Modeling Metabolism

What if we don't have a complete kinetic description?

jsedwards@salud.unm.edu

Complete kinetics?

Approach to studying behavior of defined genotypes …

Subject them to governing constraints

and

then analyze biological properties within the applicable constraints

Metabolic Constraints

- Physicochemical factors
  - Mass, energy, and redox balance
    - Systemic stoichiometry
  - osmotic pressure, electroneutrality, solvent capacity, molecular diffusion, thermodynamics
  - Non-adjustable constraints
- System specific factors
  - Capacity:
    - Maximum fluxes
  - Rates:
    - Enzyme kinetics
  - Gene Regulation
  - Adjustable constraints

What are the metabolic capabilities

- Important question.
- Genome sequencing projects were hoped to answer this question.
  - Becoming clear that cellular functions are multigenic in nature.
  - Capabilities can not be assessed by cataloging of genes.
  - Systems science must be applied to study the systemic behavior of the entire genotype.
  - Flux-balance analysis (FBA) is a method well-suited to answer many questions.

Metabolism

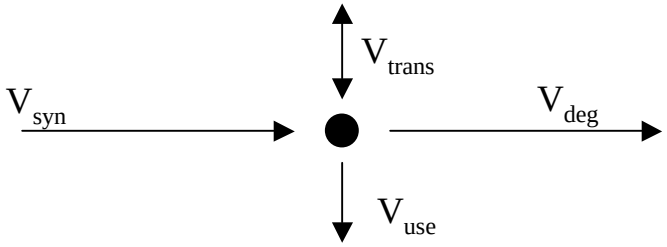
- Metabolism is the “chemical engine” that drives cellular activities.
  - Acts to convert raw materials (ie, glucose) into energy and the building blocks used to produce biological structures
  - Dynamic process
  - Obeys the laws of physics and chemistry
  - limited by the physico-chemical constraints
  - regulatory mechanisms

Description of metabolism

- Metabolic reactions (catalyzed by enzymes) are characterized by stoichiometry and the rate of conversion.
  - Stoichiometry is the most reliable information regarding metabolism.
    - Sequence of reactions
- We will discuss the mathematical description of metabolic stoichiometry

Dynamic Description

- Dynamic mass balances on each metabolite
  - Sum of rates of formation, degradation, utilization, and transport



- V_{trans}, uptake or secretion of metabolite across the cell membrane
- V_{syn}, Synthesis of the metabolite
- V_{use}, consumption of cellular constituents or maintenance requirements
- V_{deg}, degradation of metabolite

Dynamic Description

$$\frac{dX_i}{dt} = V_{syn} - V_{deg} + V_{trans} - V_{use}$$

- Typically, the uptake and secretion rates are known.
- The growth and maintenance requirements are known.

$$\frac{dX_i}{dt} = V_{syn} - V_{deg} - b$$

- More formally, one can write

$$\frac{dX_i}{dt} = S_{ij} v_j - b_i$$

- Where v_j is the jth reaction rate, S_{ij} is the moles of metabolite i produced in reaction j
- This is typically written in matrix form

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

Dynamic Description

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

- This is an important equation.
  - Gives the rate of change of the metabolite concentrations as a linear combination of the reaction rates.
  - The reaction rates are non-linear functions of the metabolite concentrations and a set of unknown parameters.

$$v_i = f(\mathbf{c};\mathbf{p})$$

  - Thus, we have a very difficult equation to solve!

Flux-Balance Analysis

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

- Make simplifications based on the properties of the system.
  - 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.
- One may thus consider the steady-state approximation to answer many questions regarding metabolism.

Flux-Balance Analysis

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$

- Removes the metabolite concentrations as a variable in the equation.
- Time is also not present in the equation.
- We are left with a simple matrix equation that contains:
  - Stoichiometry: *known*
  - Uptake rates, secretion rates, and requirements: *known*
  - Metabolic fluxes: Can be solved for!
- We will discuss the steady-state behavior now, and leave the dynamic description for later.

Stoichiometric Matrix

- The matrix, **S**, is very important in metabolic dynamics.
- It maps the reaction rates into the rates of change of metabolites.
- *mxn* matrix. The number of columns *n* (reactions) often exceeds the number of rows *m* (metabolites)
  - Will address this later

FBA

- There are 3 different situations that can occur in the stoichiometric matrix.
  - Under-determined system *(n > m)*
  - Determined system *(n = m)*
  - Over-determined system *(n < m)*

Determined System

- Most systems are under-determined, but it is sometimes possible to measure some fluxes and reduce the matrix into a square matrix as follows

$$\mathbf{v} = \begin{bmatrix} \mathbf{v_c} \\ \mathbf{v}_e \end{bmatrix}, \qquad \mathbf{S} = [\mathbf{S_c} | \mathbf{S_e}]$$

$$[\mathbf{S_c} | \mathbf{S_e}] \bullet \begin{bmatrix} \mathbf{v_c} \\ \mathbf{v}_e \end{bmatrix} = \mathbf{S_c v_c} + \mathbf{S}_e \mathbf{v}_e = \mathbf{b}$$

$$\mathbf{v_c} = \mathbf{S_c}^{-1} \left(\mathbf{b} - \mathbf{S}_e \mathbf{v}_e \right)$$

- $\mathbf{S_c}$ must be non-singular
  - Minimize the condition number of $\mathbf{S_c}$
- Fluxes in $\mathbf{v_e}$ must be measurable
- The experimentally determined fluxes are subject to experimental errors. Therefore, the condition number of $\mathbf{S_c}$ in important. The condition number is a measure of the possible error propagation in computing the flux distributions

You derive a system of linear equations using steady state mass balances for a metabolic network. You determine that you need to make a few measurements so that you can calculate all the fluxes in the system. With some work, you are able to derive the following system of equations:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$

$$\begin{bmatrix} 4.5 & 3.1 \\ 1.6 & 1.1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

You measure **b** and determine it is:

$$\mathbf{b} = \begin{bmatrix} 19.25 \\ 6.84 \end{bmatrix}$$

Using these numbers, you get the following result for the fluxes in the system:

$$\mathbf{v} = \begin{bmatrix} 2.9 \\ 2.0 \end{bmatrix}$$

Now, you decide to repeat the experiment and you make the following measurement for **b**:

$$\mathbf{b} = \begin{bmatrix} 19.24 \\ 6.85 \end{bmatrix}$$

This time you determine the fluxes in the system are:

$$\mathbf{v} = \begin{bmatrix} 7.1 \\ -4.1 \end{bmatrix}$$

Over-Determined System

When the system of flux-balance equations is over-determined, a least-squares analysis in various forms is used to determine the best steady state flux distribution. Such regression finds the best fit of the data to the flux balances, and therefore represents the best reconciliation and consistency in the data.

$$\left[\mathbf{S_c}|\mathbf{S_e}\right]\bullet\left[\frac{\mathbf{v_c}}{\mathbf{v}_e}\right]=\mathbf{S_c}\mathbf{v_c}+\mathbf{S}_e\mathbf{v}_e=\mathbf{b}$$

$$\mathbf{S}^{\mathbf{T}}\cdot\mathbf{S_c}\mathbf{v_c}=\mathbf{S}^{\mathbf{T}}\cdot\left(\mathbf{b}-\mathbf{S}_e\mathbf{v}_e\right)$$

$$\mathbf{v_c}=\left(\mathbf{S_C}^{\mathbf{T}}\cdot\mathbf{S_c}\right)^{-1}\left\{\mathbf{S_C}^{\mathbf{T}}\cdot\left(\mathbf{b}-\mathbf{S}_e\mathbf{v}_e\right)\right\}$$

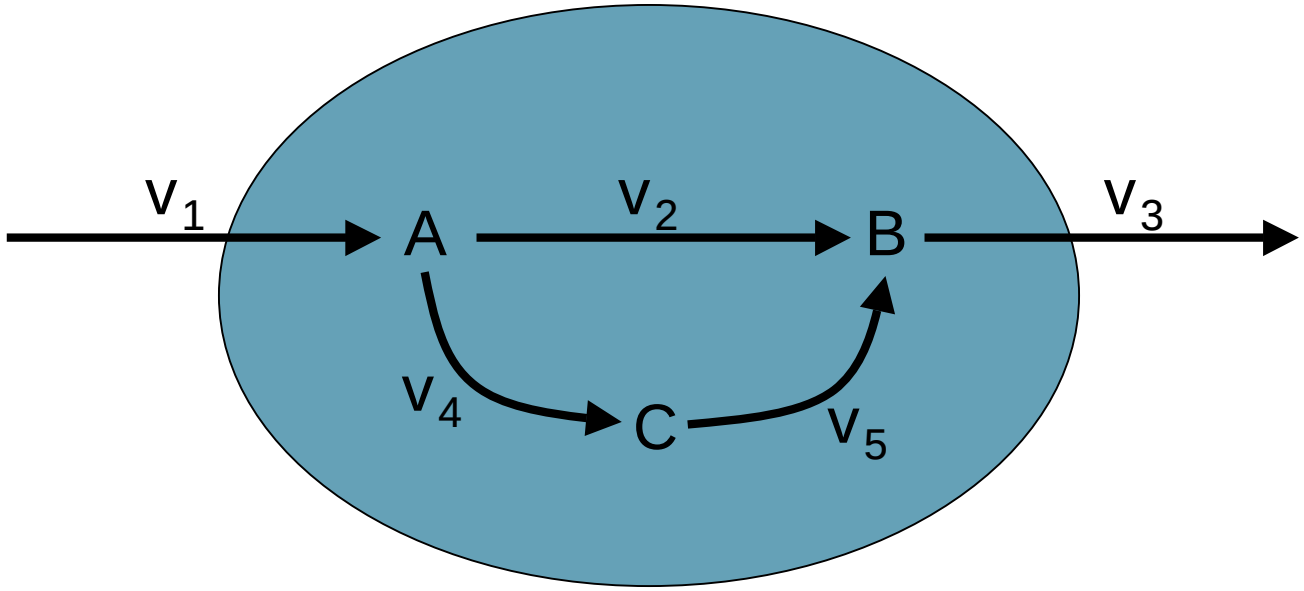
Conditions similar to the determined system are required

Example.

