Constraint-Based Workshops

6. Alternate Solutions

January 31st, 2008

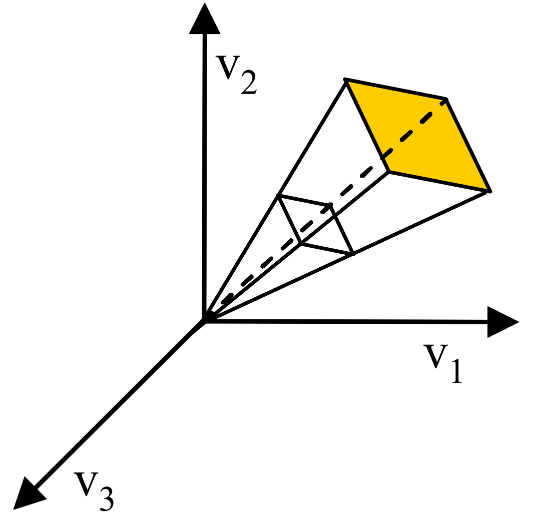
FBA Optimization Problem Statement

- Objective Function: A function that is maximized or minimized to identify optimal solutions
- Constraints: Place limits on the allowable values the solutions can take on.

Maximize: $c \cdot v$

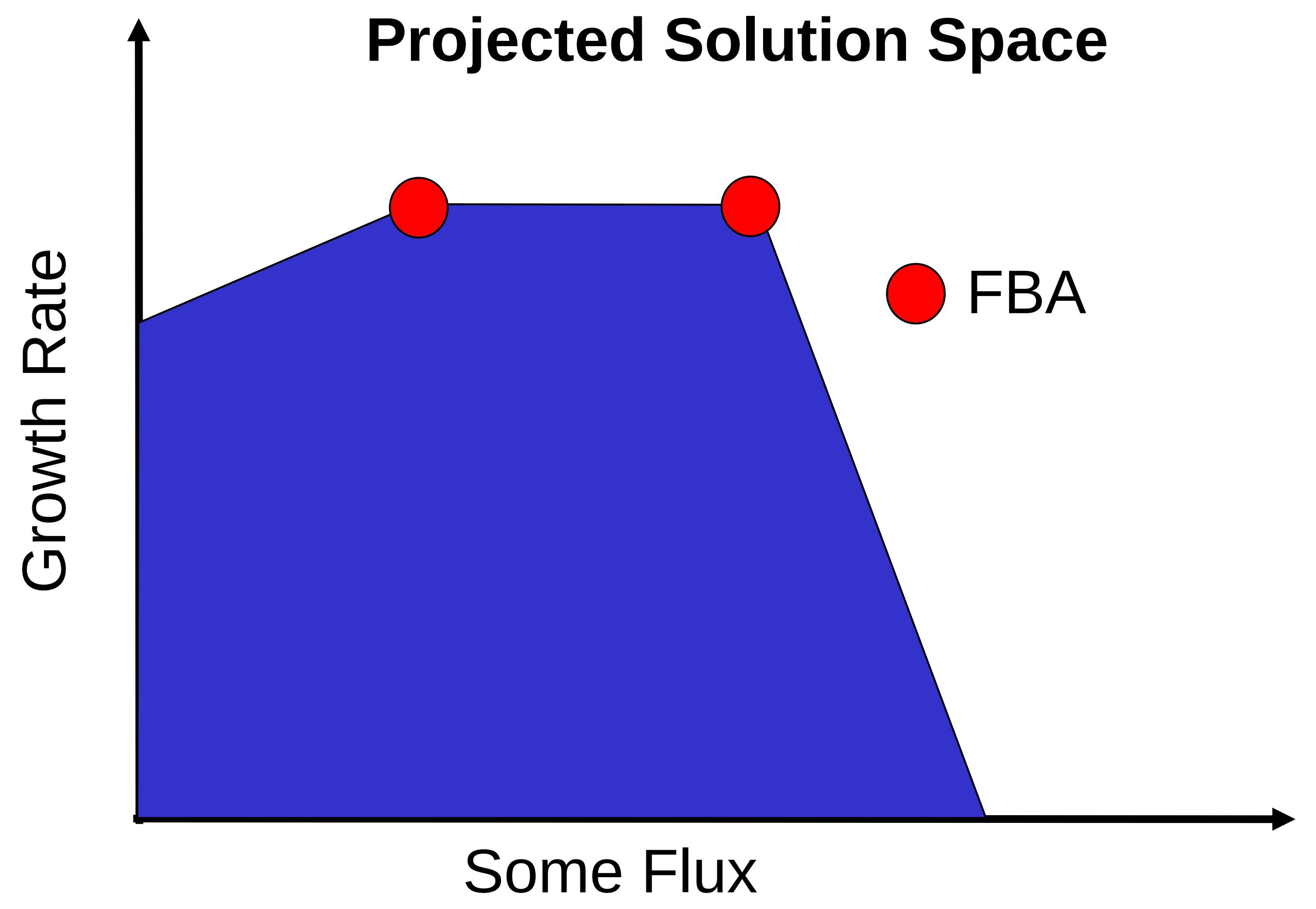
Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$



