	0	0	0	0	-0.001	0	0	0	0	0.49	0	-0.01	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	0	0	0	0	0.55	0	0	0	0
	0	0	0	0	0	0	0	0	0.08	0	0	0	0	-0.09	0	0	0	0	0	-0.59	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	1.26	0.3	0	0	0	0	0	-0.48	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.08	-0.44	0	0	0	0	0	0.4	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.46	0.1	0	0	0	0	0	0
	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.01	0	0.31	0	-0.12	0	0
	0	0.29	0	0	0	-0.13	0	0	0	0	0	0	0	0	0	0	0.1	-1.15	0.21	0	0	0	0
	0	0	0	-0.31	0	0	0	0	-0.21	0	0	0	0	0	0	0	0.38	0	0	0.36	-0.05	0	0
	0	0	0	0	0	0	0	0	-0.11	0	0	0	0	0	0.26	0	-0.18	0	0	0	0	0	0
	0	0.1	-0.09	0.04	-0.06	0	0	0.17	0.1	0	0	0.09	0	-0.01	0	0	0	-0.46	0	-0.003	0.01	0	0
	0	0	0	-0.03	0	0	0	0	0	-0.93	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	-0.06	0	0	0	0	-0.01	0	0	0	0	0	0	0	0	0.21	0	-0.25	-2.03	0	2.19
	0	0	0	2.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.11	0	0
	0	0	0	0	0	0	0	0	0.91	0	0	0	0	0	0	0	0	0	0	0	0	0	0

26 nonidentifiable parameters
40 estimated signs not significant at 95% confidence

20 signs correctly estimated
14 signs wrongly estimated

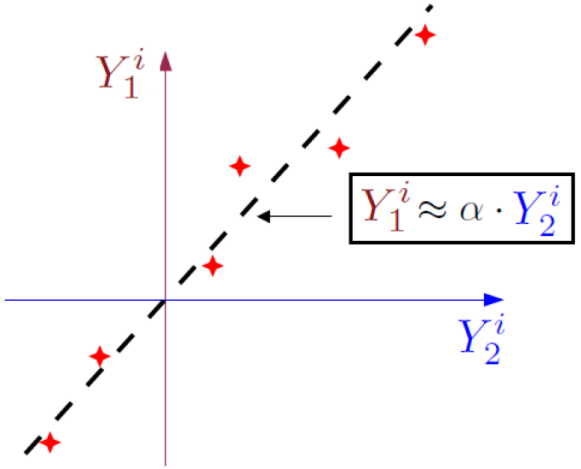
Berthoumieux *et al.* (2011), *Bioinformatics*, 27(13):i186-95

Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?
- Further analysis pinpoints **identifiability issues**:

  Different parametrizations lead to same predictions

Berthoumieux *et al.* (2012), *J. Math. Biol*, in press



$$w_i = \begin{bmatrix} Y_1^i & Y_2^i \end{bmatrix} \cdot \begin{bmatrix} c_{i1} \\ c_{i2} \end{bmatrix}$$

$$\Updownarrow$$

$$w_i = Y_1^i \cdot c_{i1}^* \quad \text{with} \quad c_{i1}^* = c_{i1} + \alpha \cdot c_{i2}$$

Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?
- Further analysis pinpoints **identifiability issues**:

Metabolites

Reactions	Glc	PEP	G6P	Pyr	F6P	FBP	DHAP	3PG	AcoA coA	6PG	Ru5P	R5P	S7P	2KG	Suc	Fum	Mal	ATP ADP	Cit	NADPH NADP	NADH NAD	FAD	Ace
	0.29	-0.89	0.79	1.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	-0.33	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	-0.16	0	0	0.04	-0.28	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0	0	0	0
	0	0	0	0	0	-0.3	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	-0.07	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	-0.18	0	-0.05	0	0	0	0	0	0	0	0	0	0.32	0	0	-0.17	0	0
	0	0.26	0	0	0	0	0	-0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0.12	0	0.49	0	-0.19	0	0	0	0	0	0	0	0	0	0	0	0.16	0	0	0	0	0
	0	0	0	0.64	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	-0.21	0	0
	0	0	-0.22	0	0	0	0	0	0	-0.24	0	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0.48	-0.09	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0	0.46	0	-0.39	0	0	0	0	0	0	0	0	0	0
	0	0	-0.74	0	0	0	0	0	0	0	0.3	-0.16	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0.01	-0.1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.32	0	0	0	0	0	0	0	0.51	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.58	0	0	0	0	0	0.35	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0.22	0	0	0	0	-0.001	0	0	0	0	0.49	0	-0.01	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	0	0	0	0	0.55	0	0	0	0
	0	0	0	0	0	0	0	0	0.08	0	0	0	0	-0.09	0	0	0	0	0	-0.59	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	1.26	0.3	0	0	0	0	0	-0.48	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.08	-0.44	0	0	0	0	0	0.4	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.46	0.1	0	0	0	0	0	0
	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.01	0	0.31	0	-0.12	0	0
	0	0.29	0	0	0	-0.13	0	0	0	0	0	0	0	0	0	0	0.1	-1.15	0.21	0	0	0	0
	0	0	0	-0.31	0	0	0	0	-0.21	0	0	0	0	0	0	0	0.38	0	0	0.36	-0.05	0	0
	0	0	0	0	0	0	0	0	-0.11	0	0	0	0	0	0.26	0	-0.18	0	0	0	0	0	0
	0	0.1	-0.09	0.04	-0.06	0	0	0.17	0.1	0	0	0.09	0	-0.01	0	0	0	-0.46	0	-0.003	0.01	0	0
	0	0	0	-0.03	0	0	0	0	0	-0.93	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	-0.06	0	0	0	0	-0.01	0	0	0	0	0	0	0	0	0.21	0	-0.25	-2.03	0	2.19
	0	0	0	2.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.11	0	0
	0	0	0	0	0	0	0	0	0.91	0	0	0	0	0	0	0	0	0	0	0	0	0	0

27/31 nonidentifiable reactions 73/100 nonidentifiable parameters

Berthoumieux *et al.* (2012), *J. Math. Biol*, in press

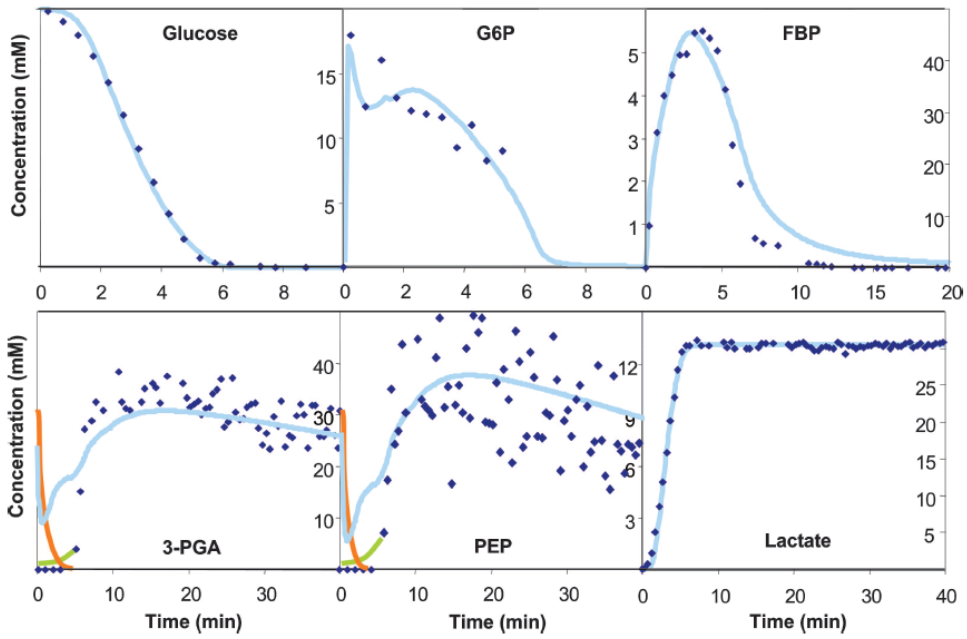
Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?
- Further analysis pinpoints **identifiability issues**
- State-of-the-art high-throughput data set does not allow to unambiguously infer most parameters in large metabolic network model
  - Noisy data (often of same order of magnitude as relevant information)
  - Little informative conditions (steady-state data, metabolism is homeostatic system)