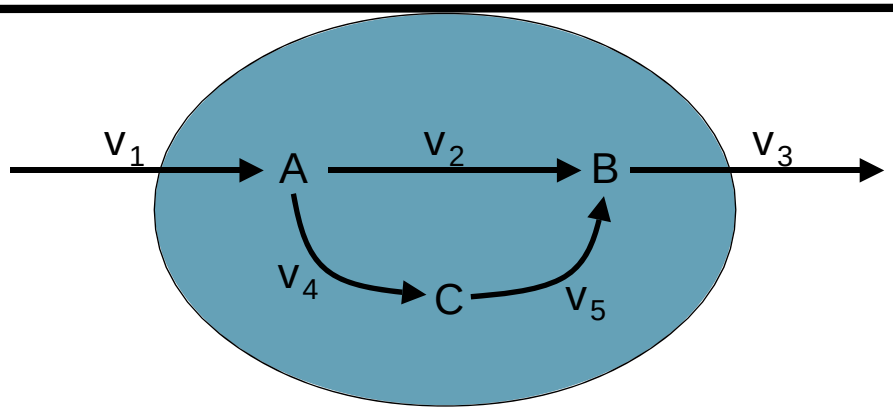
Under-Determined System

- All real metabolic systems fall into this category
- Systems are moved into the other categories by measurement of fluxes and additional assumptions.
- Infinite feasible flux distributions, however, they fall into a solution space defined by the convex polyhedral cone.
- The actual flux distribution is determined by the cells regulatory mechanisms.
- It absence of kinetic information, we can estimate the metabolic flux distribution by postulating objective functions that underlie the cell's behavior.
- Within this framework, one can address questions related to the capabilities of metabolic networks to perform functions while constrained by stoichiometry, limited thermodynamic information (reversibility), and physico-chemical constraints (ie. uptake rates)

Under-Determined System

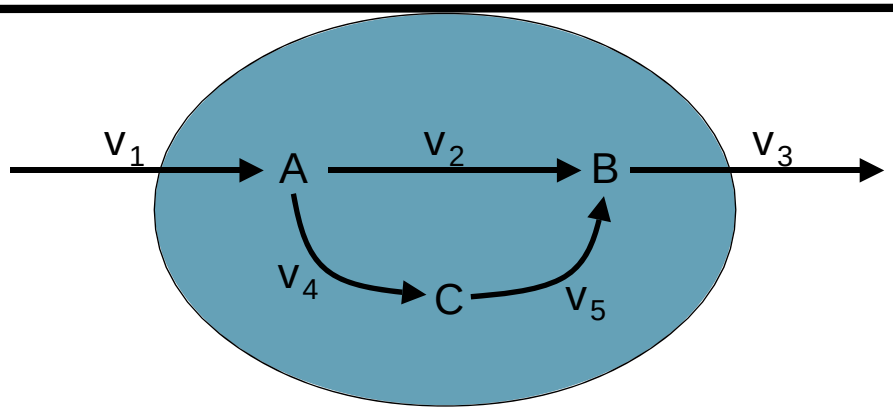


Under-Determined System



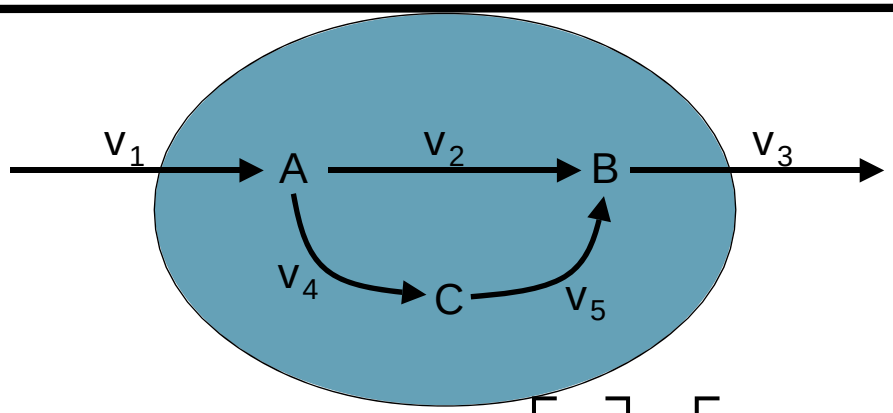
$$\frac{d}{dt}\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Under-Determined System



If $v_1 = 1$ and $v_3 = 1$ (measured), what is the relation between v_2, v_4 and v_5?