Equivalent Optimal Solutions Exist:

How can we find & characterize them?



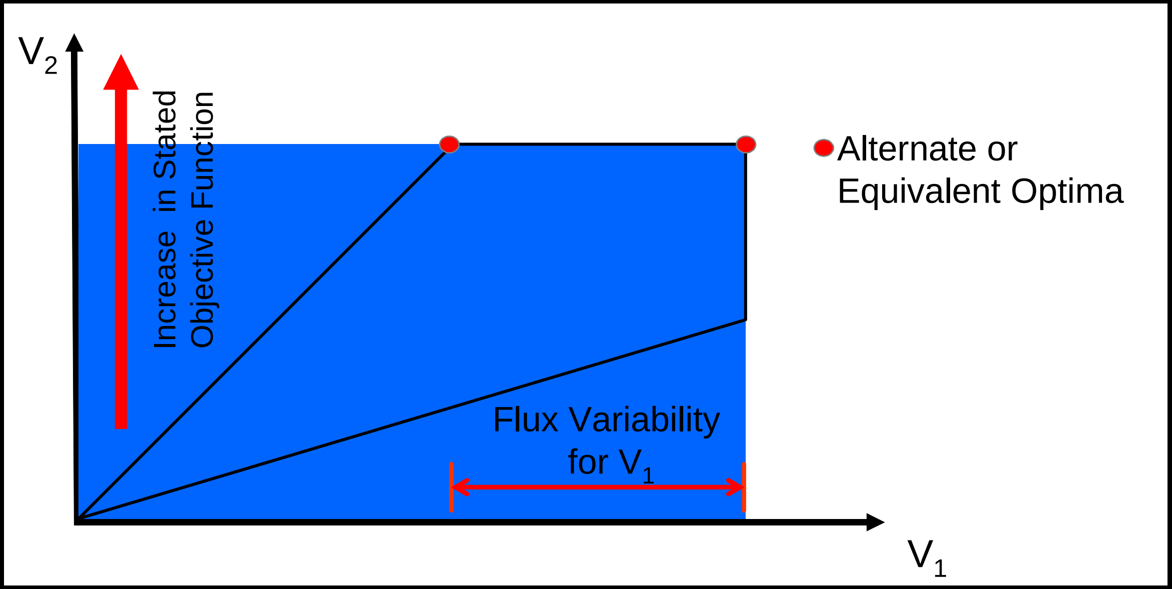
How many solutions are there?

- Most FBA solutions in genome-scale networks are not unique.
  - The value of the objective function is unique.
  - The set of fluxes giving rise to the objective function are often not unique.
- For *E. coli* optimal growth (GS network), there are likely thousands of equivalent optimal solutions.