Kinetic modeling of metabolism

- **Question:** can we reliably estimate parameter values using simplified model and state-of-the-art high-throughput data set?
- New experimental methods and techniques: **time-series data** that allow dynamic monitoring of metabolism
- Modelling glycolysis in Lactococcus lactis using approximate kinetic models

    Power-law models (cousin of linlog models)

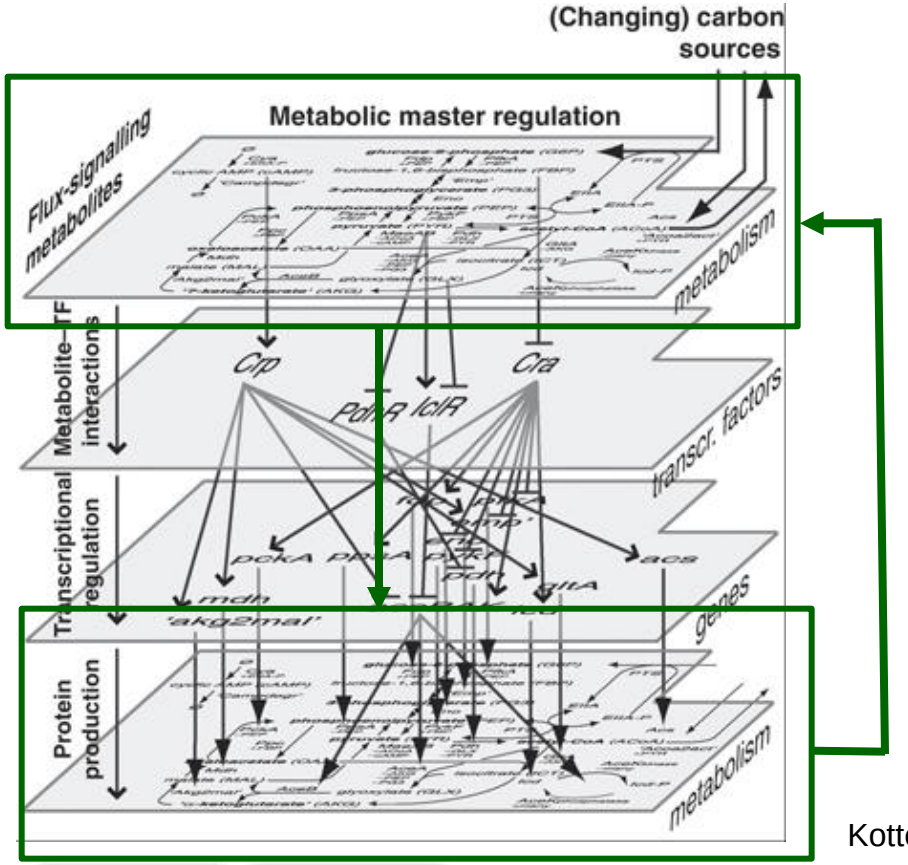


Voit *et al.* (2006), *Proc. Natl. Acad. Sci. USA*, 103(25):9452-7

Towards models of integrated networks

- Metabolic networks are integrated with gene networks and signalling networks

  Complex multi-level system with feedback across different time-scales



- **Fast response:** adaptation of metabolic fluxes and metabolite pools (metabolic network)
- **Slow response:** adaptation of enzyme and TF concentrations (gene regulatory network)
- **Feedback** across genetic and metabolic levels: complex system on different time-scales

Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6:355

Modeling regulatory networks

- Kinetic model of form $\dot{x} = N\, v(x)$
  - Concentration variables $x \in \mathbb{R}^n_+$
  - Reaction rates $v : \mathbb{R}^n_+ \to \mathbb{R}^q$
  - Stoichiometry matrix $N \in \mathbb{Z}^{n \times q}$
- Time-scale hierarchy motivates distinction between **fast** reaction rates $v^f \in \mathbb{R}^{q-p}$ and **slow** reaction rates $v^s \in \mathbb{R}^p$, such that

$$v = [v^s \; v^f]'$$

Typically, **enzymatic and complex formation** reactions are fast, **protein synthesis and degradation** are slow