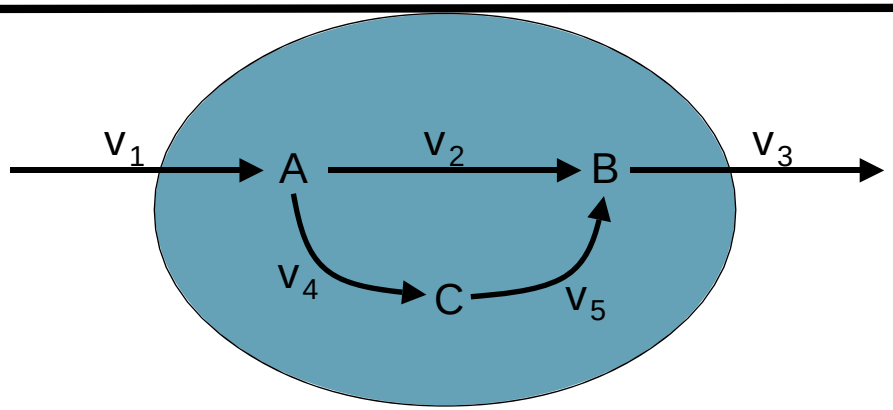
Under-Determined System



$$\frac{d}{dt}\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

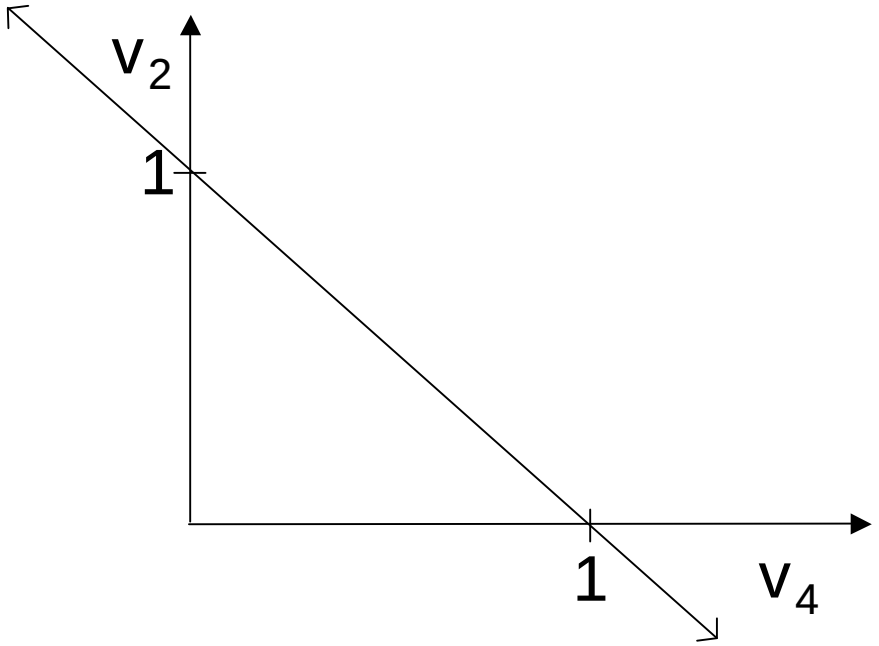
$$v_4 - v_5 = 0$$

Under-Determined System

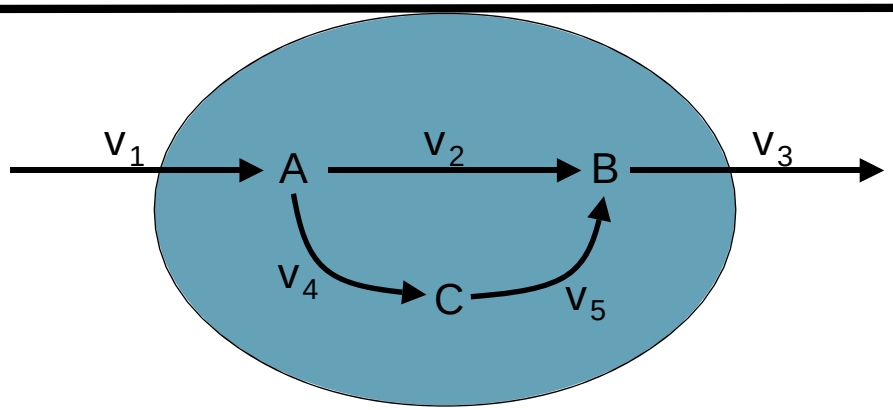


$v_4 = v_5$

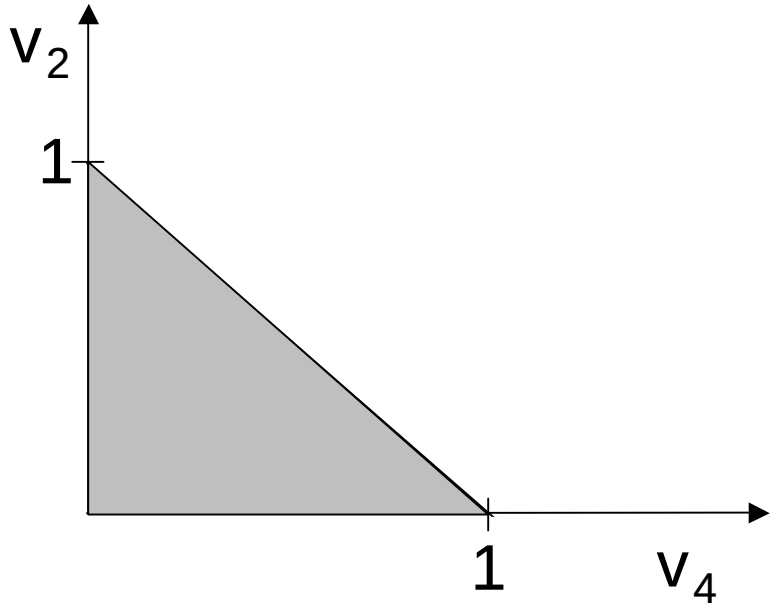
$v_1 = v_3 = 1$



Under-Determined System



$v_4 = v_5$



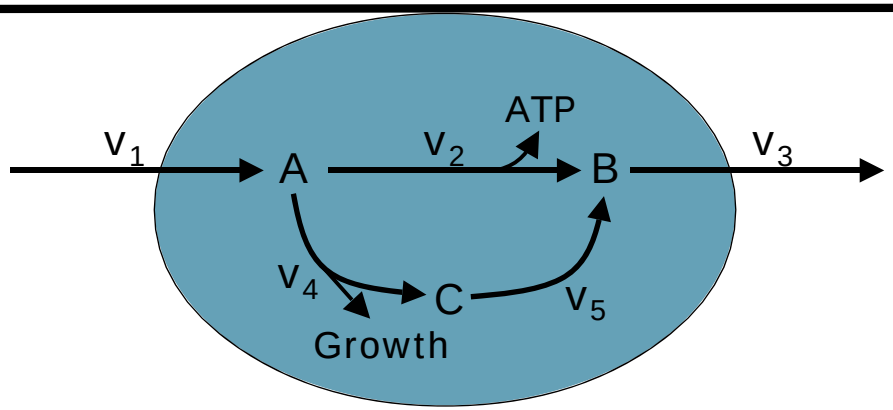
If v_4 is irreversible

If v_2 is irreversible

$v_1 \leq 1$

$v_3 \leq 1$

Under-Determined System



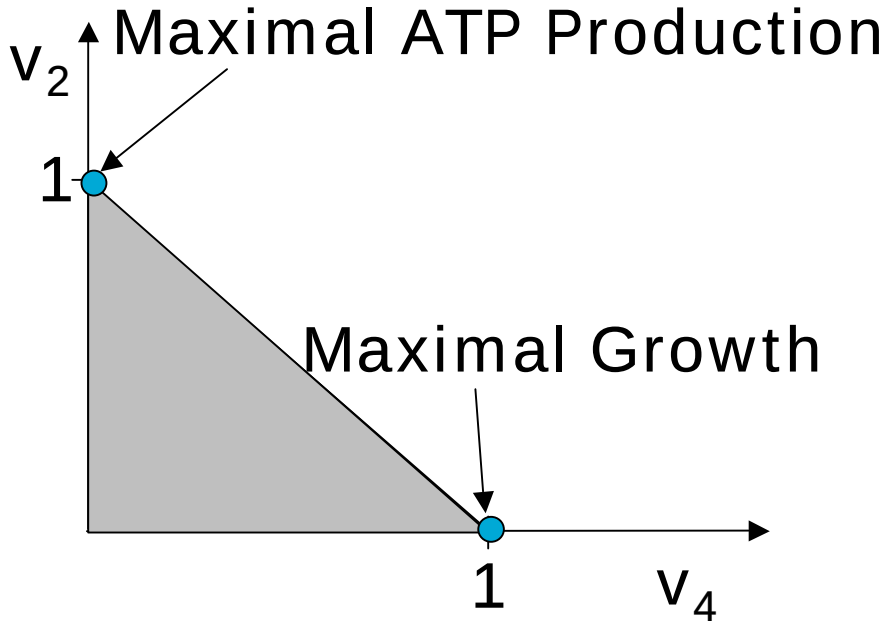
$v_4 = v_5$

If v_4 is irreversible

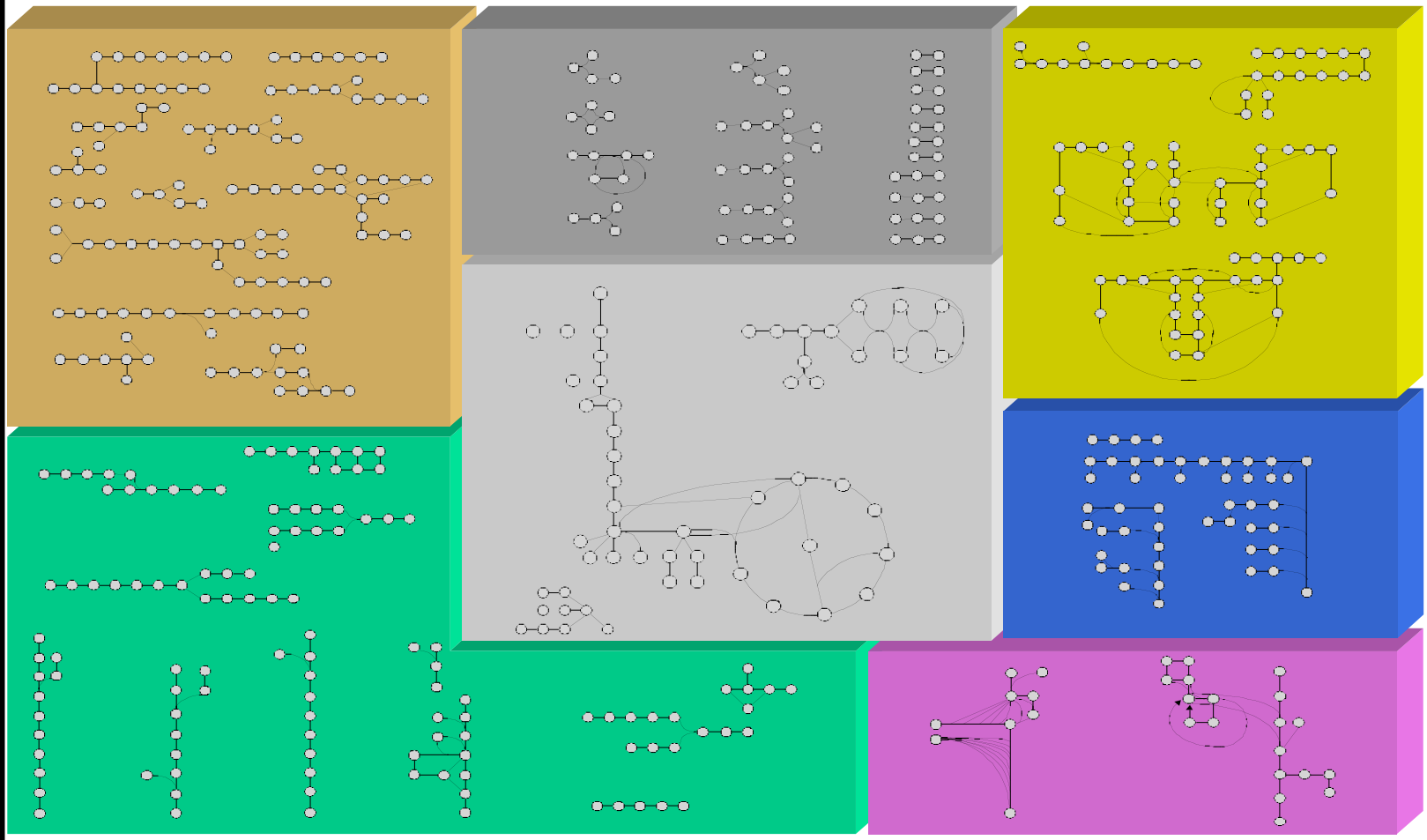
If v_2 is irreversible

$v_1 \leq 1$

$v_3 \leq 1$



Flux Balance Analysis



FBA - Linear Program

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$$

- A linear programming problem is formulated where one finds a solution to the eq. While minimizing an objective function.
  - Minimize (Z)
  - Z = **(c.v)**
- For growth, define a growth flux:

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

- Constraints to the LP problem: $\mathbf{S} \cdot \mathbf{v} = \mathbf{b}$

$$v_i \geq 0$$

$$\alpha_i \leq v_i \leq \beta_i$$

$$v_i = X_i$$

Precursors to cell growth

How to define the growth function.

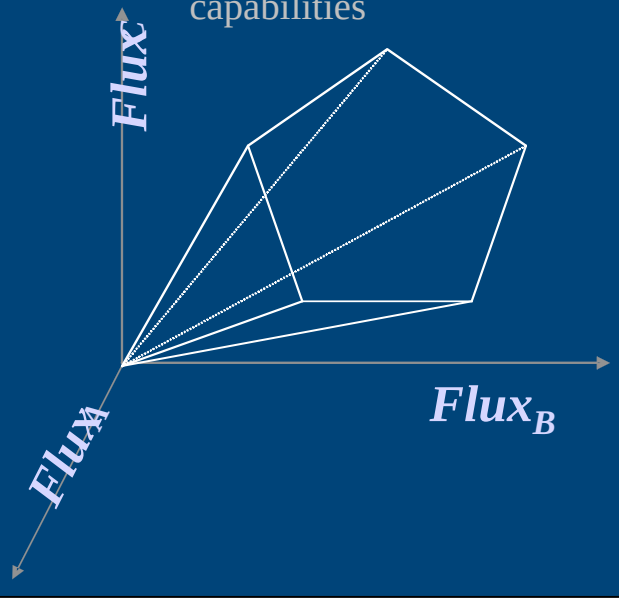
- The biomass composition has been determined for several cells, *E. coli* and *B. subtilis*.
  - This can be included in a complete metabolic network
- However, only the catabolic network can be considered that degrades the carbon source into the 12 biosynthetic precursors and generates the 3 energy and redox cofactors.

Applicability of FBA

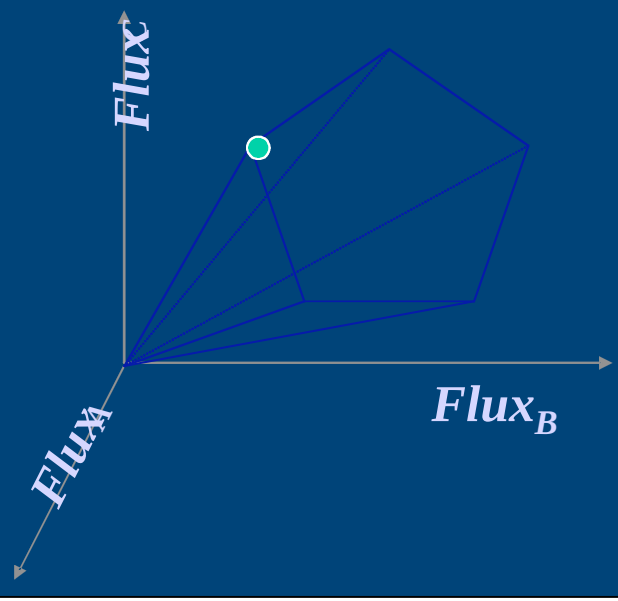
- Stoichiometry is well-known
- Limited thermodynamic information is required
  - reversibility vs. irreversibility
- Experimental knowledge can be incorporated in to the problem formulation
- Linear optimization allows the identification of the reaction pathways used to fulfil the goals of the cell if it is operating in an optimal manner.
- The relative value of the metabolites can be determined
- Flux distribution for the production of a commercial metabolite can be identified. Genetic Engineering candidates

Constraints

- Incomplete constraints
  - Physicochemical constraints
  - Feasible set is a region of flux space
    - contains flux vectors that satisfy the constraints
    - defines the metabolic capabilities

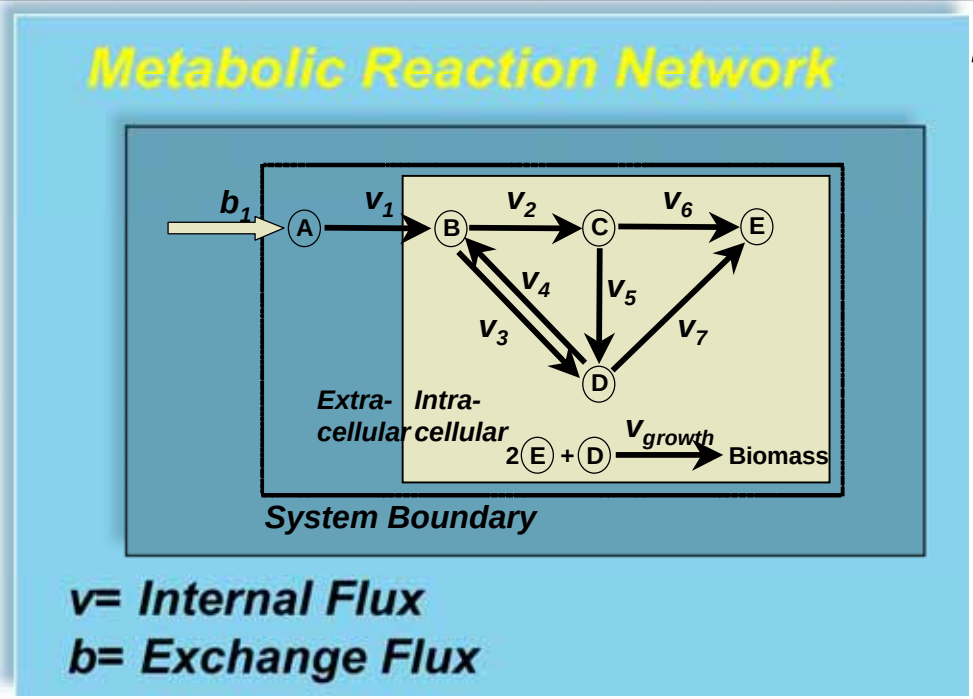


- Complete Knowledge
  - System specific constraints
    - Enzyme kinetics, gene regulation
  - Initial conditions
  - Feasible set ➍ single point



Defining the constraints

- Mass, energy, and redox balance constraints
  - Stoichiometry based
    - "hardwired"
    - well known



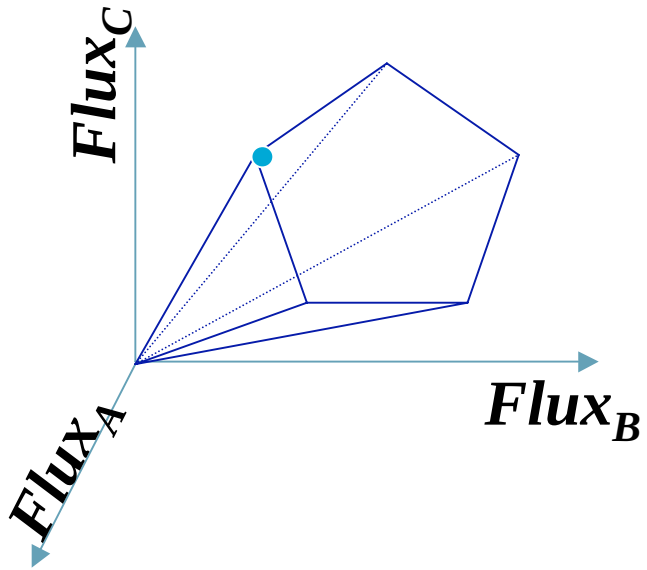
Dynamic mass balances

Steady State Conditions

S - stoichiometric matrix (*m* x *n*)
v - flux distribution vector (*n* x 1)

Defining the constraints

- Identify a specific point within the applicable constraints under any given condition
- Linear programming - Determine the optimal utilization of the metabolic network, subject to the P/C constraints, to maximize the growth of the cell



Assumption:

The cell has found the optimal solution by adjusting the system specific constraints (enzyme kinetics and gene regulation) through evolution and natural selection.

