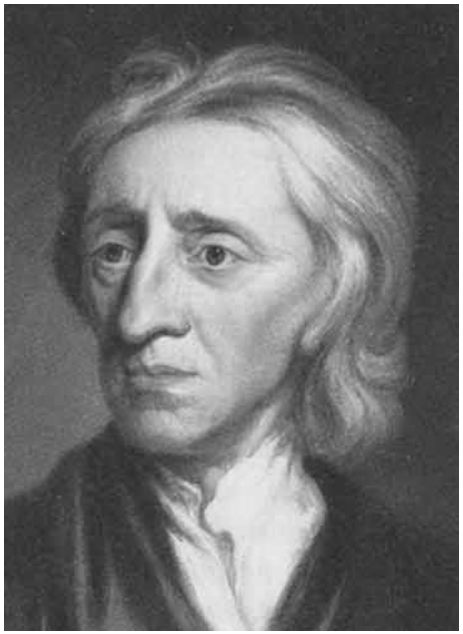
Engineering Genetic Circuits

Chris J. Myers

Lecture 13: Flux Balance Analysis

John Locke



Things of this world are in so constant a flux, that nothing remains long in the same state.

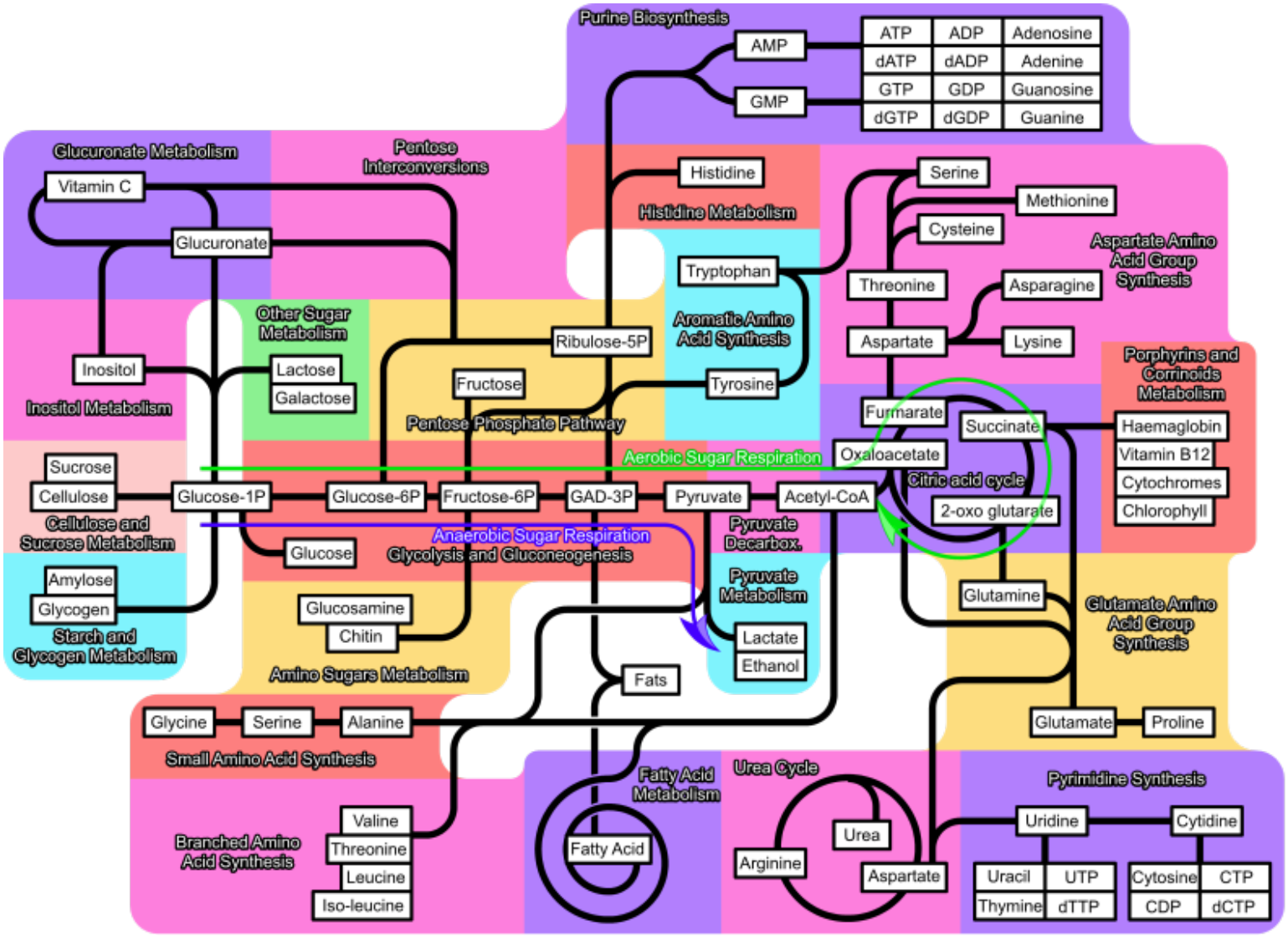
Overview

- *Metabolic networks*
- *Flux balance analysis* (FBA)

Metabolic Networks

- *Metabolic networks* consist of a series of chemical reactions that modify a *substrate* to produce a new biomolecule within a cell.
- Product may be used immediately, initiate another pathway, or stored.
- Metabolic reactions are often *catalyzed* by *enzymes*.
- Enzymes may be gene products allowing for genetic control.

Metabolic Networks

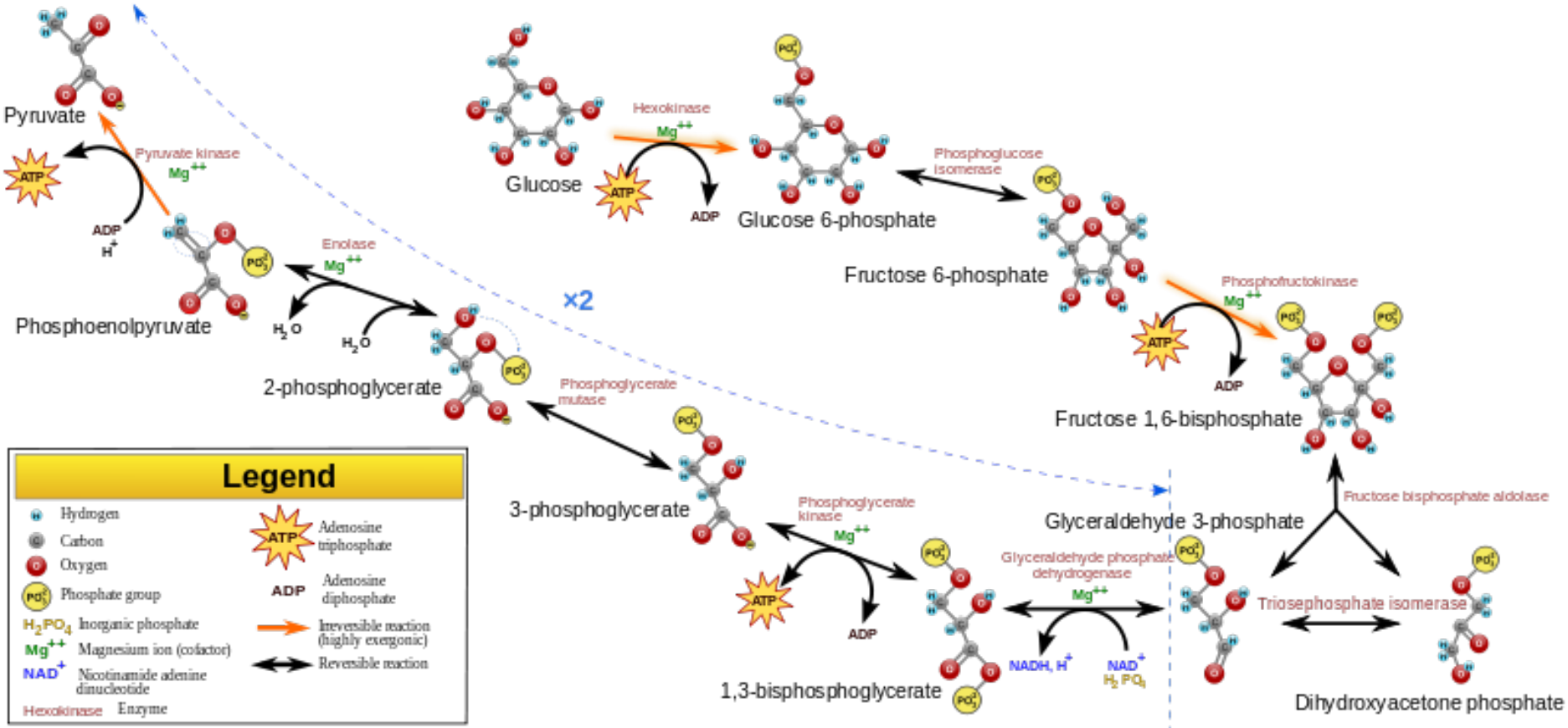


(Courtesy of Wikipedia)

Cellular Respiration

- The energy conversion pathways of a cell.
- Complex sugars broken down into *glucose* that enters the cell through *glucose transporters* in the cell's membrane.
- Cellular respiration breaks down glucose to make ATP using:
  - *Glycolysis*
  - *Citric acid (Kreb's) cycle*
- Products of each pathway are reactants in following pathway.
- Glycolysis occurs in the cytoplasm while later steps occur in mitochondria.
- Capable of generating enough ATP to run all the cell functions.