Flux Variability Analysis

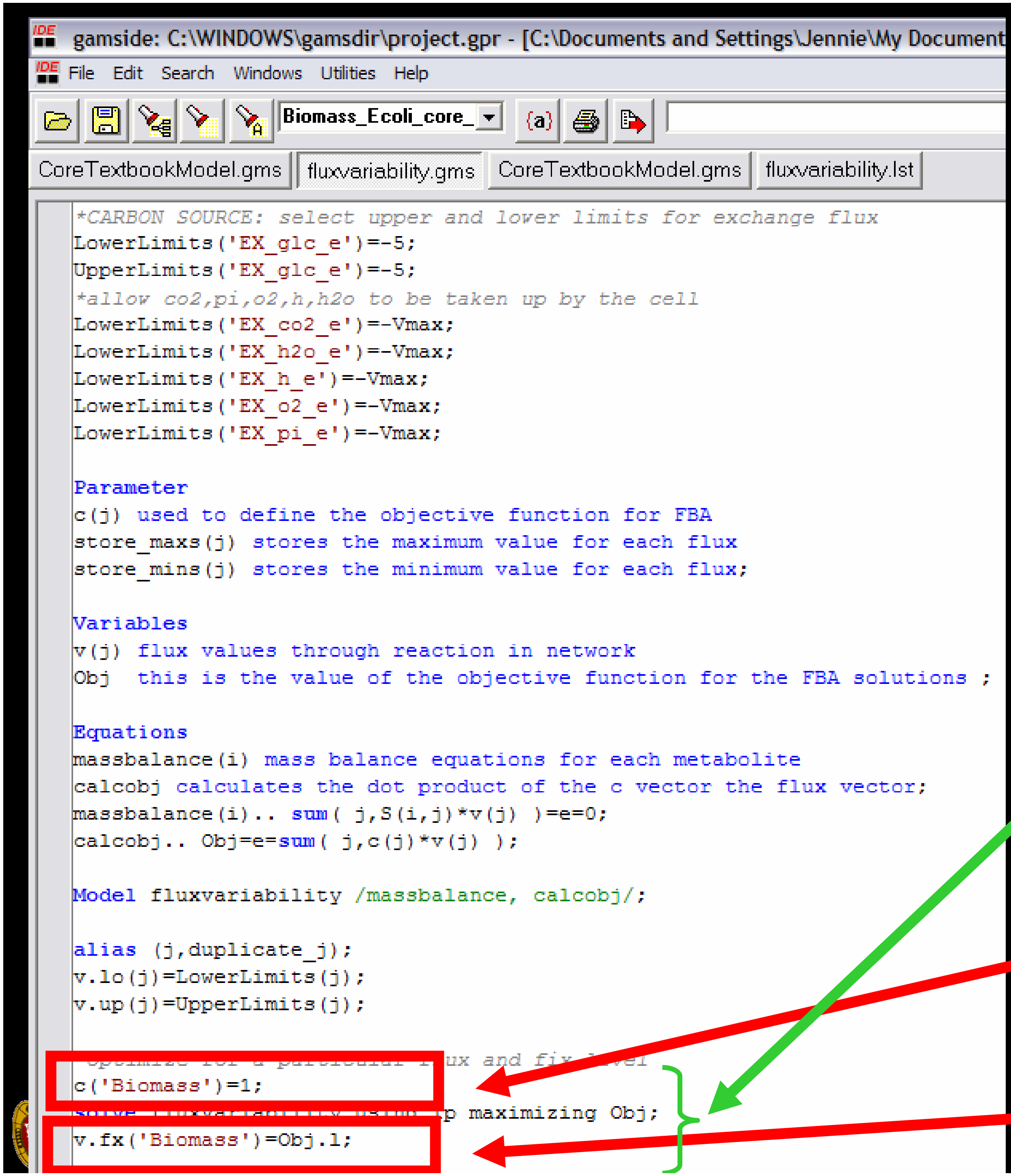
Flux Variability Analysis:

- First, identify the maximum value of the objective function and constrain objective function to this value.
- Second, minimize and maximize each flux independently to identify flexibility in the fluxes across alternate optima.



If we have *n* fluxes, we basically solve 1+2*n* FBA problems

Flux Variability Analysis



```
*CARBON SOURCE: select upper and lower limits for exchange flux
LowerLimits('EX_glc_e')=-5;
UpperLimits('EX_glc_e')=-5;
*allow co2,pi,o2,h,h2o to be taken up by the cell
LowerLimits('EX_co2_e')=-Vmax;
LowerLimits('EX_h2o_e')=-Vmax;
LowerLimits('EX_h_e')=-Vmax;
LowerLimits('EX_o2_e')=-Vmax;
LowerLimits('EX_pi_e')=-Vmax;

Parameter
c(j) used to define the objective function for FBA
store_maxs(j) stores the maximum value for each flux
store_mins(j) stores the minimum value for each flux;

Variables
v(j) flux values through reaction in network
Obj  this is the value of the objective function for the FBA solutions ;

Equations
massbalance(i) mass balance equations for each metabolite
calcobj calculates the dot product of the c vector the flux vector;
massbalance(i).. sum( j,S(i,j)*v(j) )=e=0;
calcobj.. Obj=e=sum( j,c(j)*v(j) );

Model fluxvariability /massbalance, calcobj/;

alias (j,duplicate_j);
v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);

*optimize for a particular flux and fix level
c('Biomass')=1;
solve fluxvariability using lp maximizing Obj;
v.fx('Biomass')=Obj.l;
```

Note: to calculate the variability across all solutions not just optimal ones just comment out the three lines with a *

Define Flux to Optimize

Fix the level of flux to optimal value

Flux Variability Calculations: Max μ

- How many fluxes vary for anaerobic optimal growth on glucose (where you are maximizing biomass).