I will find the optimal solution by linear programming

Map Check

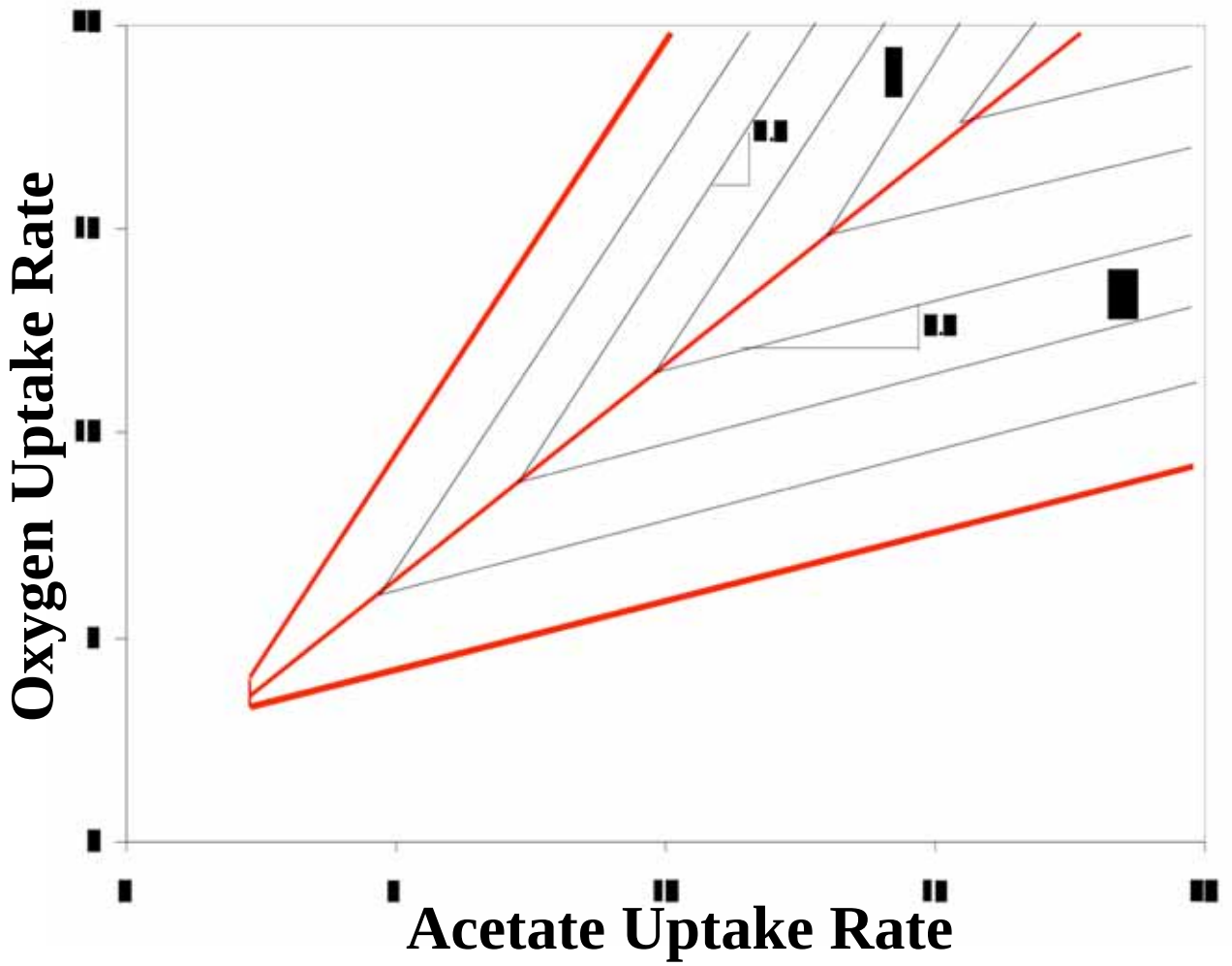
Flux balance analysis:

Quantitative Analysis of the Metabolic Flux

Acetate Carbon Source

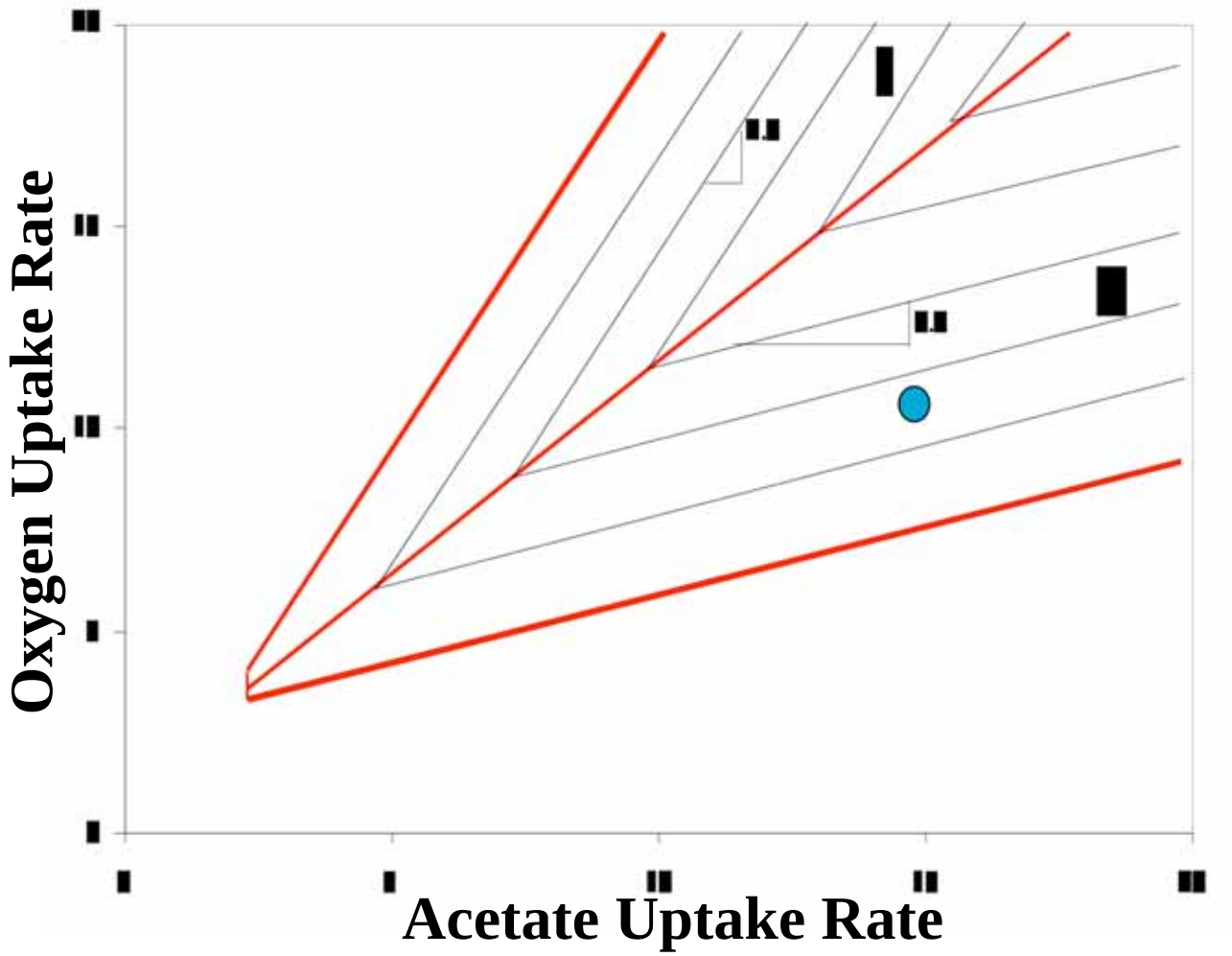
Experimental reconstruction
of the flux cone