Glycolysis

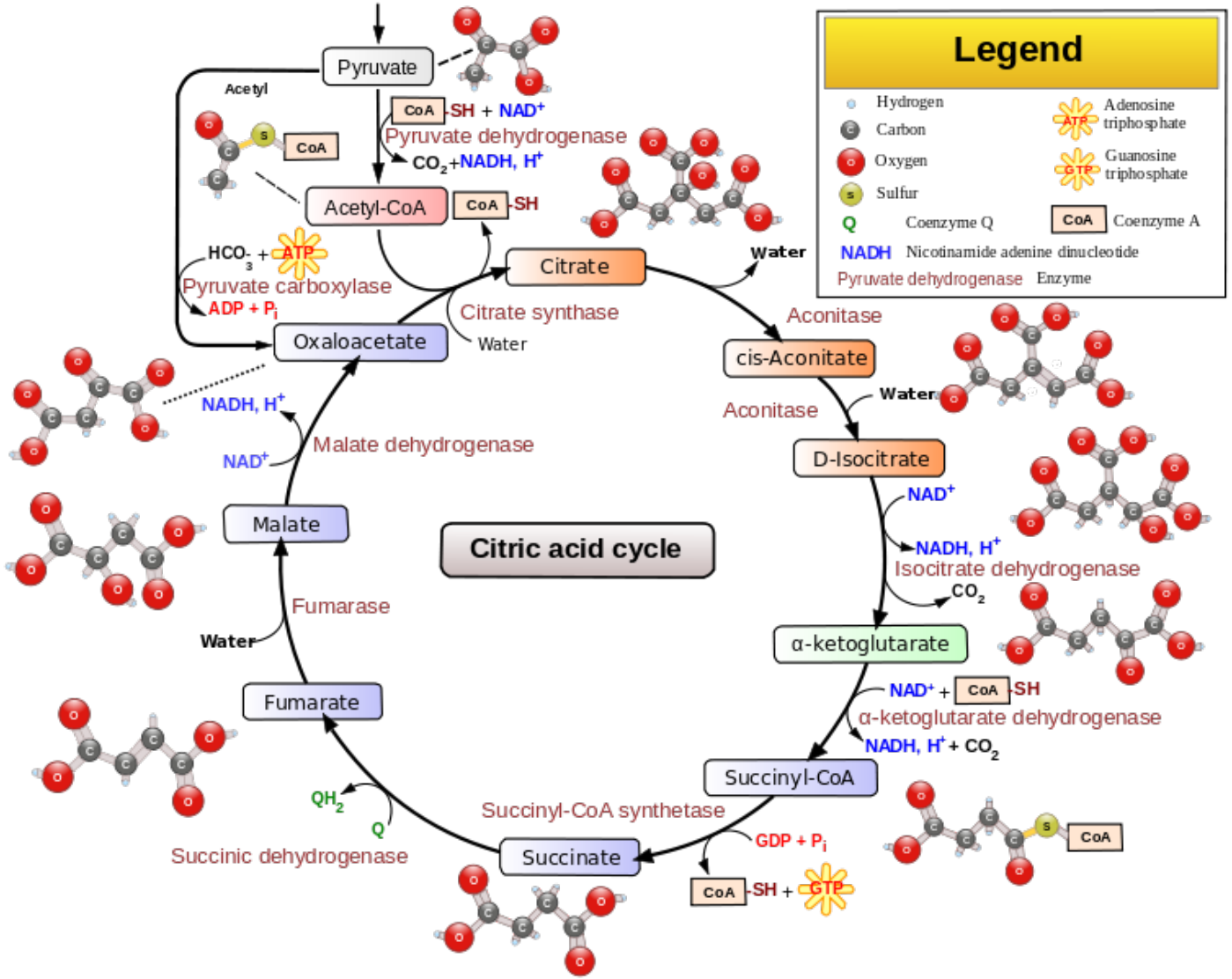


(Courtesy of Wikipedia)

Glycolysis

- Requires no oxygen and is referred to as *anaerobic metabolism*.
- Glycolysis occurs in the cytoplasm outside the mitochondria.
- Glucose is broken down into a molecule called *pyruvate*.
- Each reaction produces hydrogen ions to make ATP.
- Only four ATP molecules can be made from one of glucose.
- In prokaryotes, glycolysis is the only method to produce energy.

Citric Acid (Kreb's) Cycle



(Courtesy of Wikipedia)

Pyruvate Oxidation

- Begins with two molecules of pyruvic acid.
- Next, pyruvic acid is altered by the removal of a carbon and two oxygens, which go on to form carbon dioxide.
- When CO_2 is removed, energy is given off, and $NAD+$ molecule is converted into the higher energy form *NADH*.
- Another molecule, *coenzyme A*, then attaches to the remaining acetyl unit, forming *acetyl CoA*.

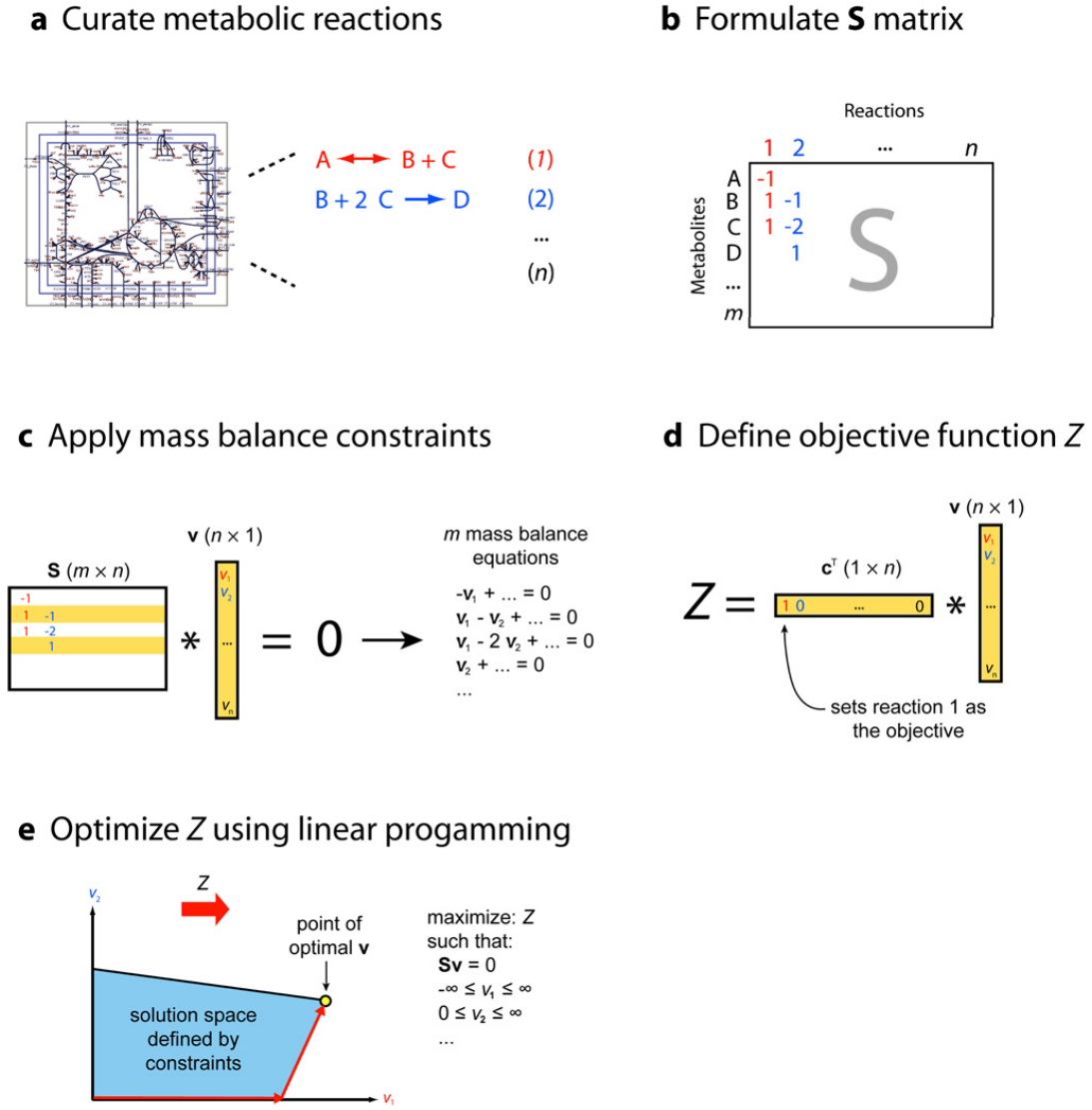
Citric Acid (Kreb's) Cycle

- Acetyl CoA binds to a four-carbon molecule called *oxaloacetate* to make a six-carbon molecule called *citric acid*.
- Citric acid is then broken down and modified in a stepwise fashion releasing hydrogen ions and carbon molecules.
- The carbon molecules are used to make more carbon dioxide.
- The hydrogen ions are picked up by NAD and another molecule called *flavin-adenine dinucleotide* (FAD).
- Eventually, the process produces the four-carbon oxaloacetate again, ending up where it started off.
- Further processing of the ions released can result in 24 to 28 ATP molecules from one molecule of glucose.

Flux Balance Analysis (FBA)

- Most analysis methods require detailed kinetic information.
- Metabolic networks can be reconstructed from an annotated genome and literature.
- FBA determines theoretical capabilities of metabolic networks using only stoichiometry and fundamental physiochemical capacity constraints.
- Capacity constraints include maximum uptake rates of oxygen and substrates, such as, glucose, acetate, lactose, etc.
- FBA determines optimal flux distribution for given conditions.
- Cell growth is used as the objective function approximated by production of growth precursors in certain ratios.

FBA Procedure



(Courtesy of Orth et al., Nature Biotechnology, 2010)

Metabolic Network Reconstruction

- Identify *homologous genes* in sequence data to assign functions.
- Examine metabolic reaction database to find:
  - The substrates, products, and stoichiometry of each metabolic reaction,
  - The name of the enzyme catalyzing the reaction, and
  - The genes that code for the respective enzymes.
- Review literature for metabolic genes/reactions not found in the database.