- What does it imply about the number of alternate optima if there are no varying fluxes?

- How many fluxes can vary if you look at solutions which have at least 90% of the maximum growth rate (ie. biomass flux)?
  - HINT: Change line 50 to be:

```
v.lo('Biomass')=0.9*Obj.l;
```

Flux Variability Calculations: Max Ethanol Production

- How many fluxes vary for anaerobic production of ethanol from glucose (where now you first optimize for the EX_etoh_e flux)?

- How many fluxes are fixed to non-zero value?

- How many reactions are not needed for ethanol production? Could fluxes through these reactions reduce ethanol production?

Flux Variability Calculations: Max μ

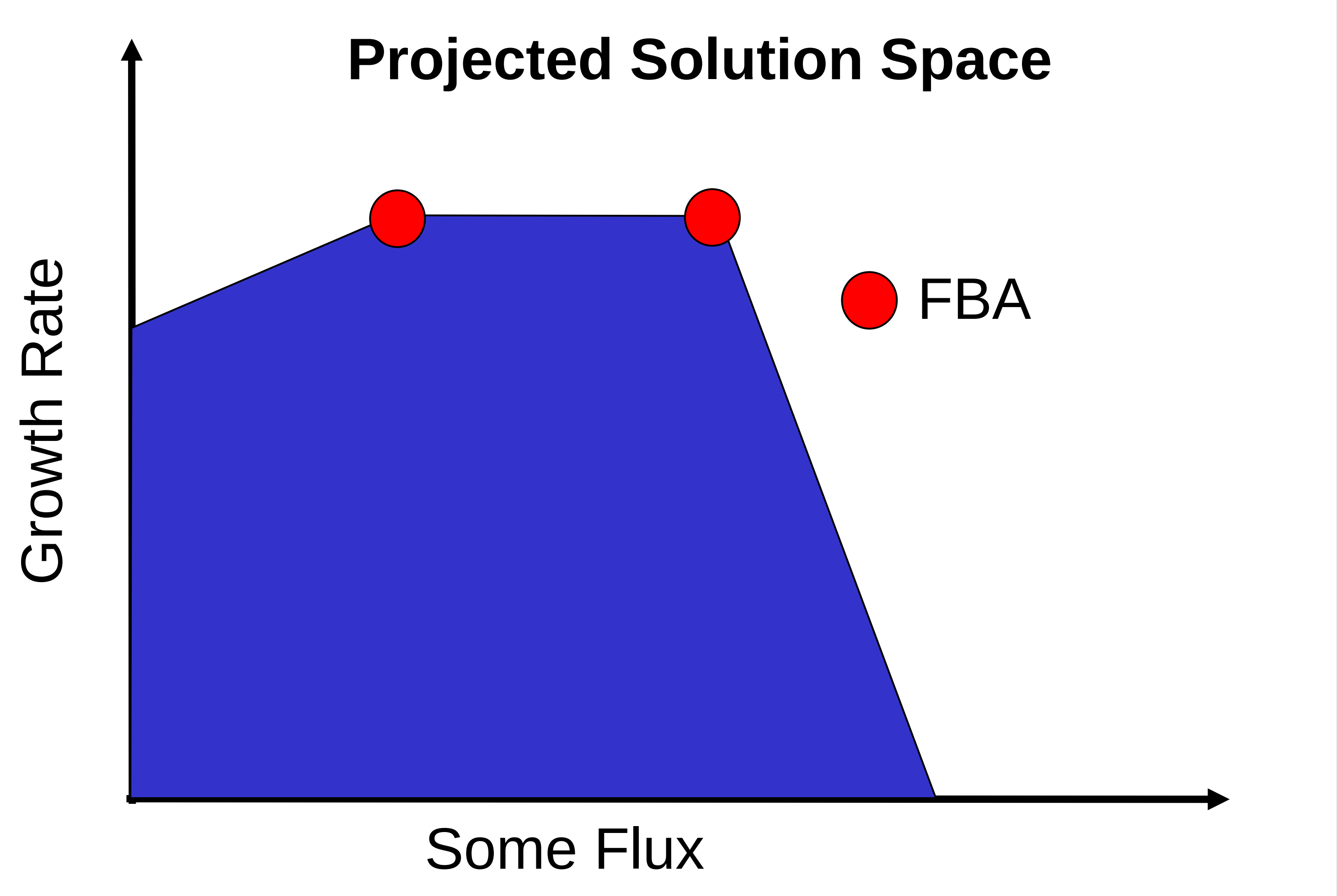
- How many fluxes vary for anaerobic optimal growth on glucose (where you are maximizing biomass).
  - ANS: 2
- What does it imply about the number of alternate optima if there are no varying fluxes?
  - ANS: It means there is only one solution and it is unique.
- How many fluxes can vary if you look at solutions which have at least 90% of the maximum growth rate (ie. biomass flux)?
  - ANS: 70