Acetate-Oxygen PhPP

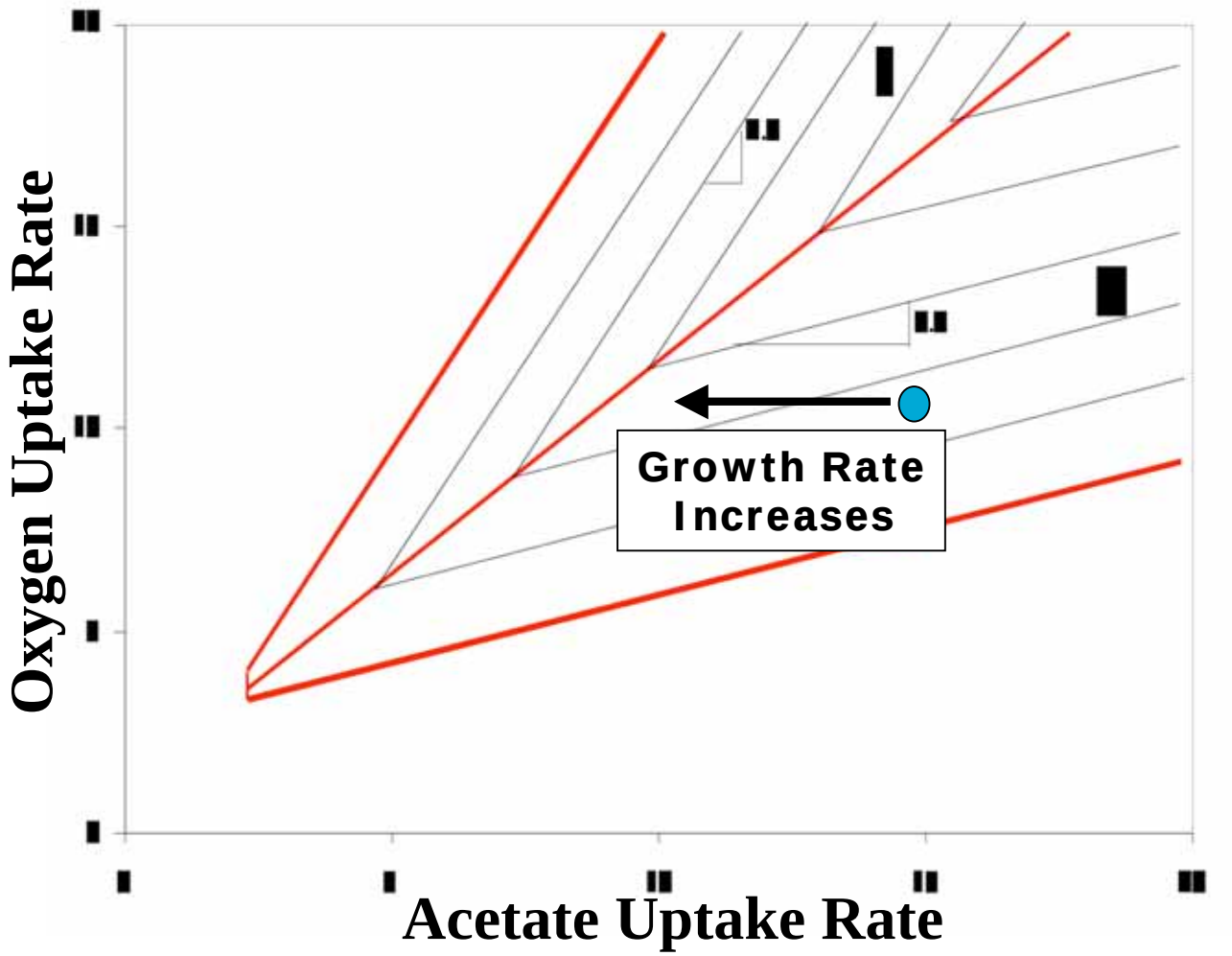


jsedwards@salud.unm.edu

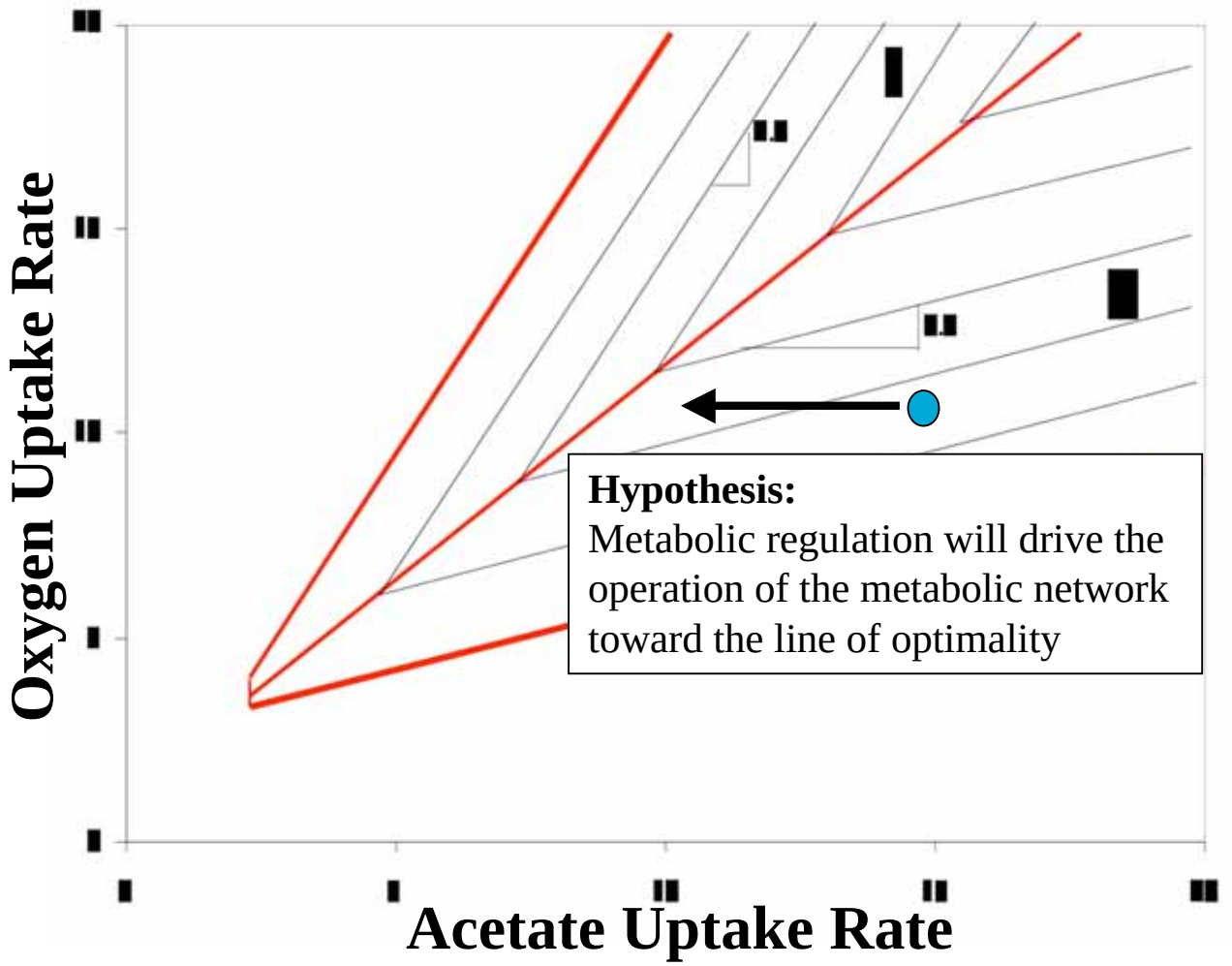
Acetate-Oxygen PhPP



Acetate-Oxygen PhPP



Acetate-Oxygen PhPP



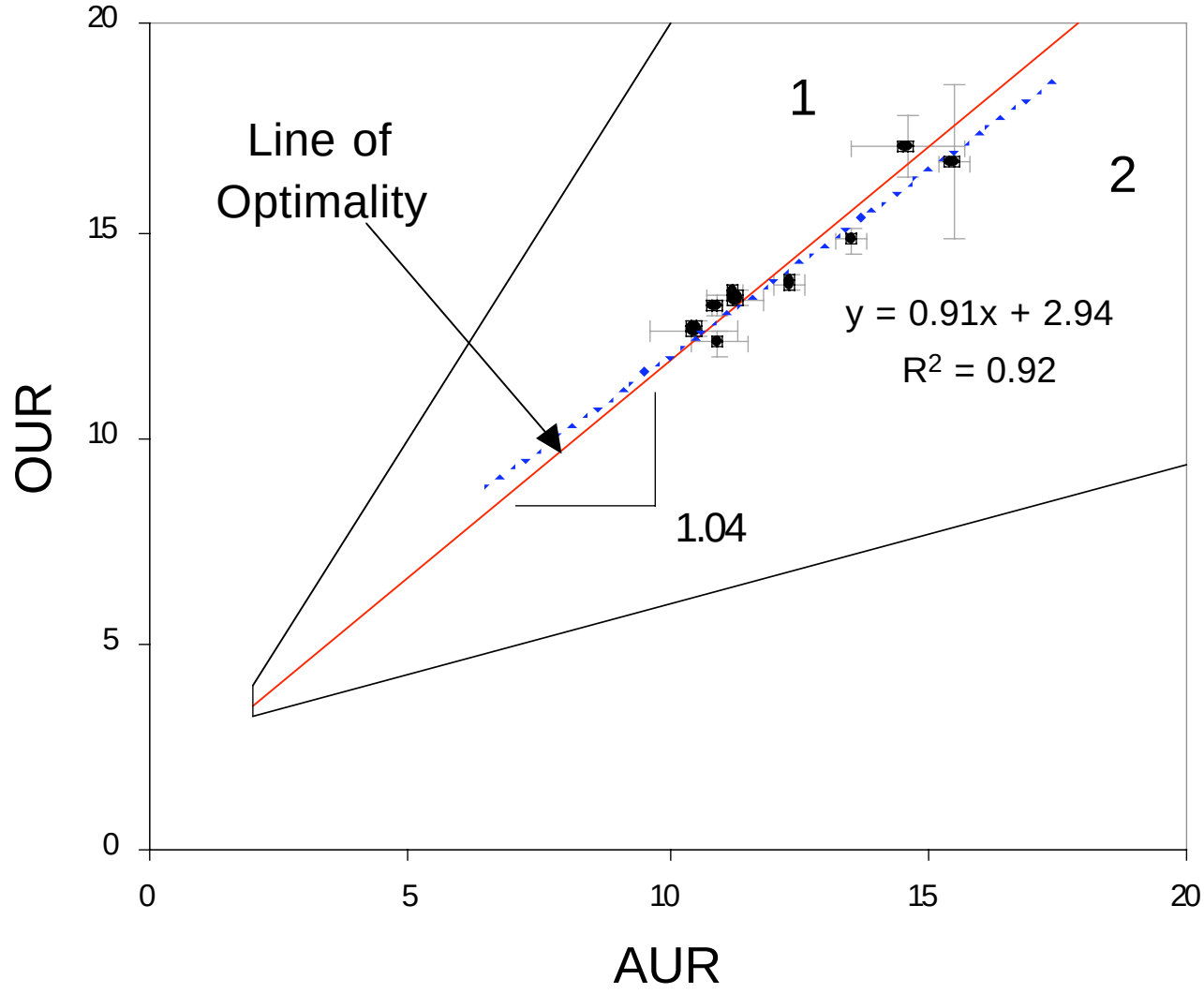
Experimental program

to test the *in silico* derived hypothesis

Methods

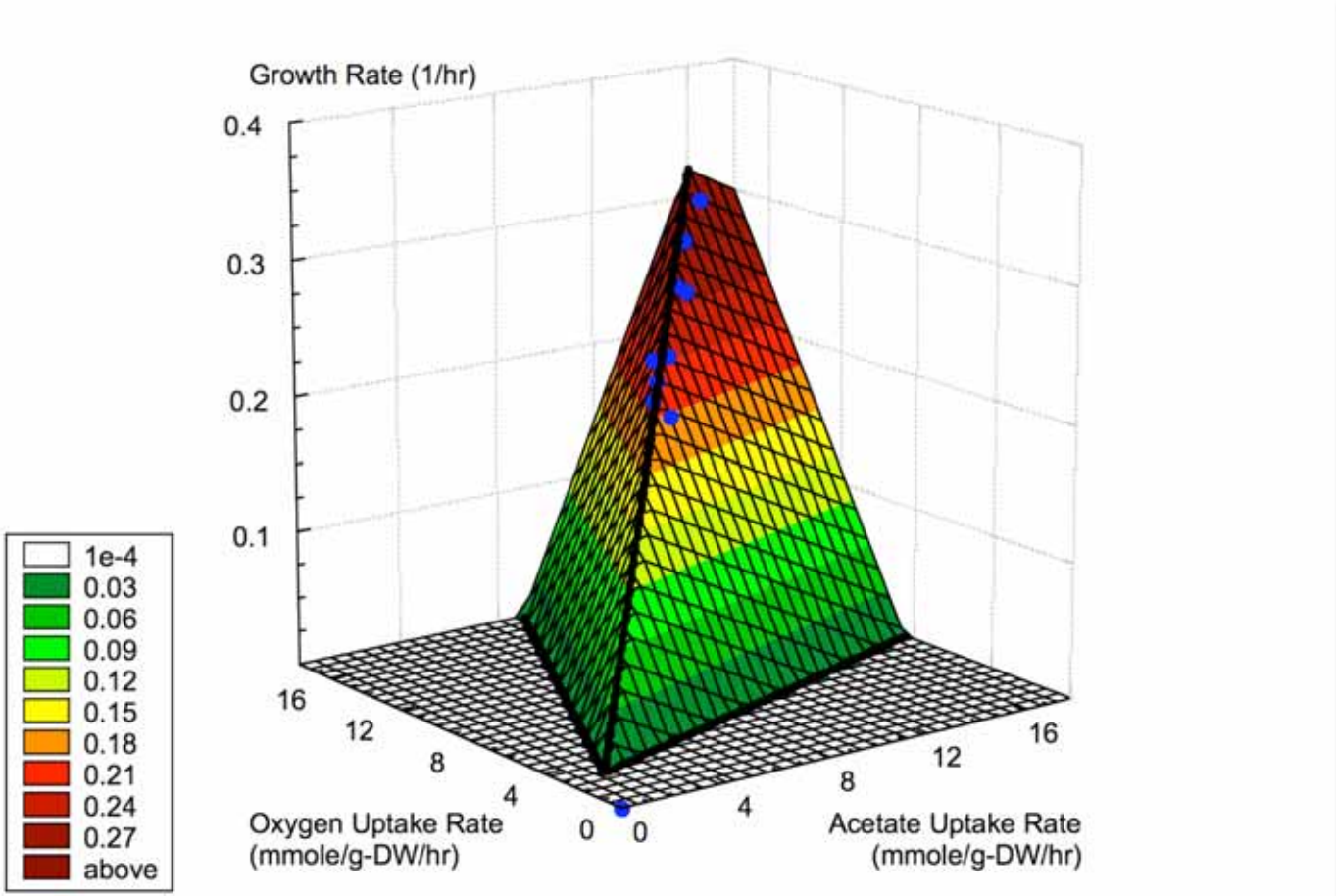
- Batch E. coli K12 on acetate M9 media at 37ºC.
- Titration of the initial acetate concentration to control the acetate uptake rate (0.3 – 4 g/L)
- Simultaneously measured the parameters to reconstruct the phenotype phase plane
  - Acetate uptake rate
    - HPLC
  - Oxygen uptake rate
    - Mass transfer measurement,Respirometer, Gas analyzer
  - Growth rate
    - Turbidity (A600 & A420) and Cell counts (Coulter Counter)
  - By-product production (Only CO2 – Not measured)