Hypothetical Metabolic Genotype

Gene	Enzyme	Flux
$Gene_1$	A Transporter	R_1
$Gene_2$	$Enzyme_2$	R_2
$Gene_3$	$Enzyme_3$	R_3
$Gene_4$	D Transporter	R_4
$Gene_5$	$Enzyme_5$	R_5
$Gene_6$	$Enzyme_6$	R_6
$Gene_7$	F Transporter	R_7
$Gene_8$	$Enzyme_8$	R_8
$Gene_9$	$Enzyme_9$	R_9
$Gene_{10}$	H Transporter	R_{10}
-	A Exchange	A_{xt}
-	D Exchange	D_{xt}
-	F Exchange	F_{xt}
-	H Exchange	H_{xt}

Figure 2a: Hypothetical metabolic genotype

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Hypothetical Metabolic Network

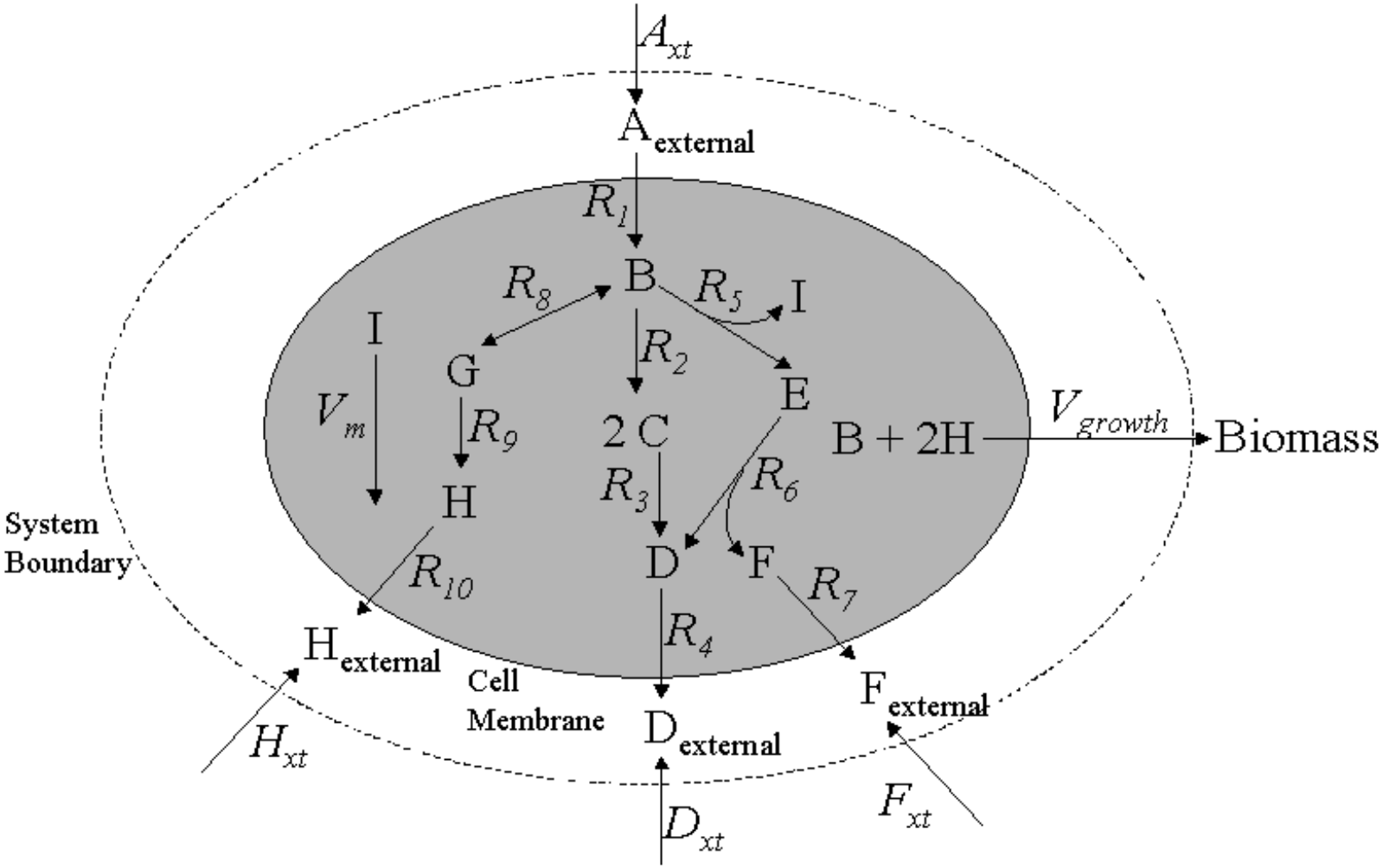


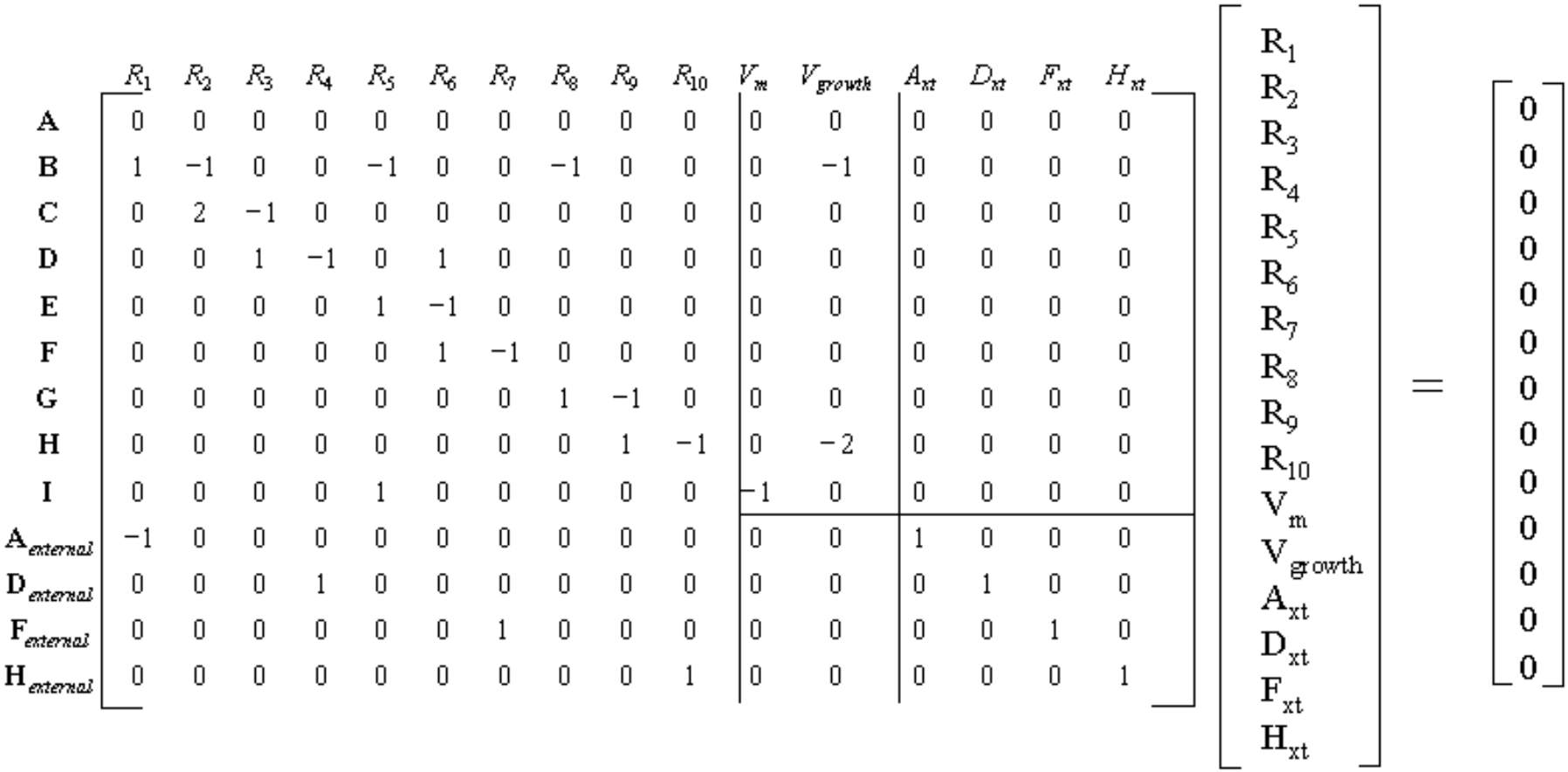
Figure 2b: Hypothetical metabolic network

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Formulate Stoichiometric Matrix

- Metabolic reactions defined by a $m \times n$ *stoichiometric matrix*, **S**.
  - Each of the m rows represents a chemical species.
  - Each of the n columns represents a chemical reaction.
  - The $S_{i,j}$ entry is the net stoichiometry for species i in reaction j.

Stoichiometric Matrix for Hypothetical Metabolic Network



	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	V_m	V_{growth}	A_{xt}	D_{xt}	F_{xt}	H_{xt}
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	1	−1	0	0	−1	0	0	−1	0	0	0	−1	0	0	0	0
C	0	2	−1	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	−1	0	1	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	1	−1	0	−2	0	0	0	0
I	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	0
$A_{external}$	−1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
$D_{external}$	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
$F_{external}$	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
$H_{external}$	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1

$$\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \\ R_9 \\ R_{10} \\ V_m \\ V_{growth} \\ A_{xt} \\ D_{xt} \\ F_{xt} \\ H_{xt} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Figure 4: Stoichiometric matrix for the hypothetical metabolic network shown in Figure 2b.

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Defining Mass Balance Constraints

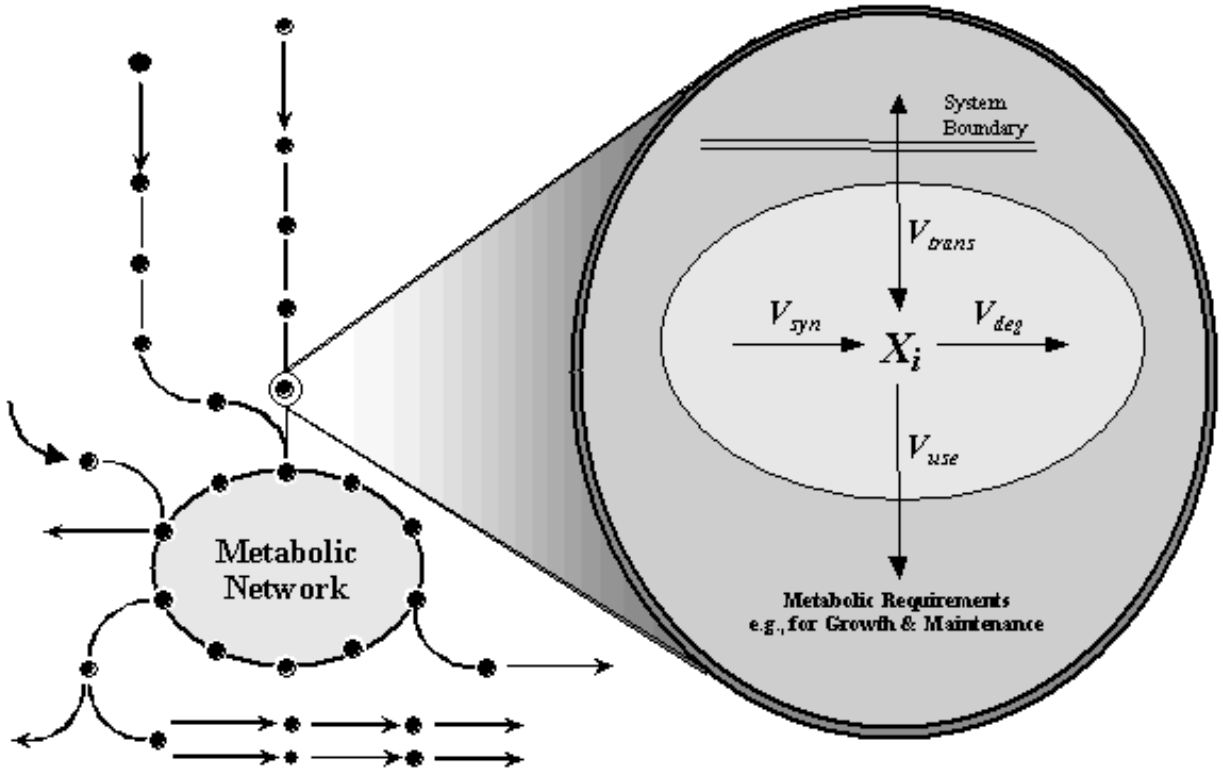


Figure 1: Defining the mass balance constraints. The mass balance constraints are defined by summing the rates of production and degradation for each metabolite in the network.