Modeling regulatory networks

- Separation of fast and slow reactions motivates a linear transformation $T \in \mathbb{Z}^n \times \mathbb{Z}^n$ of the variables

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = T\,x \qquad \begin{bmatrix} N^s & 0 \\ N^{s'} & N^f \end{bmatrix} = T\,N$$

- We call $x^s \in \mathbb{R}_+^m$ **slow variables** and $x^f \in \mathbb{R}_+^{n-m}$ **fast variables**
- Separation of fast and slow variables allows $\dot{x} = N\,v(x)$ to be rewritten as coupled slow (genetic) and fast (metabolic) subsystems

$$\dot{x}^s = N^s\,v^s(x^s, x^f)$$
$$\dot{x}^f = N^{s'}\,v^s(x^s, x^f) + N^f\,v^f(x^s, x^f) \approx N^f\,v^f(x^s, x^f)$$

Modeling regulatory networks

- Separation of fast and slow reactions motivates a linear transformation $T \in \mathbb{Z}^n \times \mathbb{Z}^n$ of the variables

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = T\,x \qquad \begin{bmatrix} N^s & 0 \\ N^{s'} & N^f \end{bmatrix} = T\,N$$

- We call $x^s \in \mathbb{R}_+^m$ **slow variables** and $x^f \in \mathbb{R}_+^{n-m}$ **fast variables**
- Separation of fast and slow variables allows $\dot{x} = N\,v(x)$ to be rewritten as coupled slow (genetic) and fast (metabolic) subsystems

  Slow variables are typically **total protein concentrations**, fast variables **metabolites and biochemical complexes**

Modeling regulatory networks

- Separation of fast and slow reactions motivates a linear transformation $T \in \mathbb{Z}^n \times \mathbb{Z}^n$ of the variables

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = T\,x \qquad \begin{bmatrix} N^s & 0 \\ N^{s'} & N^f \end{bmatrix} = T\,N$$

- We call $x^s \in \mathbb{R}_+^m$ **slow variables** and $x^f \in \mathbb{R}_+^{n-m}$ **fast variables**
- Separation of fast and slow variables allows $\dot{x} = N\,v(x)$ to be rewritten as coupled slow (genetic) and fast (metabolic) subsystems

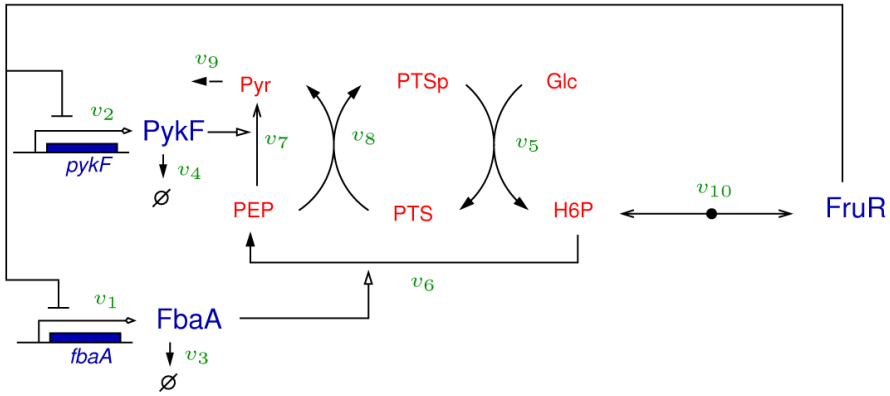
  Slow variables are typically **total protein concentrations**, fast variables **metabolites and biochemical complexes**