Acetate Data

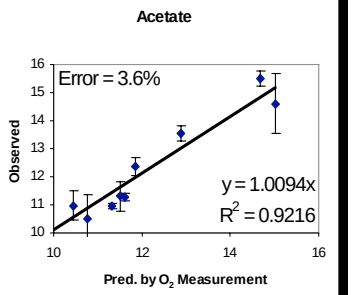
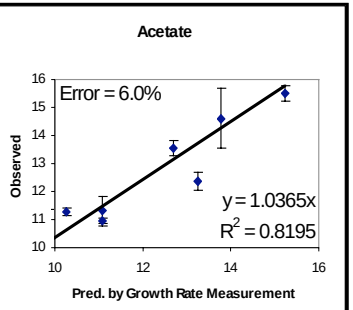
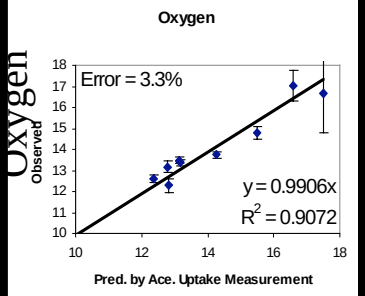
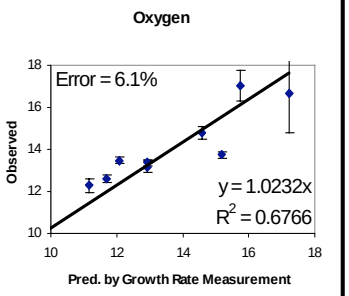
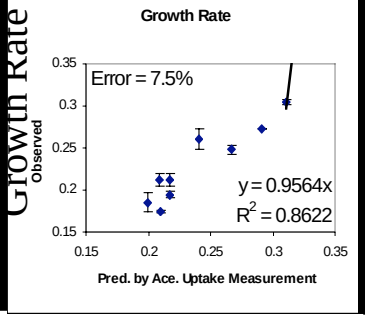
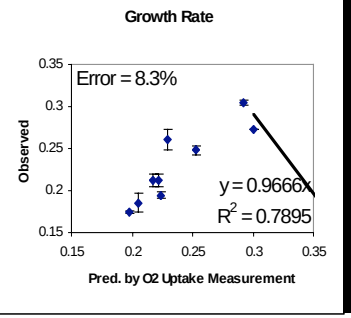


jsedwards@salud.unm.edu

Acetate 3-D PhPP



Predictive Capability

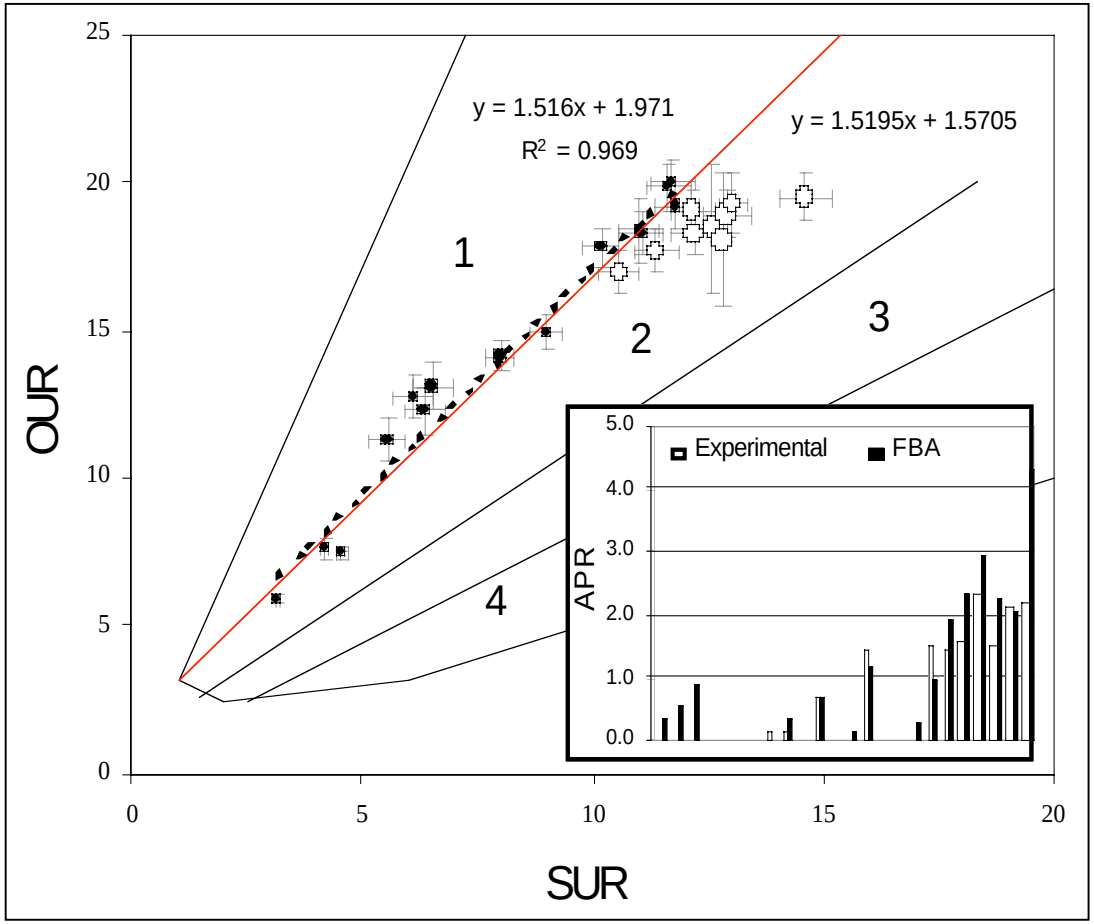
	Acetate	Oxygen	Growth Rate
Acetate	in silico analysis predicted a single degree of freedom ❹ Experimentally confirmed		
Oxygen		Measurement of one flux allows for the prediction of the remaining metabolic fluxes	
Growth Rate			AUR ❹ OUR slope diff. = 1% R^2 = 91% Avg. Error = 3.3% Overall AE = 5.8%

jsedwards@salud.unm.edu

Succinate

Experimental reconstruction
of the phenotype phase plane

Succinate PhPP



jsedwards@salud.unm.edu

Map Check

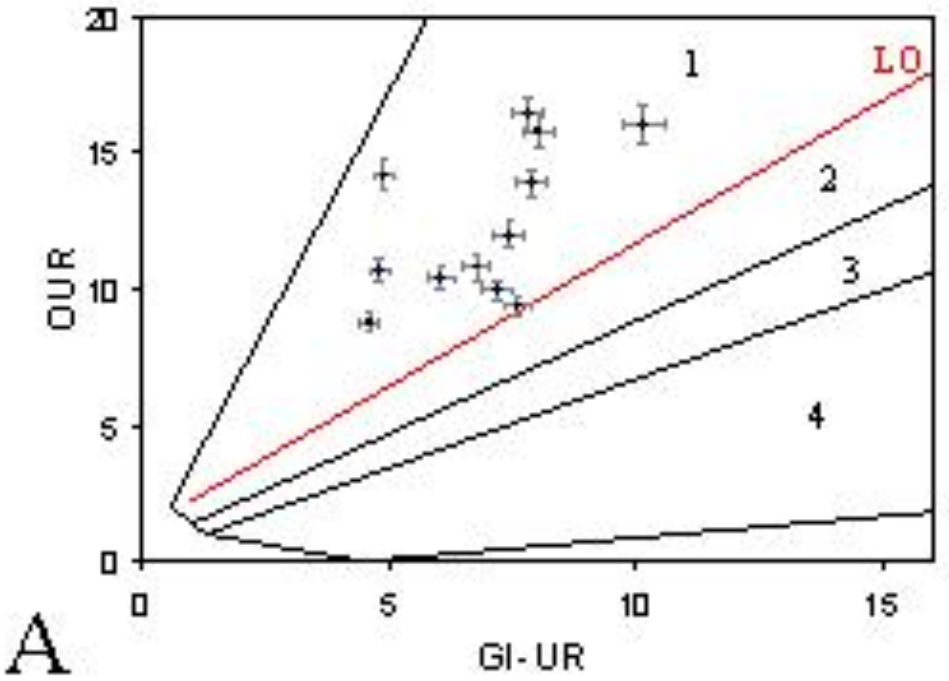
Flux balance analysis:

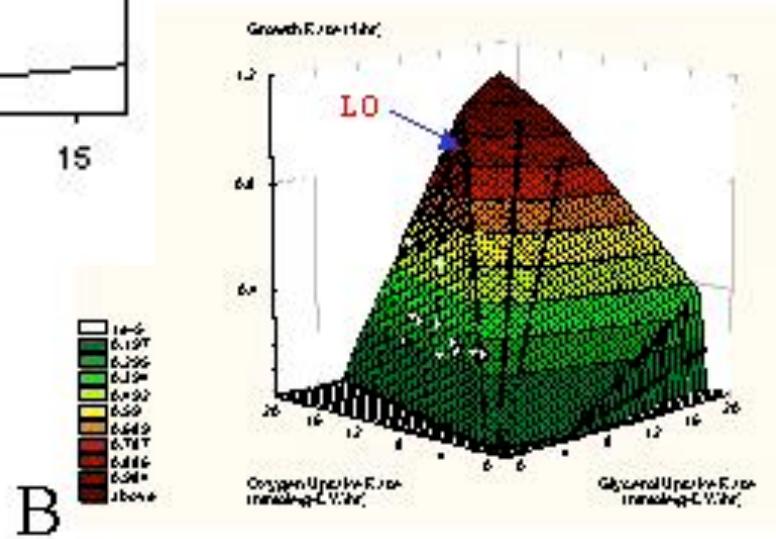
What if we are wrong?

Always valid?

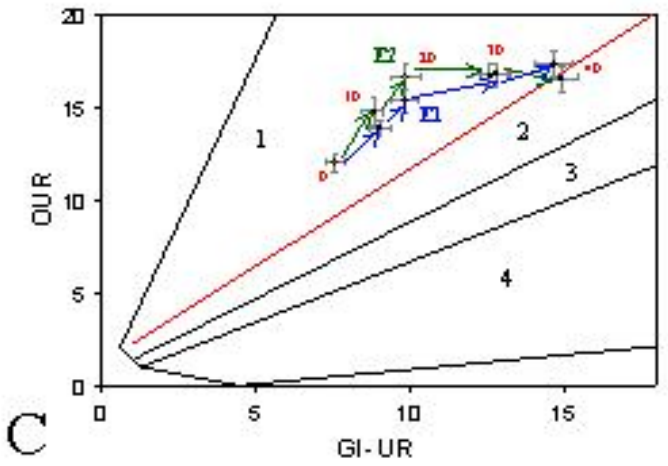
- FBA and linear optimization does not always correctly predict the behavior of *E. coli*
- Why???
- How can we test the FBA framework???