Flux Variability Calculations: Max Ethanol Production

- How many fluxes vary for anaerobic production of ethanol from glucose (where now you first optimize for the EX_etoh_e flux)?
  - ANS: 18
- How many fluxes are fixed to non-zero value?
  - ANS: 15
- How many reactions are not needed for ethanol production? Could fluxes through these reactions reduce ethanol production?
  - ANS: 44
  - Non-zero fluxes through these reactions will reduce ethanol production or make your problem infeasible.

Enumerating Alternate Optima

Equivalent Optimal Solutions Exist:
We can used Mixed Integer Linear Programming to Find Them.



Algorithm For Identifying Different “Corner” Points

- GOAL: given your past solutions, find a new one that uses a different set of non-zero fluxes in the solution.

- The result is that you will identify all the different corner point solutions that have the same objective function value.

- Any optimal solution, can be written as the weighted sum of the corner point optimal solutions.

Metabolic Network Example

Reaction List

v_1: $A \rightarrow B$

v_2: $A \rightarrow C$

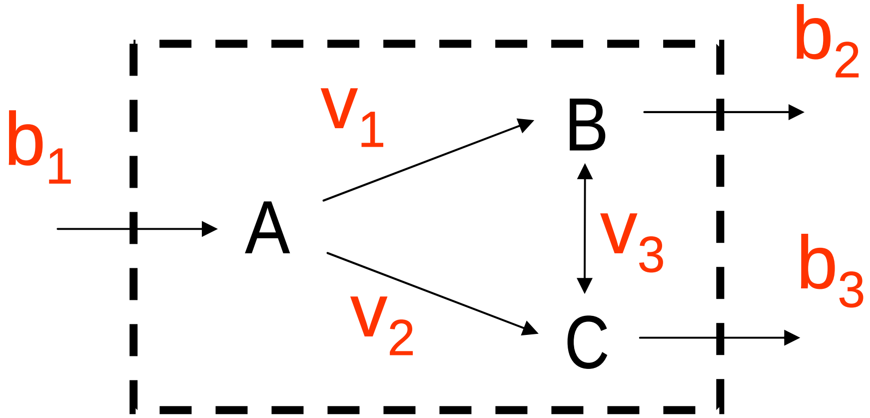
v_3: $B \leftrightarrow C$