Modeling regulatory networks: example

- Separation of fast and slow variables allows original model to be rewritten as coupled slow (genetic) and fast (metabolic) subsystems

$$\dot{x}^s = N^s\, v^s(x^s, x^f)$$

$$\dot{x}^f = N^{s'}\, v^s(x^s, x^f) + N^f\, v^f(x^s, x^f) \approx N^f\, v^f(x^s, x^f)$$



$$\begin{bmatrix} \dot{x}_{FbaA} \\ \dot{x}_{PykF} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1(x_{FruR\cdot free}) \\ v_2(x_{FruR\cdot free}) \\ v_3(x_{FbaA}) \\ v_4(x_{PykF}) \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_{H6P} \\ \dot{x}_{PEP} \\ \dot{x}_{Pyr} \\ \dot{x}_{PTSp} \\ \dot{x}_{FruR\cdot free} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 2 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_5(x_{Glc}, x_{PTSp}) \\ v_6(x_{H6P}, x_{PEP}, x_{FbaA}) \\ v_7(x_{Pyr}, x_{PEP}, x_{PykF}) \\ v_8(x_{PEP}, x_{Pyr}, x_{PTSp}) \\ v_9(x_{Pyr}) \\ v_{10}(x_{H6P}, x_{FruR\cdot free}) \end{bmatrix}$$

Baldazzi *et al.* (2010), *PLoS Comput. Biol.*, 6(6):e1000812

Modeling regulatory networks

- Separation of fast and slow variables allows original model to be rewritten as coupled slow (genetic) and fast (metabolic) subsystems

$$\dot{x}^s = N^s \, v^s(x^s, x^f)$$

$$\dot{x}^f = N^{s'} \, v^s(x^s, x^f) + N^f \, v^f(x^s, x^f) \approx N^f \, v^f(x^s, x^f)$$

- Under **quasi-steady-state approximation (QSSA)**, fast variables are assumed to instantly adapt to slow dynamics

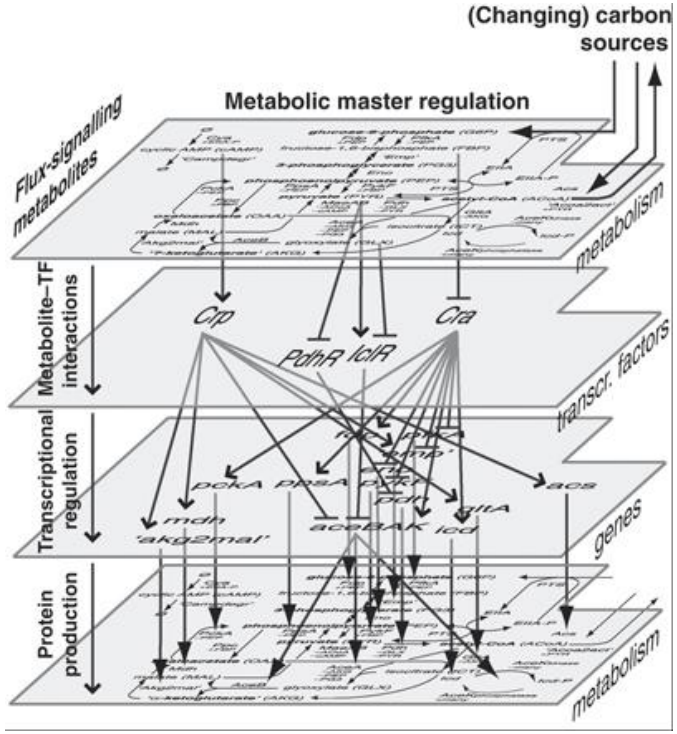
$$\dot{x}^f = 0 \;\Rightarrow\; N^f \, v^f(x^s, x^f) = 0$$