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Apply Mass Balance Constraints

- Rate of accumulation of X_i:

$$\begin{aligned} \frac{dX_i}{dt} &= V_{syn} - V_{deg} - V_{use} \pm V_{trans} \\ &= V_{syn} - V_{deg} - V_{use} + b_i \end{aligned}$$

- Single equation for complete metabolic network:

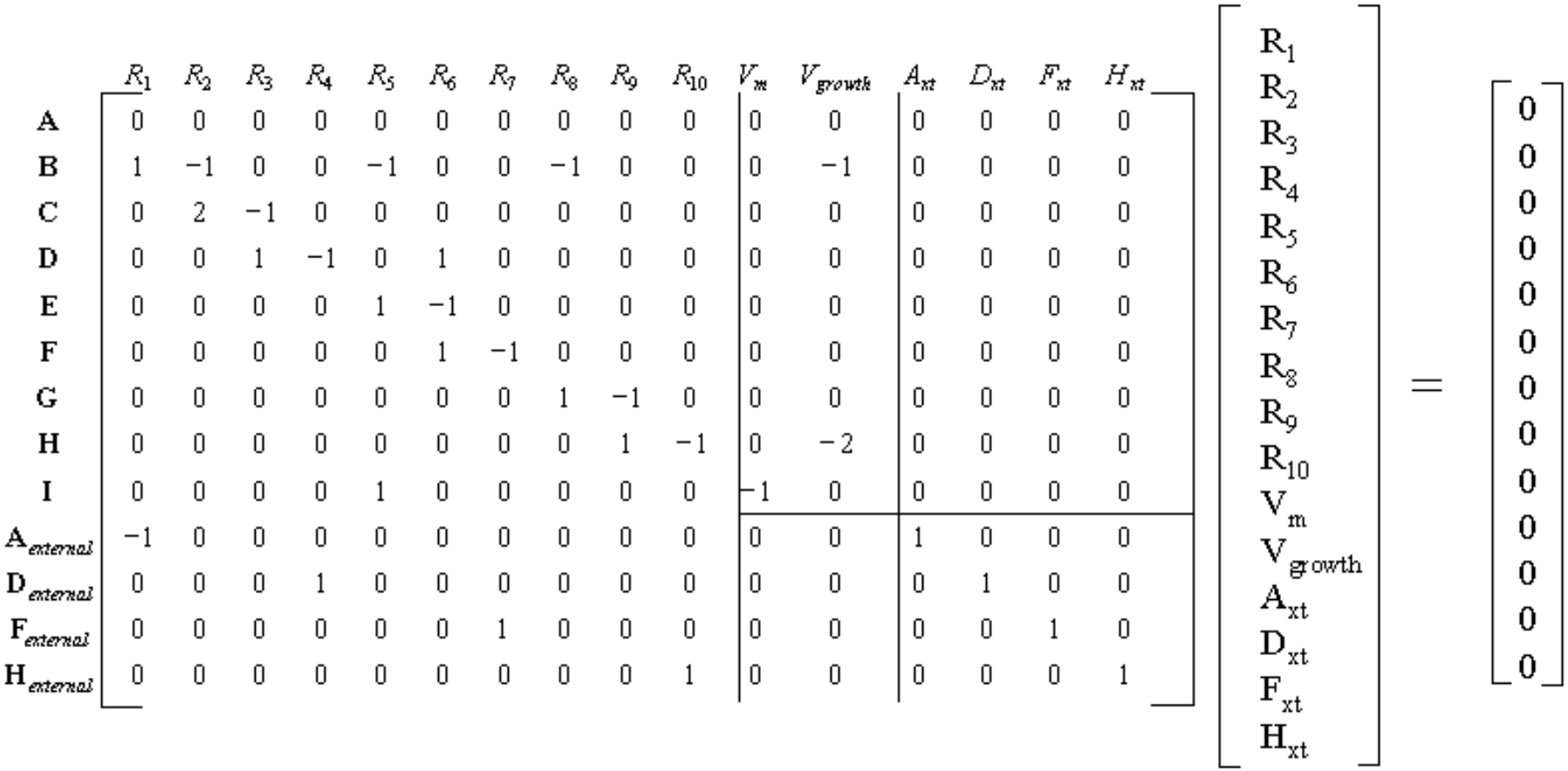
$$\frac{\mathbf{dX}}{dt} = \mathbf{S} \cdot \mathbf{v} + \mathbf{b}$$

where $\mathbf{S}$ is the $m \times n$ stoichiometric matrix, $\mathbf{v}$ is the vector of metabolic fluxes, and $\mathbf{b}$ is the vector of metabolic exchange fluxes.

- Assuming steady-state yields:

$$\begin{aligned} \mathbf{S} \cdot \mathbf{v} + \mathbf{I} \cdot \mathbf{b} &= 0 \\ [\mathbf{S}_{\text{reactions}} \mid \mathbf{S}_{\text{use}} \mid \mathbf{U}] \begin{bmatrix} \mathbf{v}_{\text{reactions}} \\ \mathbf{v}_{\text{use}} \\ \mathbf{b}_r \end{bmatrix} &= 0 \\ \mathbf{S}' \cdot \mathbf{v}' &= 0 \end{aligned}$$

Mass Balance for Hypothetical Metabolic Network



	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	V_m	V_{growth}	A_{xt}	D_{xt}	F_{xt}	H_{xt}			
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		R_1	0
B	1	−1	0	0	−1	0	0	−1	0	0	0	−1	0	0	0	0		R_2	0
C	0	2	−1	0	0	0	0	0	0	0	0	0	0	0	0	0		R_3	0
D	0	0	1	−1	0	1	0	0	0	0	0	0	0	0	0	0		R_4	0
E	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0		R_5	0
F	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0		R_6	0
G	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0		R_7	0
H	0	0	0	0	0	0	0	0	1	−1	0	−2	0	0	0	0		R_8	0
I	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	0		R_9	0
$A_{external}$	−1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		R_{10}	0
$D_{external}$	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0		V_m	0
$F_{external}$	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0		V_{growth}	0
$H_{external}$	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1		A_{xt}	0
																		D_{xt}	0
																		F_{xt}	0
																		H_{xt}	0

Figure 4: Stoichiometric matrix for the hypothetical metabolic network shown in Figure 2b.

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Additional Constraints

- Solutions of mass balance equation are metabolic flux distributions that do not violate mass, energy, and redox balance constraints.
- Many vectors though within this *nullspace* are not physiologically feasible.
- Capacity and thermodynamic constraints can be expressed as inequalities of the form: $\alpha_j \leq v_j \leq \beta_j$ where $j = 1 \ldots n$.

Constraints on the Hypothetical Metabolic Network

Mass Balances	Flux Constraints
$\mathbf{B}: R_1 - R_2 - R_5 - R_8 - V_{growth} = 0$	$0 \le R_1 \le \infty$
$\mathbf{C}: 2R_2 - R_3 = 0$	$0 \le R_2 \le \infty$
$\mathbf{D}: R_3 + R_6 - R_4 = 0$	$0 \le R_3 \le \infty$
$\mathbf{E}: R_5 - R_6 = 0$	$0 \le R_4 \le \infty$
$\mathbf{F}: R_6 - R_7 = 0$	$0 \le R_5 \le \infty$
$\mathbf{G}: R_8 - R_9 = 0$	$0 \le R_6 \le \infty$
$\mathbf{H}: R_9 - R_{10} - 2V_{growth} = 0$	$0 \le R_7 \le \infty$
$\mathbf{I}: R_5 - V_m = 0$	$-\infty \le R_8 \le \infty$
$\mathbf{A}_{external}: A_{xt} - R_1 = 0$	$0 \le R_9 \le \infty$
$\mathbf{D}_{external}: D_{xt} + R_4 = 0$	$0 \le R_{10} \le \infty$
$\mathbf{F}_{external}: F_{xt} + R_7 = 0$	$V_1 \le V_m \le V_1$
$\mathbf{H}_{external}: H_{xt} + R_{10} = 0$	$0 \le V_{growth} \le \infty$
	$V_2 \le A_{up} \le V_2$
	$-\infty \le D_{up} \le 0$
	$-\infty \le F_{up} \le 0$
	$-\infty \le H_{up} \le 0$

Objective Function

$$Z = V_{growth}$$

Figure 3: Constraints on the metabolic network. The constraints on the metabolic network consist of mass balance constraints and flux constraints (reversibility constraints). Linear programming can be used to determine the optimal use of the metabolic network subject to the imposed constraints.

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Define Objective Function

- Many metabolic flux vectors satisfy all constraints.
- Goal is to find one that maximizes (or minimizes) an *objective function*.

$$\text{Maximize } Z$$
$$\text{where } Z = \sum c_i \cdot v_i = \langle \mathbf{c} \cdot \mathbf{v} \rangle$$

- Objective for hypothetical metabolic network: $Z = V_{\text{growth}}$.