b_1: $\rightarrow A$

b_2: $B \rightarrow$

b_3: $C \rightarrow$

Metabolic Map



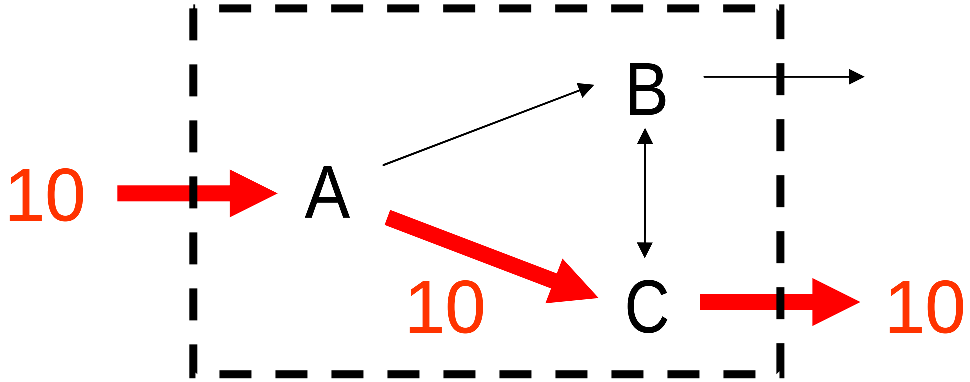
Maximize $Z = c \cdot v = b_3$

Such that $S \cdot v = 0$

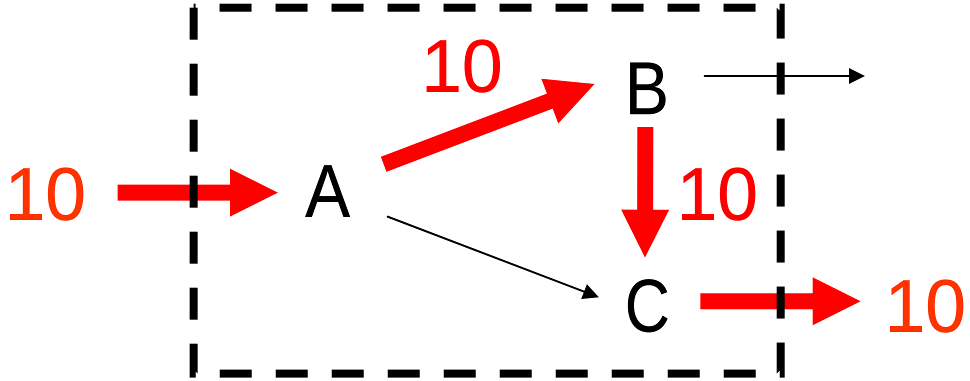
$0 \leq v_1, v_2, b_1, b_2, b_3 \leq 10$

$-10 \leq v_3 \leq 10$

Solution 1:



Solution 2:



1. Find non-zero fluxes (NZ^{J-1}) at current solution, J-1.
2. Pick at least one NZ^{J-1} flux to become zero at next solution ($y_i=1$).

$$\sum_{i \in NZ^{J-1}} y_i \geq 1$$

3. Make sure that the set of non-zero fluxes haven't been visited at previous k iterations.

$$y_i + w_i \leq 1$$

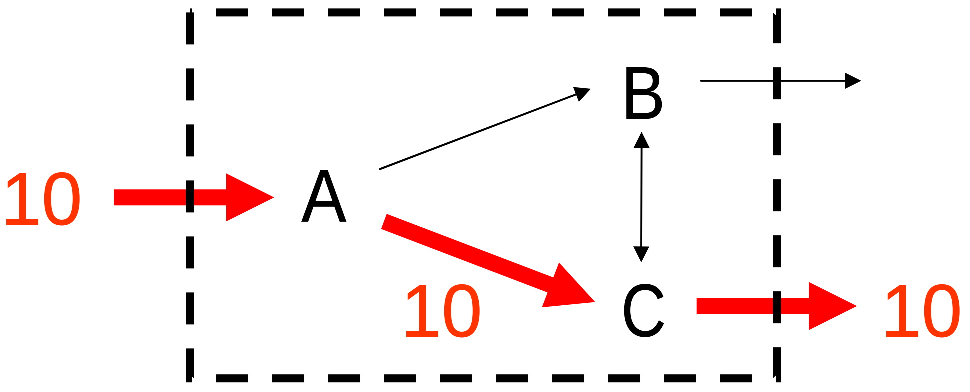
$$\sum_{i \in NZ^{k}} w_i \leq |NZ^k| - 1$$

4. Constrain those selected fluxes to have zero flux.

$$w_i \cdot v_{min} \leq v_i \leq w_i \cdot v_{min}$$

5. Find solution & repeat.

Solution 1:



1. My NZ^1 fluxes are v_2,b_1,b_3.
2. I will pick v_2 to become zero at next solution: $y_{v2}=1$ & $y_{b1}=y_{b3}=0$;
3. w_{v2} = 0; lets assume that w_{b1} and w_{b3} = 1.

$$\sum_{i \in NZ^1} w_i \le |NZ^1|-1 \quad (2 \le 2)$$

4. v_2 = 0 and the other fluxes can be between their normal v_{min} and v_{max} values.