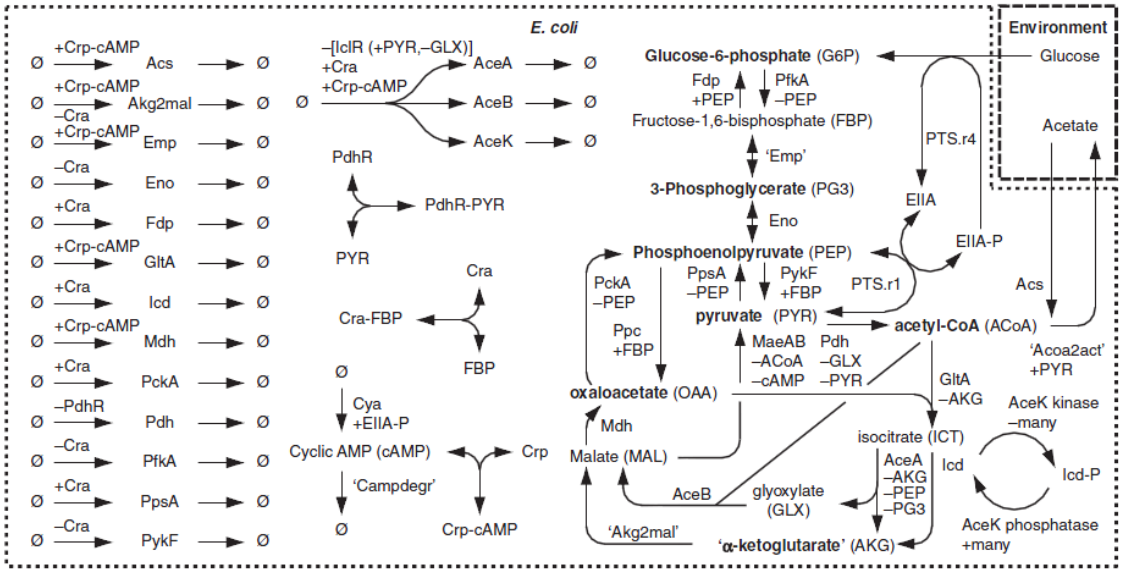
Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall