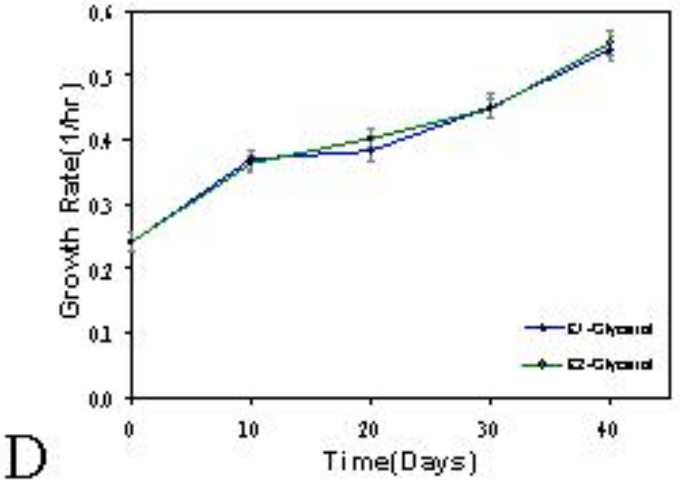
We are wrong

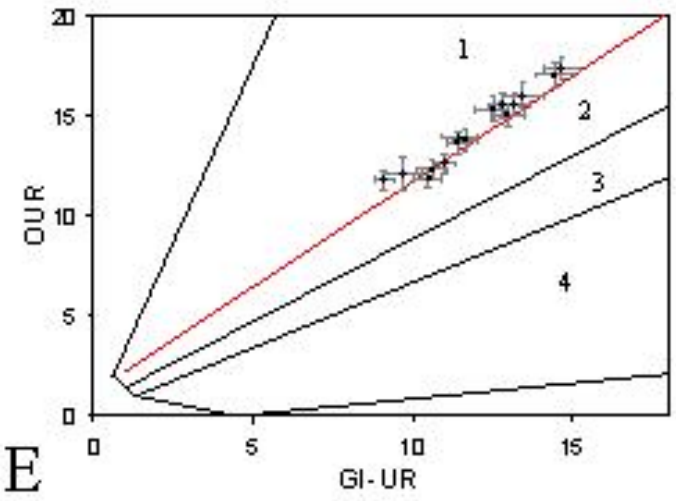


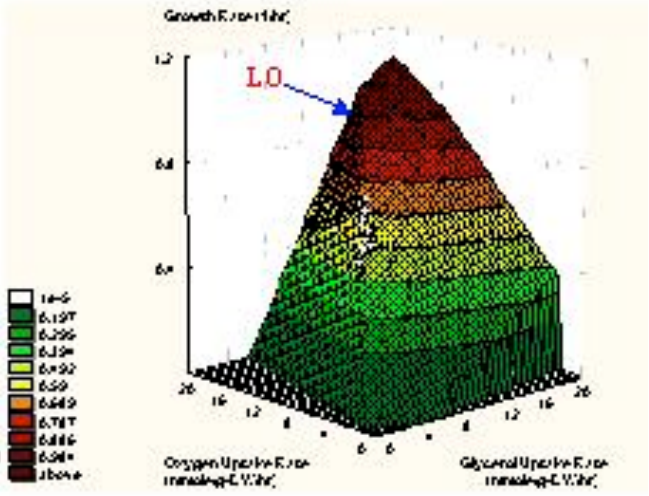


But we are also right!!!









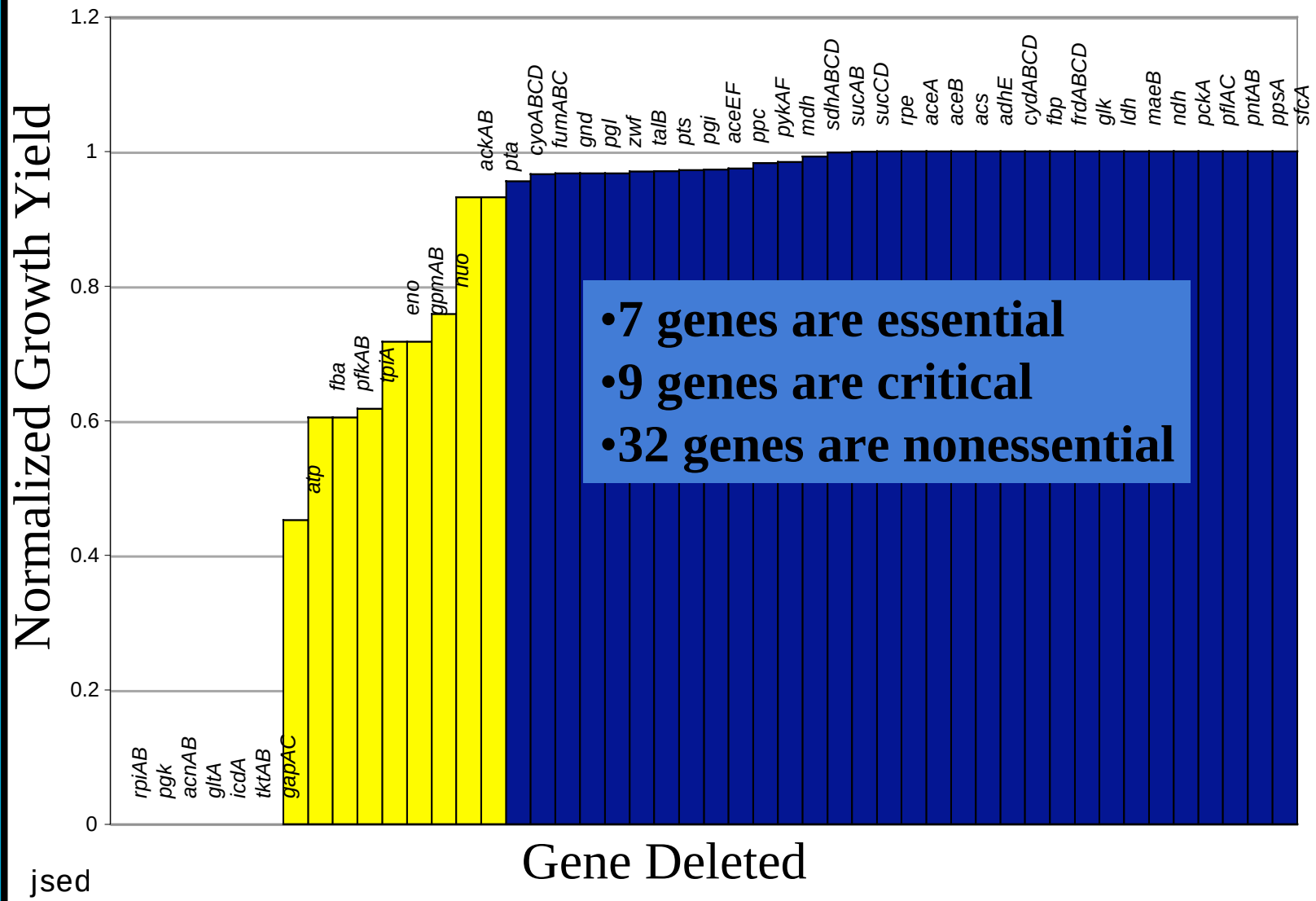
Map Check

Testing our predictions

Mutant Analysis

Using flux balance analysis
to study the effect of
gene deletions

Predicted E. coli mutant growth



Edwards and Palsson (2000) PNAS

Prediction Accuracy

Experimental/Predictions

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*	-/-			

Edwards and Palsson (2000) PNAS

Map Check

Testing our predictions:
High throughput analysis of FBA gene deletion results

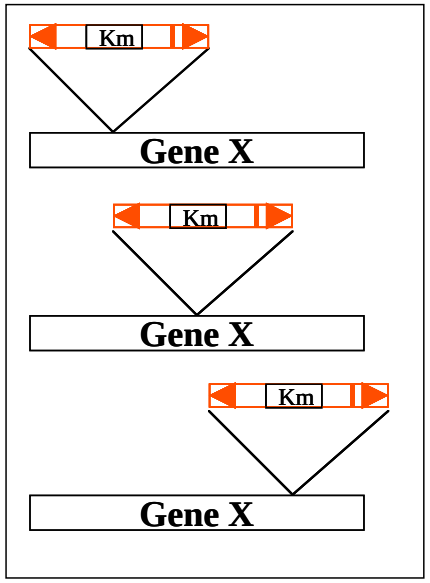
Gene deletion analysis

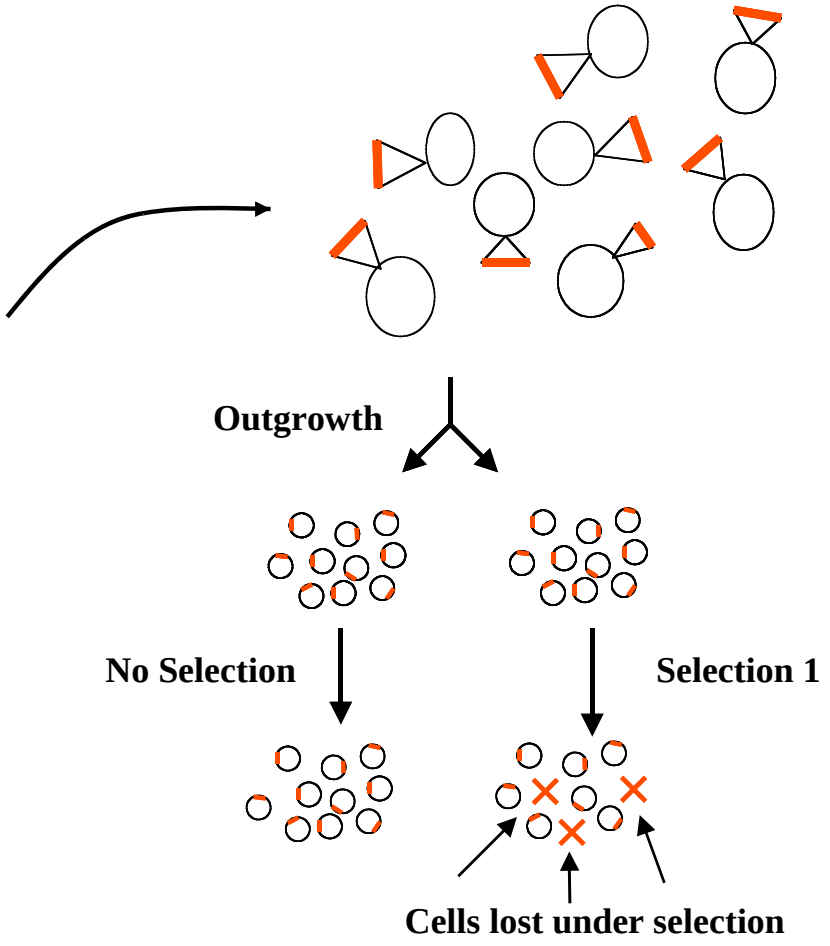
Badarinarayana, V., Estep, P.W., 3rd, Shendure, J., Edwards, J., Tavazoie, S., Lam, F. and Church, G.M. (2001) **Selection analyses of insertional mutants using subgenic-resolution arrays.** *Nat Biotechnol*, **19,** 1060-1065.

Gene deletion analysis

- Test the FBA predictions for mutant growth rate for ALL gene mutants at one time.
- Random, high-density, tagged insertional mutagenesis of the *E. coli* genome.
- Negative selection on the library of mutants.
- Read-out to determine population-wide changes in representation... Under a specific negative selection, disruption of which genomic sequences results in reduced growth rates?