Optimize Using Linear Programming

- Canonical form of a flux balance analysis problem:

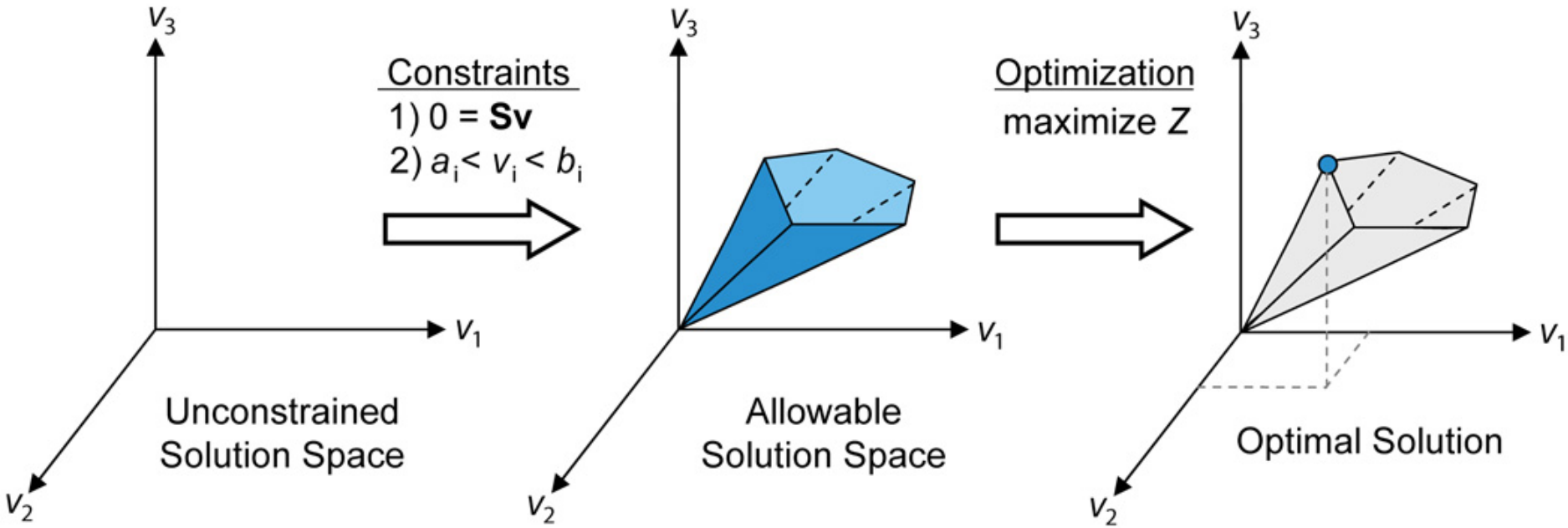
  maximize $\langle \mathbf{c} \cdot \mathbf{v} \rangle$

  subject to $\mathbf{S}' \cdot \mathbf{v}' = 0$

  and $\alpha \leq \mathbf{v}' \leq \beta$

- Can be solved using *linear programming*, a technique for finding an optimal solution in a convex space defined by linear equalities and inequalities.
- Many efficient algorithms and software packages exist.

Conceptional Basis for FBA



(Courtesy of Orth et al., Nature Biotechnology, 2010)

Phenotype Phase Plane Example

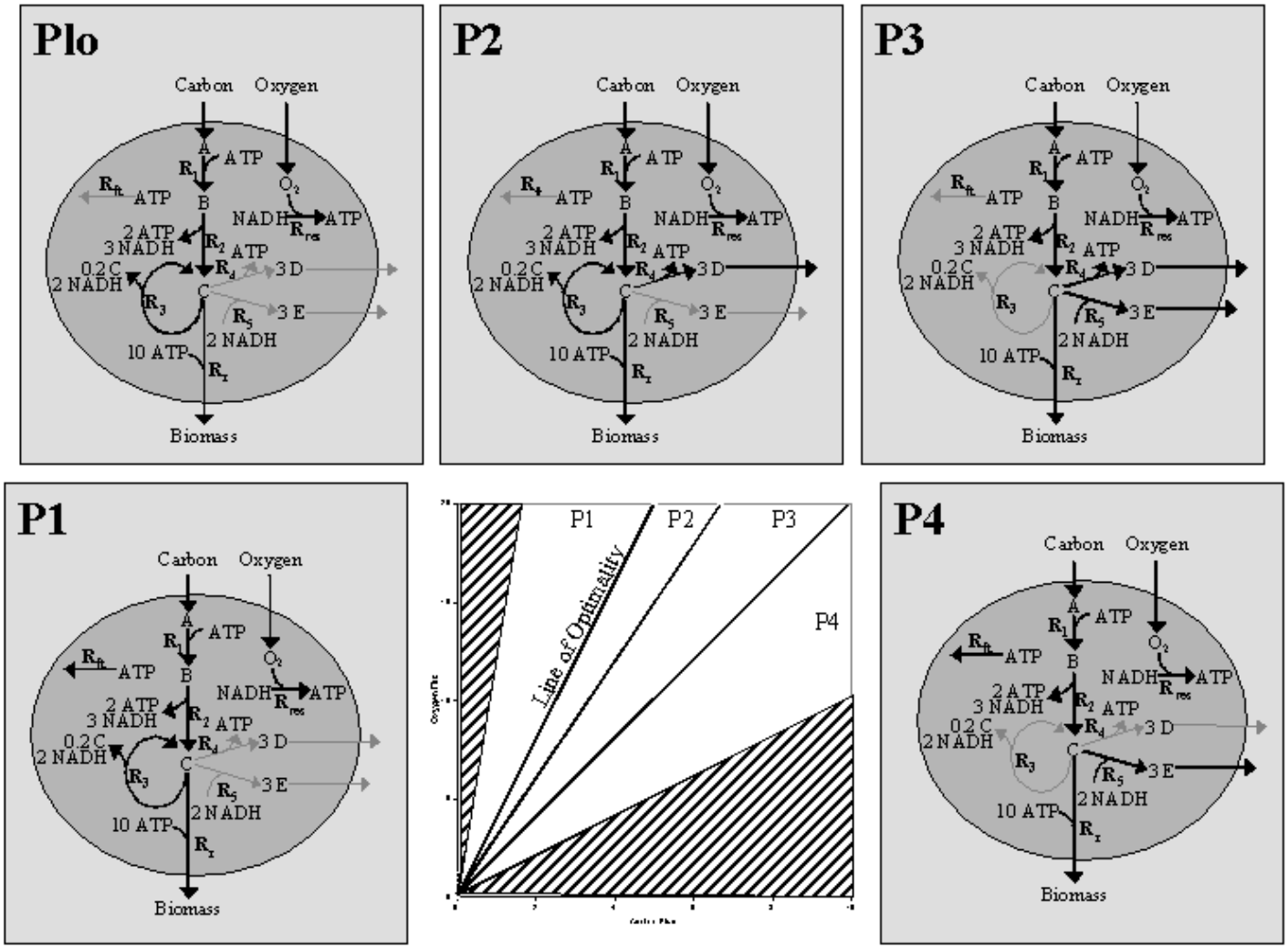


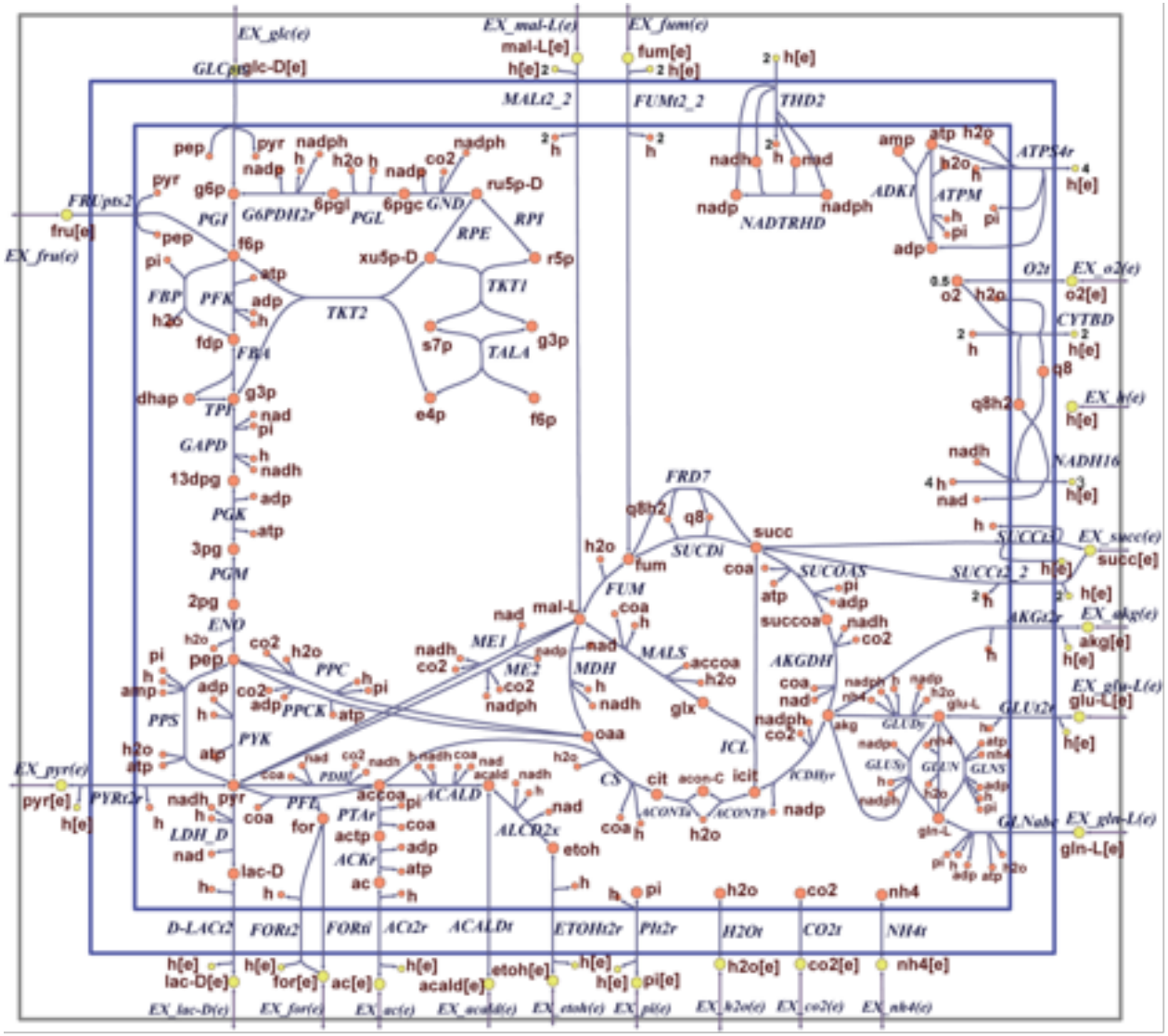
Figure 5: Phenotype phase plane example figure.