Towards models of integrated networks

- Coupling of gene expression and metabolism into a single integrated model

$$\dot{x}^s = N^s \, v^s(x^s, x^f)$$
$$N^f \, v^f(x^s, x^f) = 0$$

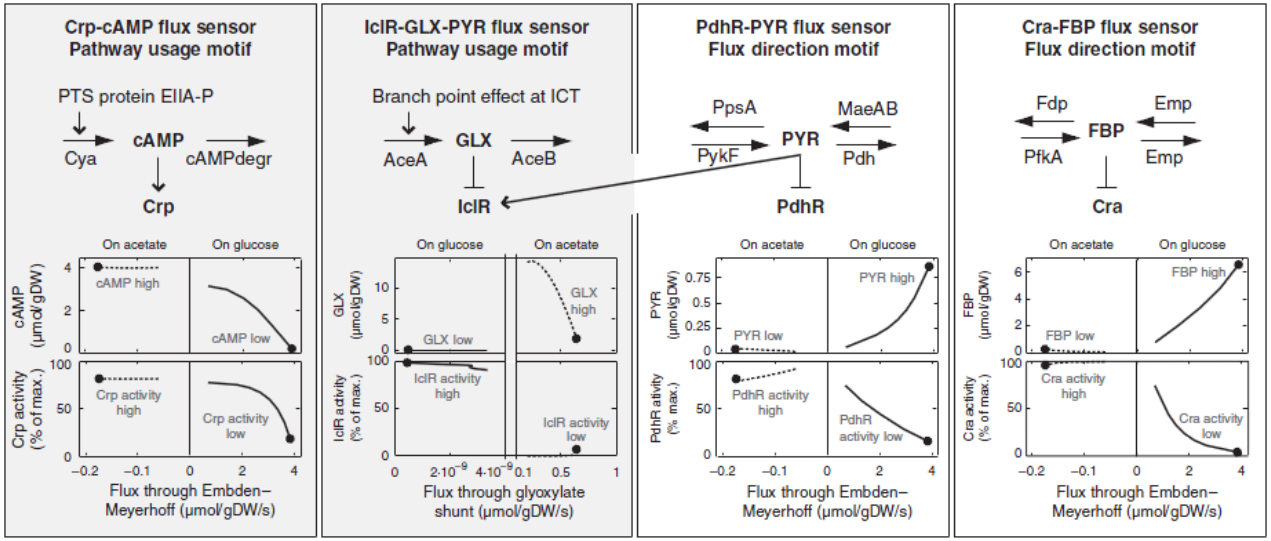




Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6: 355

Towards models of integrated networks

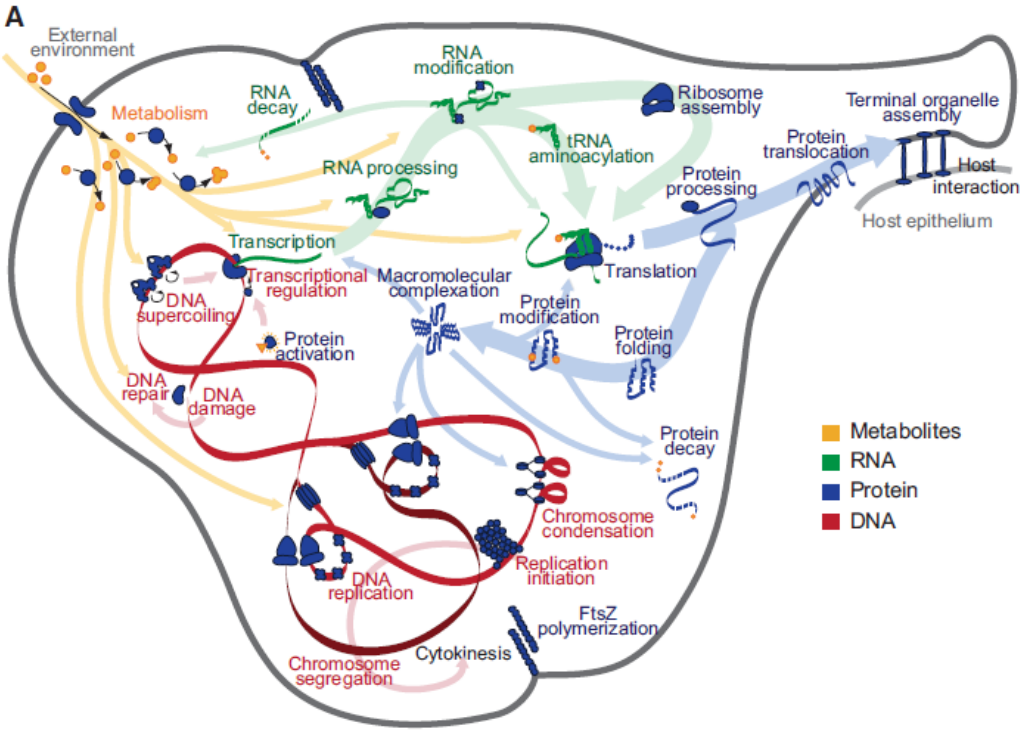
- Kinetic model with 47 variables and 193 parameters
  - Parameters estimated from published experimental steady-state data sets for balanced growth on either glucose or acetate
- Analysis of model shows that adaptation to change in carbon source is achieved by distributed sensing of intracellular fluxes



Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6: 355

Towards models of integrated networks

- Metabolic networks are integrated with gene networks and signalling networks
  - Complex multi-level system with feedback across different time-scales



Whole-cell model of *Mycoplasma genitalium*

Karr *et al.* (2012), *Cell*, 150: 389-401

Towards models of integrated networks

- Metabolic networks are integrated with gene networks and signalling networks
  - Complex multi-level system with feedback across different time-scales
- Upscaling of analysis to large networks of dozens or even hundreds of genes, proteins, metabolites, …
  - Model reduction, qualitative models, and formal verification tools

Conclusions

- Adaptation of bacteria to their environment involves reorganization of cellular physiology
- Increasingly powerful methods have become available to experimentally quantify cellular adaptation
  - Transcriptomics, proteomics, fluxomics, metabolomics, …
- Adaptation process achieved by large and complex regulatory networks
  - Nonlinear dynamical systems with feedback across different time-scales
- Mathematical modeling and computer simulation helpful in making sense out of multiple sources of data
  - Systems biology

Conclusions

- Issues for integration of different data sources into quantitative, predictive mathematical models:
  - Noisy data, obtained with heterogeneous methods in different experimental conditions
  - Parameter estimation and model identifiability
  - Complex models with many variables evolving on different time-scales
- How can we make useful and reliable explanations/ predictions with the help of such large mathematical/ computational objects?
- To which extent can these approaches be applied to more complex multicellular organisms? Evolving in communities and ecosystems?

Merci !

www.inrialpes.fr/ibis