1. Find non-zero fluxes (NZ^{J-1}) at current solution, J-1.
2. Pick at least one NZ^{J-1} flux to become zero at next solution (y_i=1).

$$\sum_{i \in NZ^{J-1}} y_i \ge 1$$

3. Make sure that the set of non-zero fluxes haven't been visited at previous k iterations.

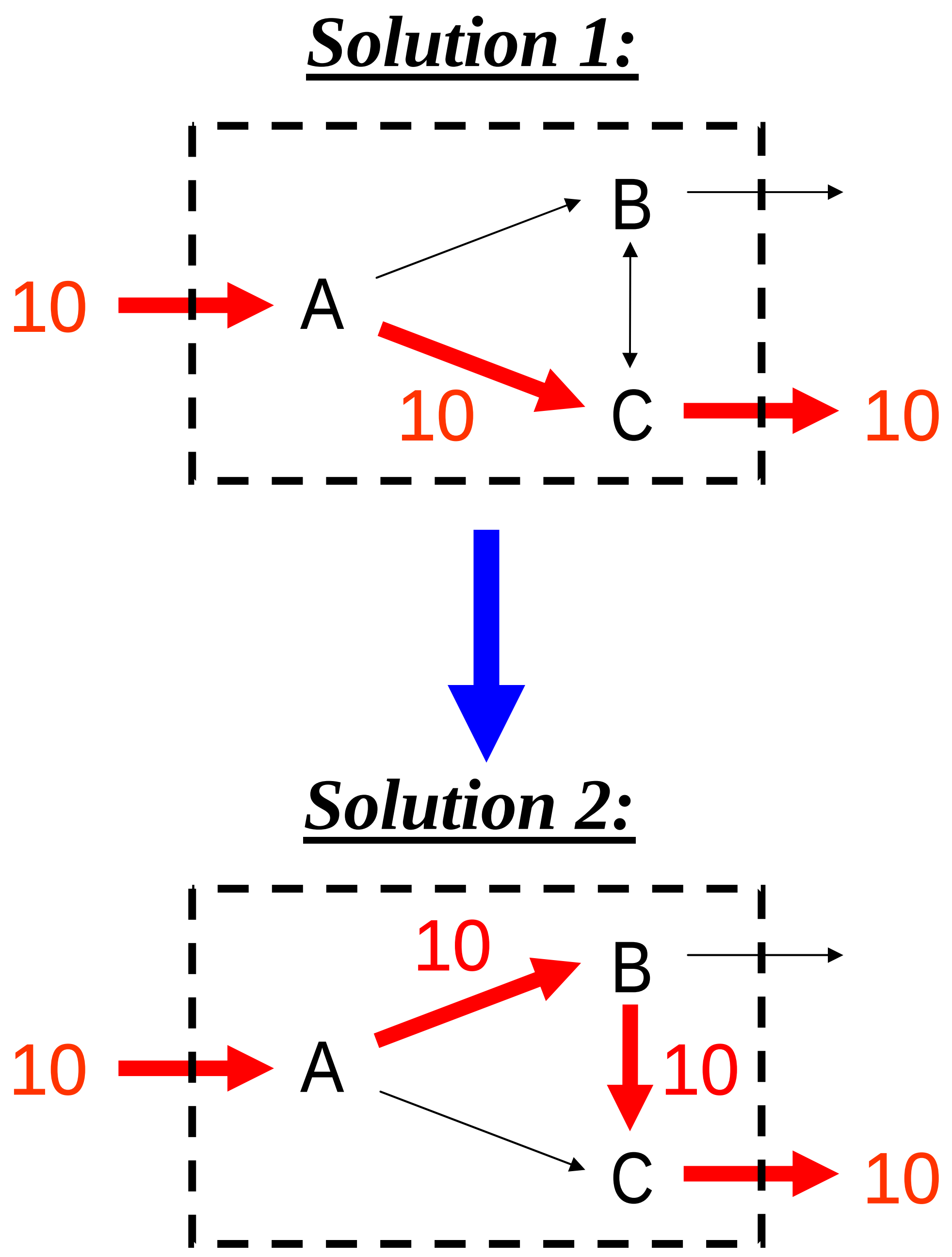
$$y_i + w_i \le 1$$

$$\sum_{i \in NZ^k} w_i \le |NZ^k|-1$$

4. Constrain those selected fluxes to have zero flux.

$$w_i \cdot v_{min} \le v_i \le w_i \cdot v_{min}$$

5. Find solution & repeat.



1. Find non-zero fluxes (NZ^{J-1}) at current solution, J-1.
2. Pick at least one NZ^{J-1} flux to become zero at next solution (y_i=1).

$$\sum_{i \in NZ^{J-1}} y_i \geq 1$$

3. Make sure that the set of non-zero fluxes haven't been visited at previous k iterations.

$$y_i + w_i \leq 1$$

$$\sum_{i \in NZ^k} w_i \leq |NZ^k| - 1$$

4. Constrain those selected fluxes to have zero flux.

$$w_i \cdot v_{min} \leq v_i \leq w_i \cdot v_{min}$$

5. Find solution & repeat.

Alternate Optima Calculations

- How many alternate solutions are there for glucose aerobic growth, where you maximize for biomass production?