(Courtesy of Palsson et al., Nature Biotechnology, 2001)

Applications of Flux Balance Analysis

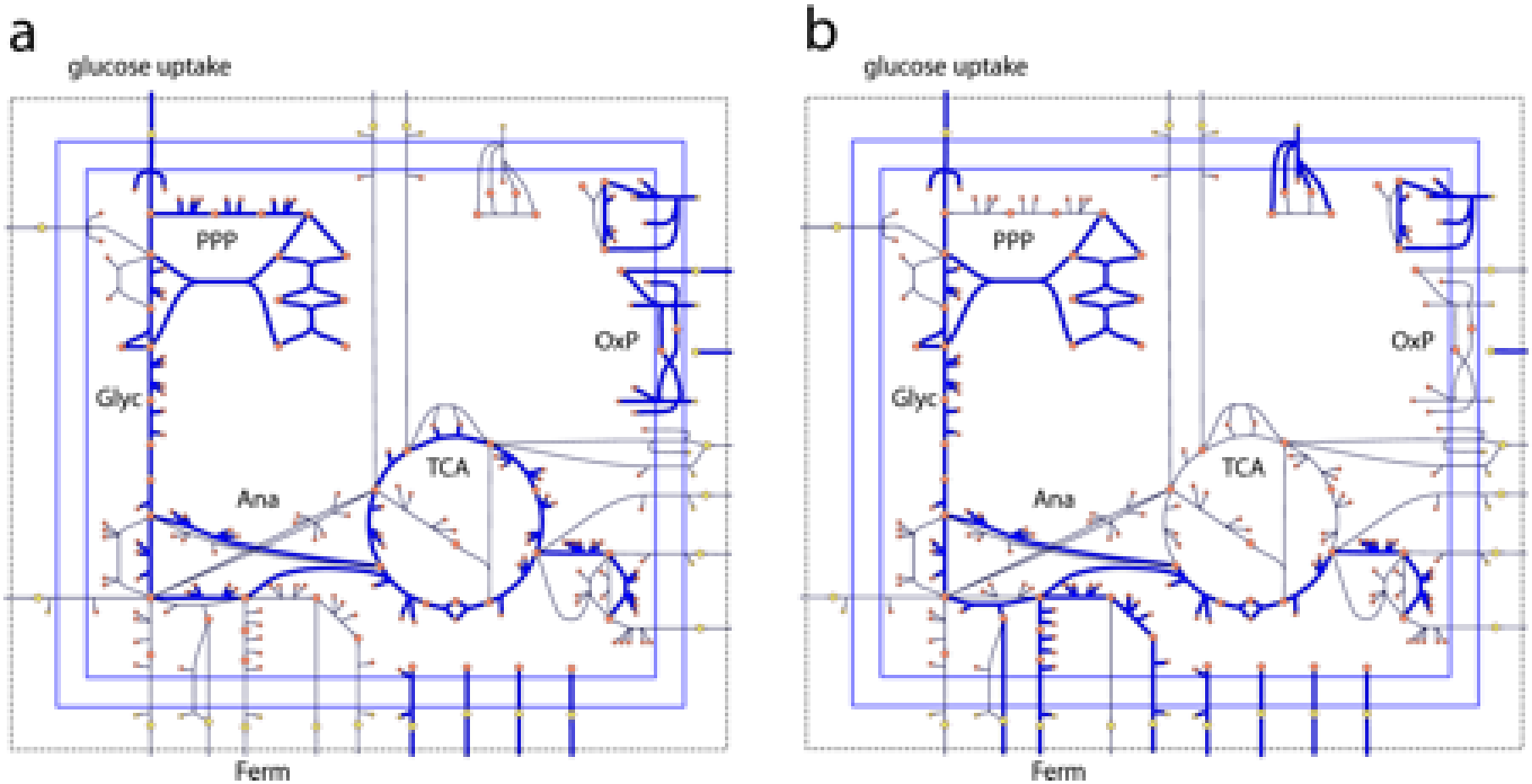
- Predict the growth or yields on different mediums.
- Study the effects of gene/reaction deletions and other perturbations.
- Fill in gaps in genome-scale metabolic reconstructions.
- Improve the efficiency of metabolic engineering.

Map of the *E. coli* Metabolic Network



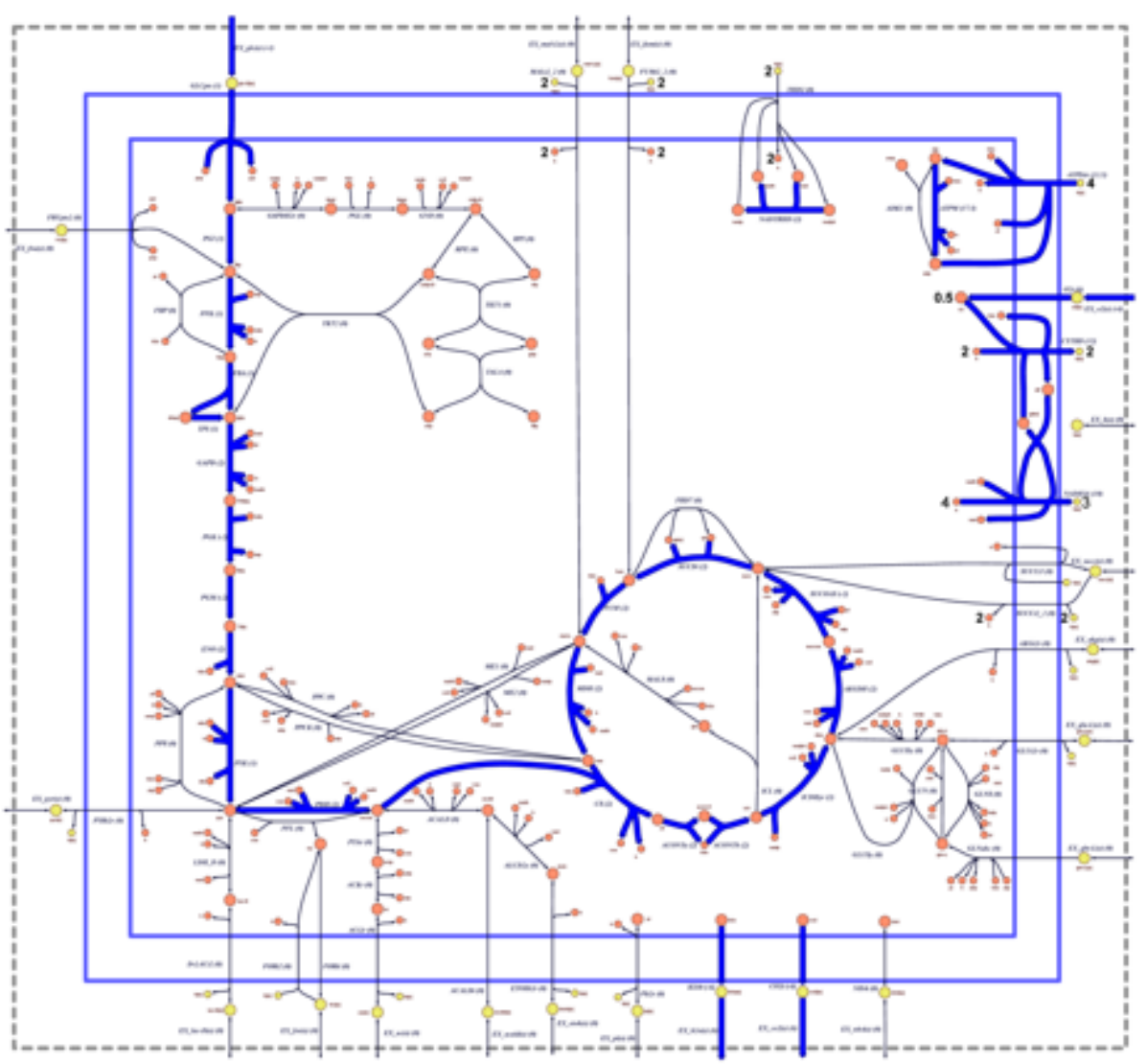
(Courtesy of Orth et al., Nature Biotechnology, 2010)

Flux Map for Aerobic and Anaerobic Growth



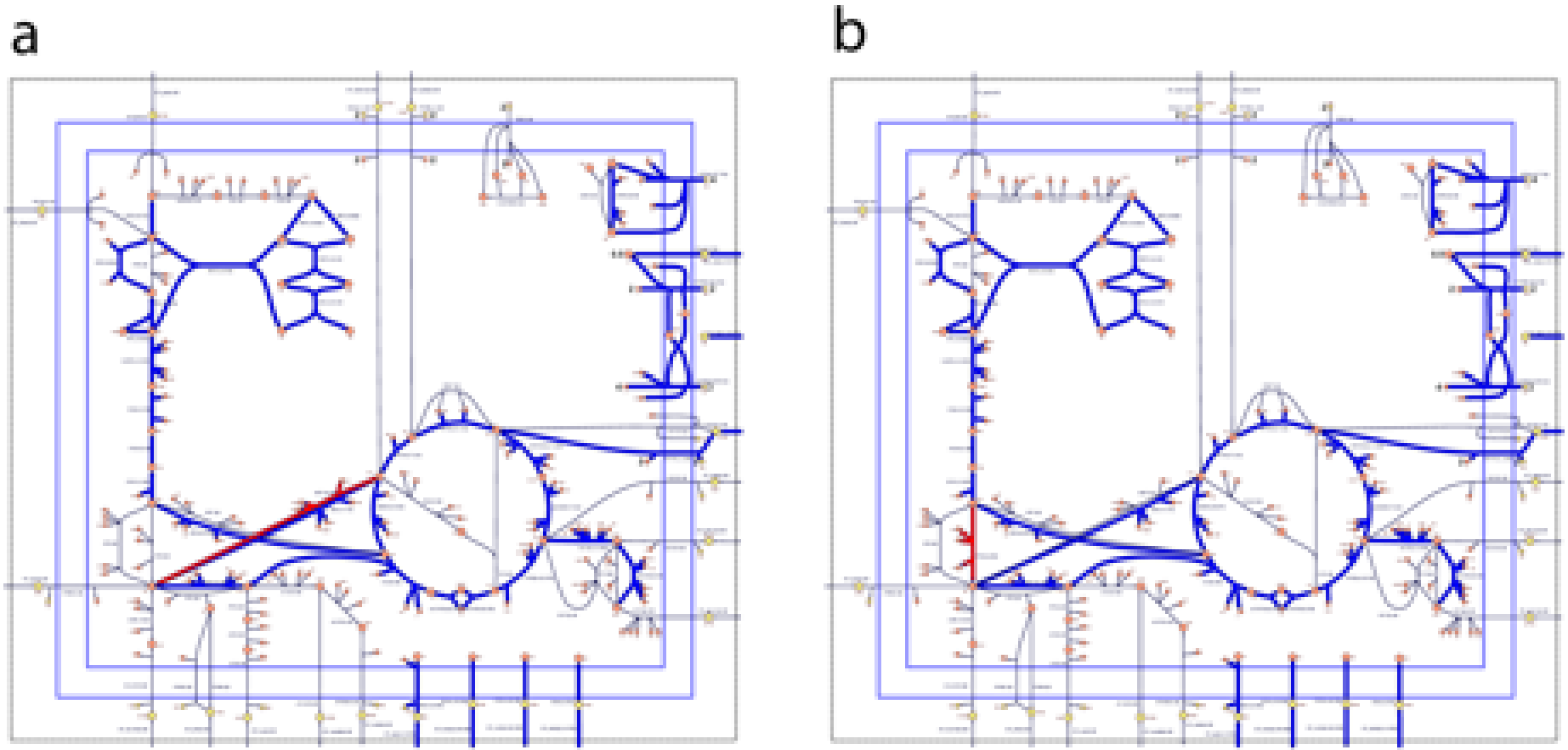
(Courtesy of Orth et al., Nature Biotechnology, 2010)