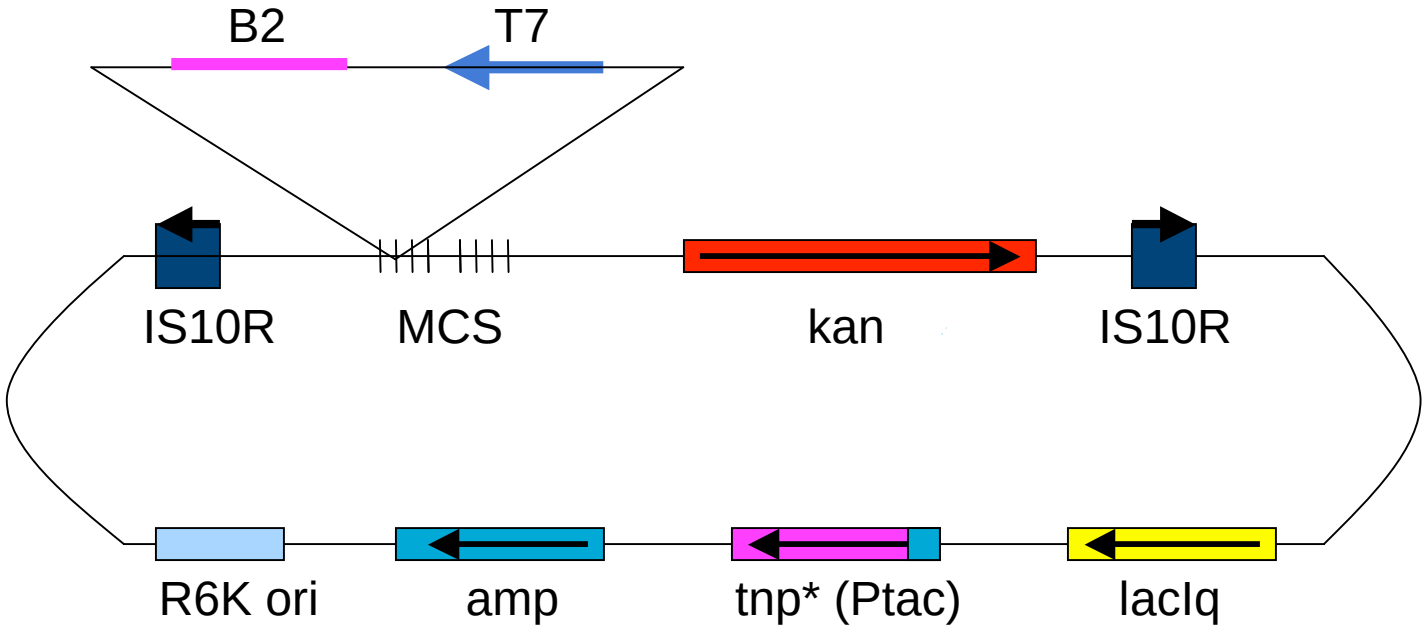
Transposon mutagenesis





Gene deletion analysis

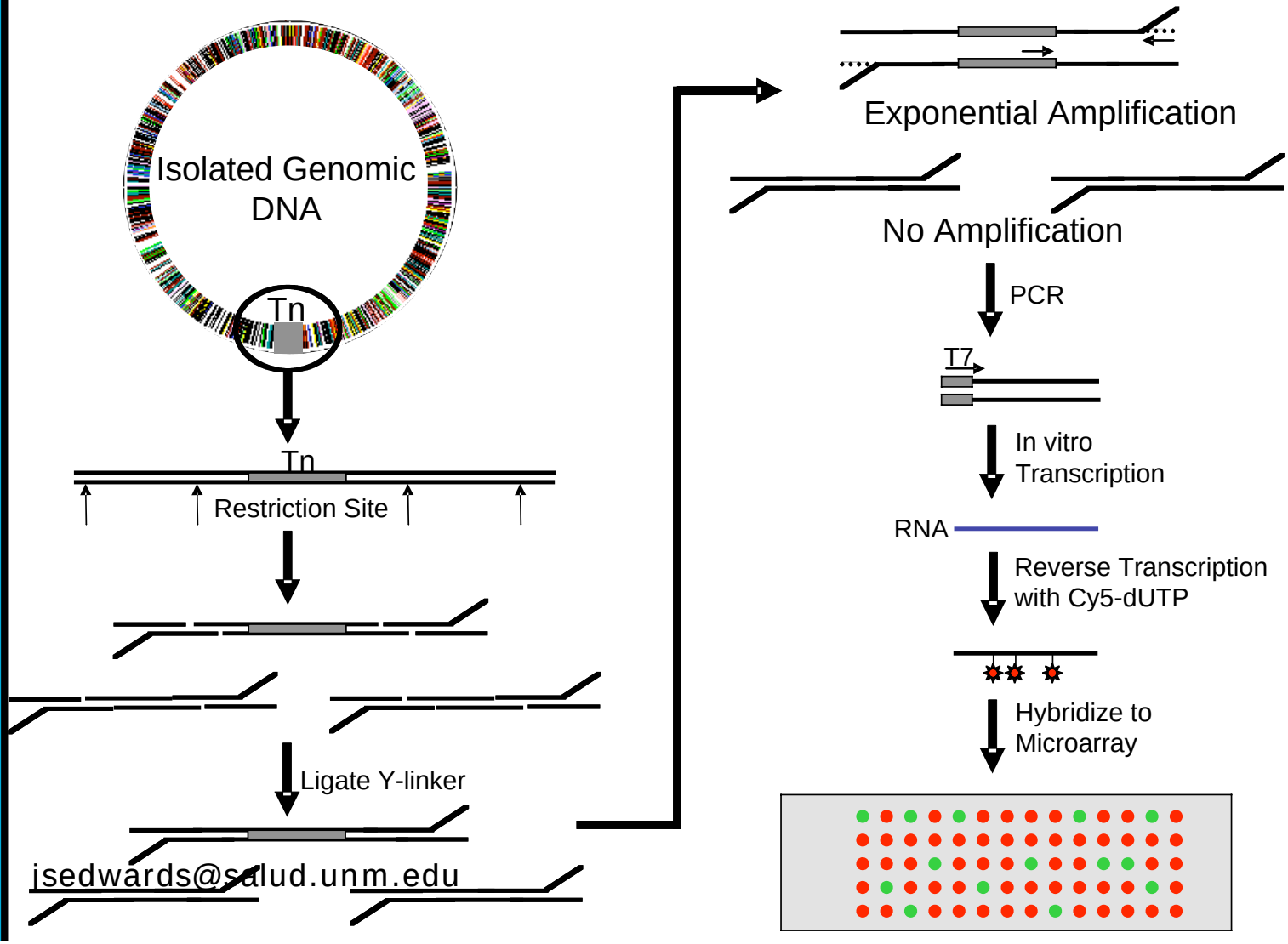
- "suicide" vector (R6K γ-ori; pir-strain restricted)
- Encodes variant of the Tn10 transposase with reduced specificity for hot spots.
- transposon element carries *kan* marker & MCS.
- T7 promoter



Badarinarayana, et al. (2001) Nat Biotechnol

Labeling the DNA

Badarinarayana, et al. (2001) Nat Biotechnol



Gene deletion analysis

Table 1. *Escherichia coli* genes exhibiting largest fold decrease in signal

Gene[a]	Fold decrease[b]	Functional subcategory	Functional category
rfaC	140	Lipopolysaccharide	Macromolecule synthesis, modification
speF	125	Polyamine biosynthesis	Central intermediary metabolism
metB	114	Methionine	Amino acid biosynthesis
cysK	106	Cysteine	Amino acid biosynthesis
cydD	69	ABC superfamily (membrane)	Transport/binding proteins
gltA	64	TCA cycle	Energy metabolism, carbon
iciA	64	DNA replication, repair	Macromolecule synthesis, modification
aroA	63	Chorismate	Amino acid biosynthesis
purN	58	Purine ribonucleotide biosynthesis	Nucleotide biosynthesis
xylB	58	Carbon compounds	Degradation of small molecules

[a]Indicates the gene containing the insertion.
[b]Indicates the fold change derived from the ratio of the competitively selected library to the initial library.
[c]The complete list of genes analyzed is included in the supplemental data (http://arep.med.harvard.edu/).

Badarinarayana, et al. (2001) Nat Biotechnol

Gene Deletions and FBA

Table 5. Comparison of genetic footprinting data with FBA model predictions

Predictions from model	Number of genes within prediction class	Negatively selected[a]	Not negatively selected[b]
Essential	143	80	63
Reduced growth rate	46	24	22
Nonessential	299	119	180

[a]The number of genes within each class that contain negatively selected insertions. [b]The number of insertion containing genes within each class that were not negatively selected. The numbers in the last two columns were used to compute the χ^2 number and compute the *P* value. *P* value from $\chi^2 = 0.0039$.

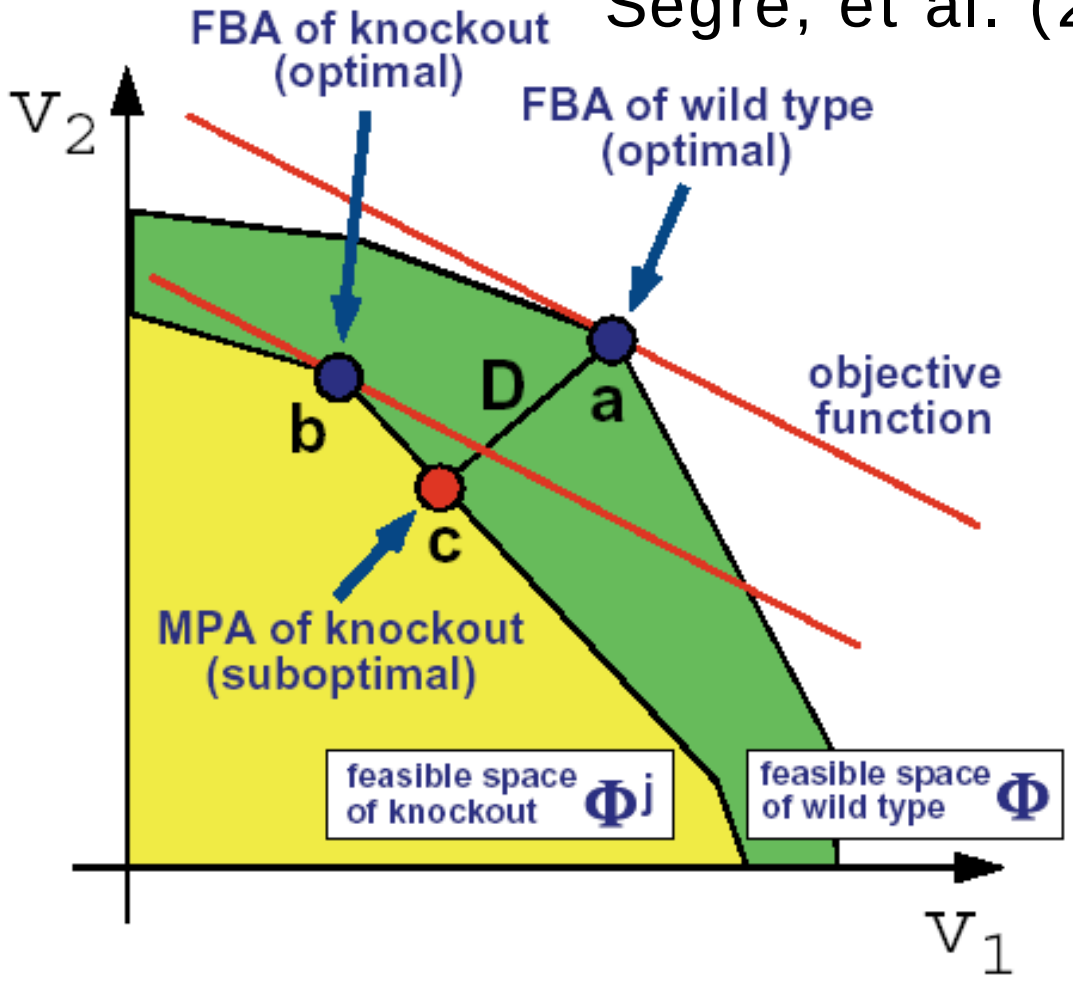
Badarinarayana, et al. (2001) Nat Biotechnol

Suboptimal mutants

- Mutants will not behave optimally
- Regulatory constraints can be adjusted to optimize the system subject to the physicochemical constraints
- Predictions of the initial behavior of mutants

Improved Growth Predictions

Segre, et al. (2002) PNAS



Improved Growth Predictions

Segre, et al. (2002) PNAS