Flux Map for Maximum ATP Yield in Aerobic Conditions



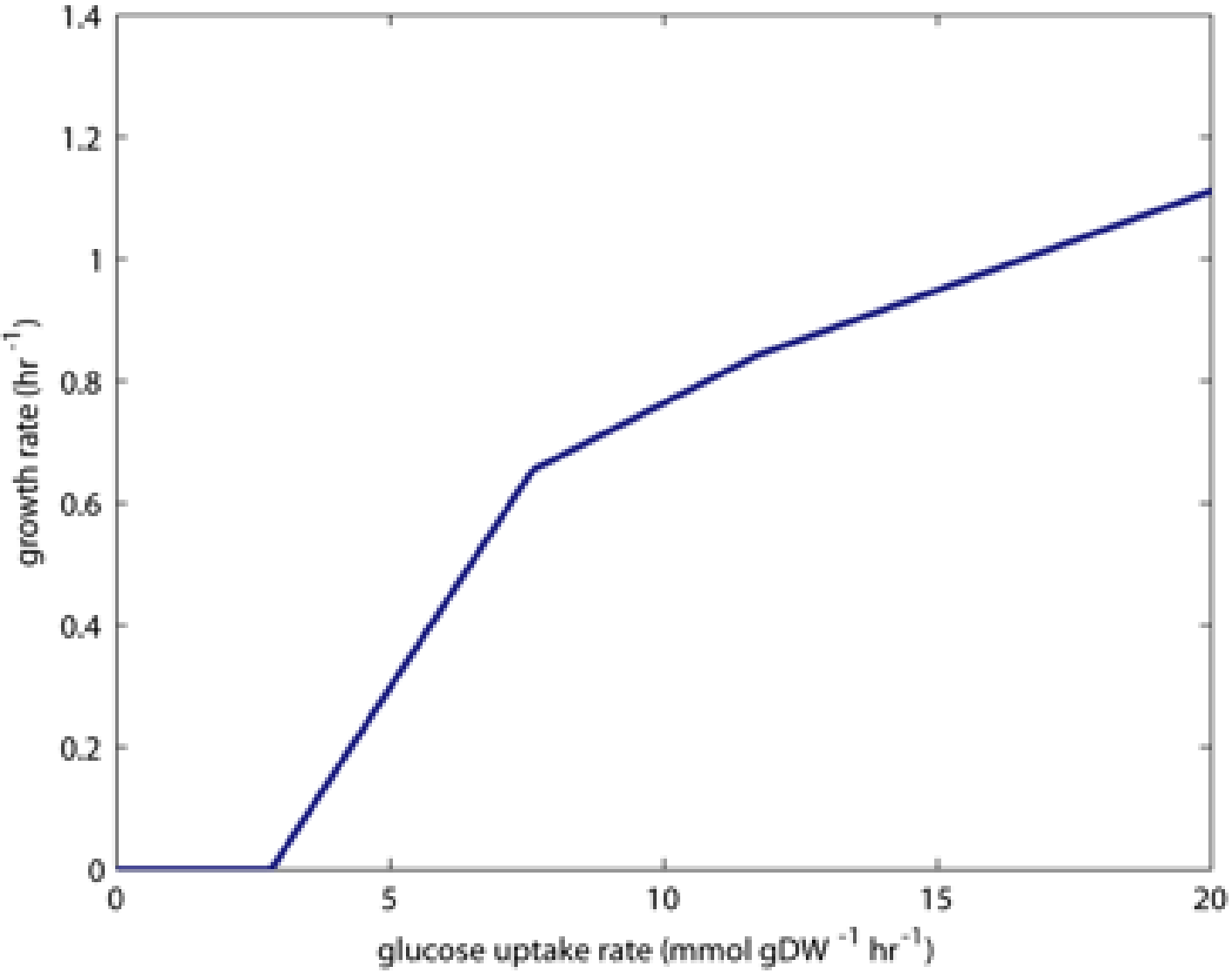
(Courtesy of Orth et al., Nature Biotechnology, 2010)

Flux Maps for Aerobic Growth on Succinate



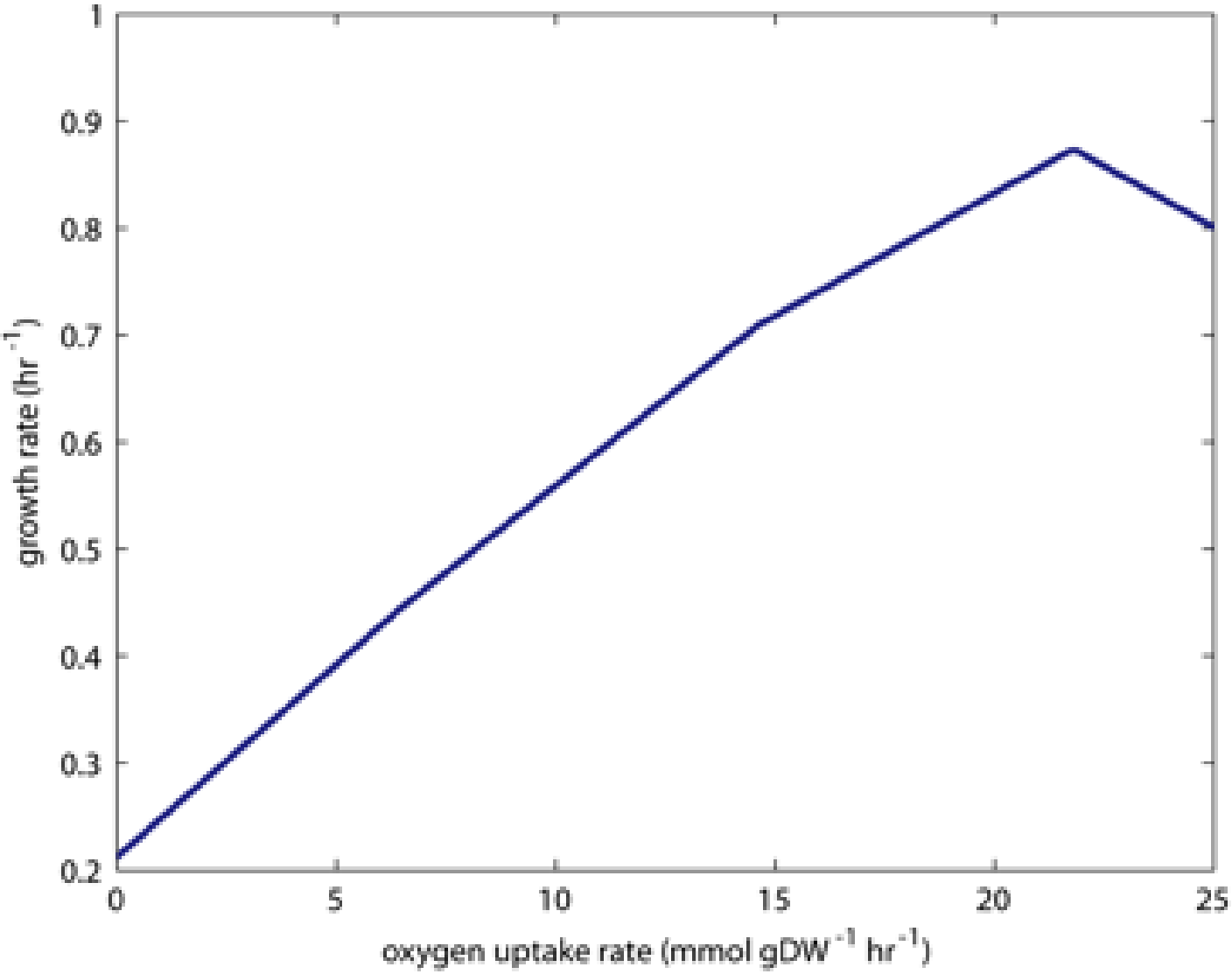
(Courtesy of Orth et al., Nature Biotechnology, 2010)

Maximum Growth Rate Versus Glucose Uptake



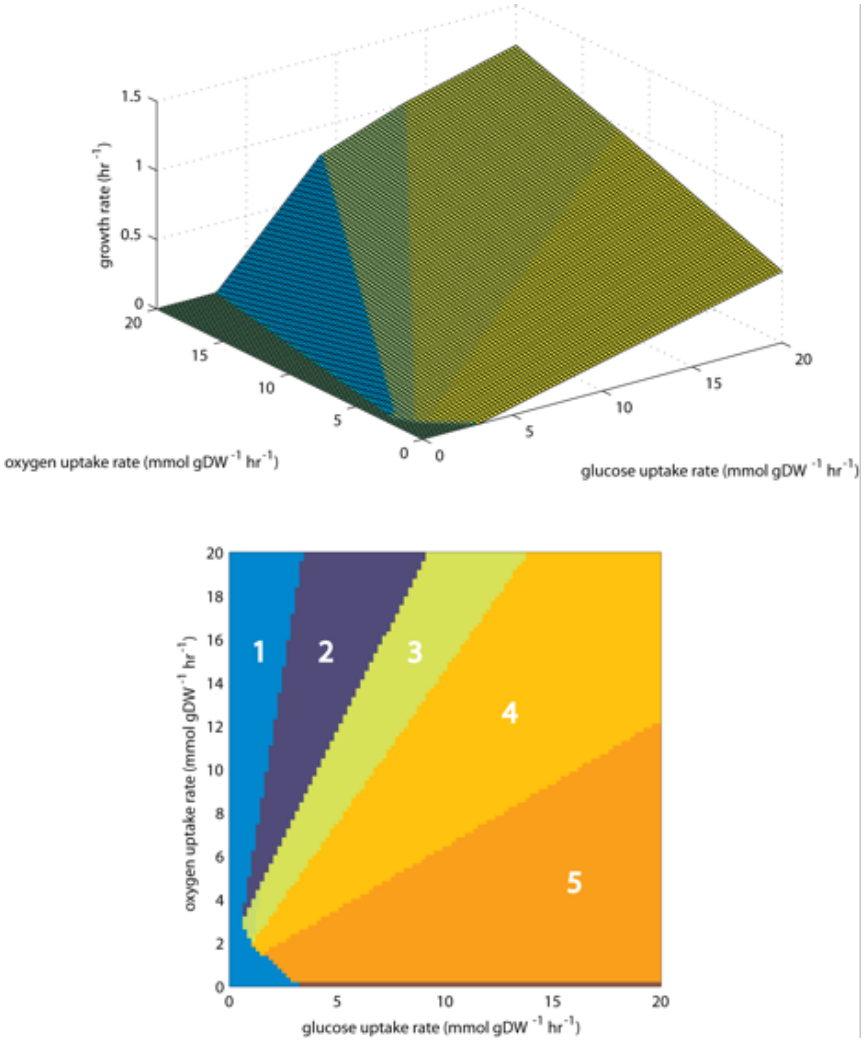
(Courtesy of Orth et al., Nature Biotechnology, 2010)

Maximum Growth Rate Versus Oxygen Uptake



(Courtesy of Orth et al., Nature Biotechnology, 2010)

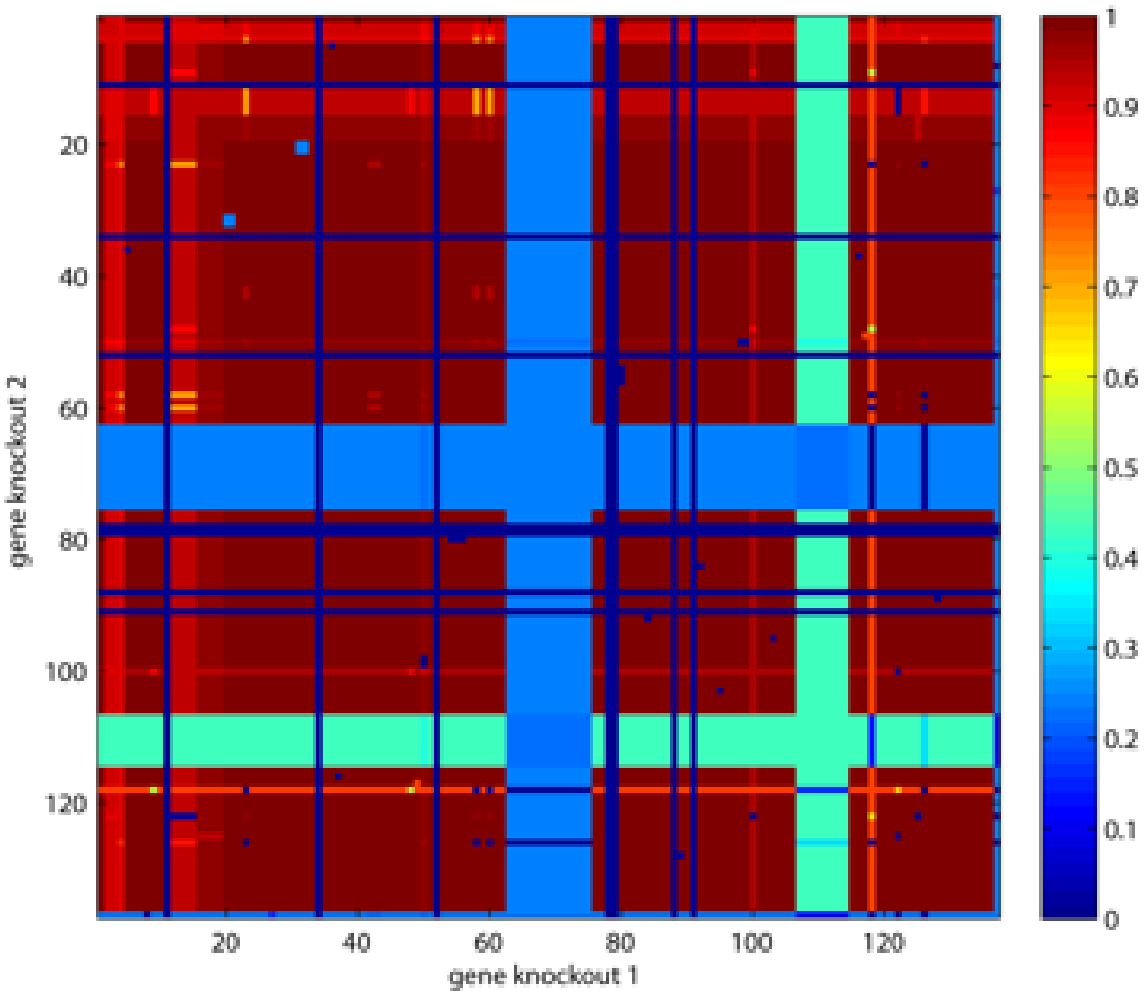
Phenotypic Phase Planes for Growth



(Courtesy of Orth et al., NBT, 2010)

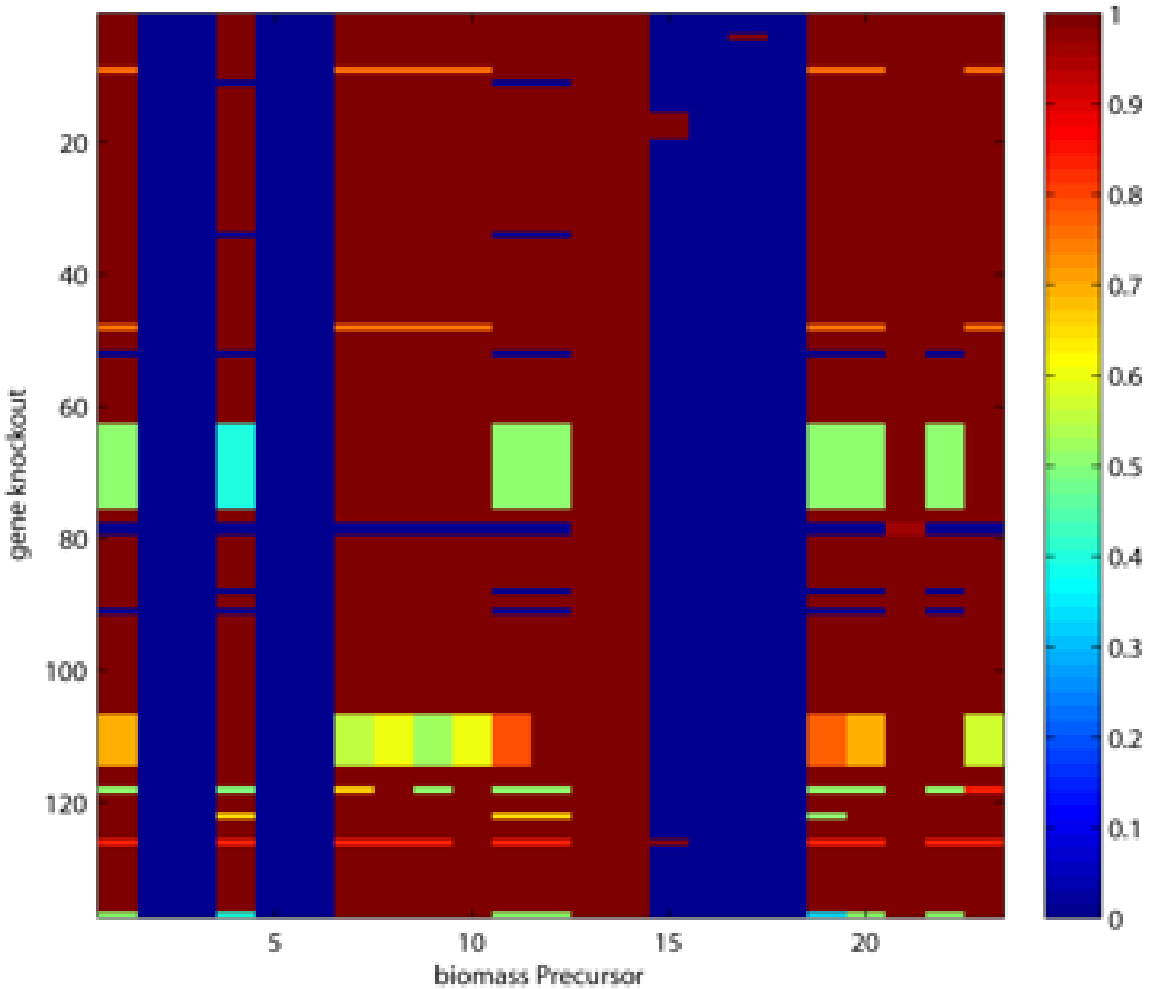
1. No growth
2. Growth limited by excess oxygen
3. Acetate is secreted
4. Acetate/formate secreted
5. Acetate/formate/ethanol secreted

Effects of Gene Knockouts on Growth



(Courtesy of Orth et al., Nature Biotechnology, 2010)

Effects of Gene Knockouts on Biomass Precursors



(Courtesy of Orth et al., Nature Biotechnology, 2010)

Advanced Flux Balance Analysis Methods

- *Flux variability analysis* - returns boundaries of reaction fluxes.
- *Minimization of Metabolic Adjustment* (MOMA) - determines flux distribution immediately following a perturbation.
- *Regulatory On-Off Minimization (ROOM)* - minimizes regulatory changes needed to adapt after a perturbation.
- *Dynamic FBA* - runs FBA, changes model using dynamic simulation, and reruns FBA.

FBA Discussion

- Advantages:
  - Does not require any kinetic parameters.
  - Only requires network and its stoichiometric matrix.
  - Results have been shown to agree with experimental data.
- Disadvantages:
  - Difficult to know true objective function for optimal fluxes.
  - Does not describe dynamic behavior.
  - Results do not always agree with experimental data.

Sources

- Early works by Papoutsakis (1984), Watson (1984), Fell/Small (1986).
- FBA has been championed by Palsson and his collaborators (see Orth et al., NBT, 2010 for a good overview).
- Palsson's book, Systems Biology: Properties of Reconstructed Networks (2006), is another good reference.
- Example used in this lecture can be found at:

  `http://www.nature.com/nbt/web_extras/supp_info/nbt0201_125/info_frame.html`