- If you allow ATPM to be reversible, by setting lower limit to –Vmax:
  - How many alternate solutions are there?
  - Does biomass production improve if you allow free synthesis of ATP? Is this realistic?

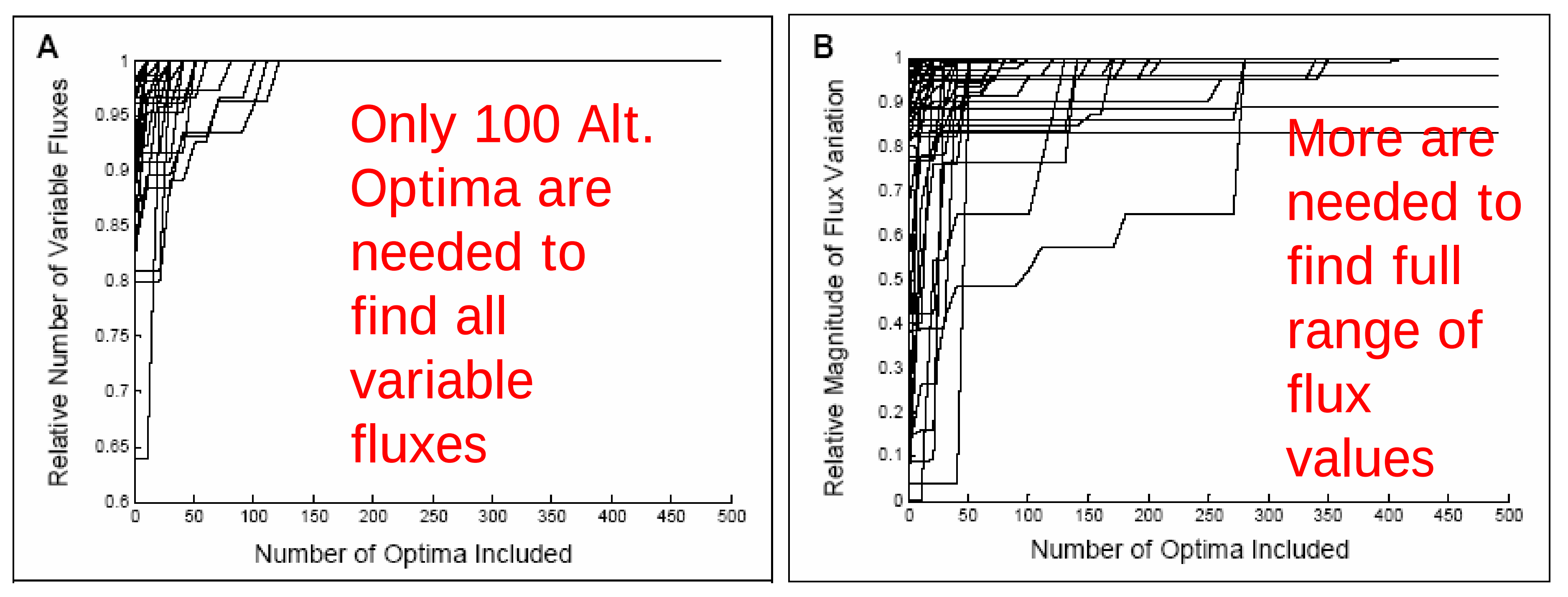
Alternate Optima Calculations

- How many alternate solutions are there for glucose aerobic growth, where you maximize for biomass production?
  - ANS: Only 1 solution
- If you allow ATPM to be reversible, by setting lower limit to –Vmax:
  - How many alternate solutions are there?

    ANS: three solutions
  - Does biomass production improve if you allow free synthesis of ATP? Is this realistic?

    ANS: Yes (0.49 vs. 0.66), but not realistic

Flux Variability vs. Alternate Optima



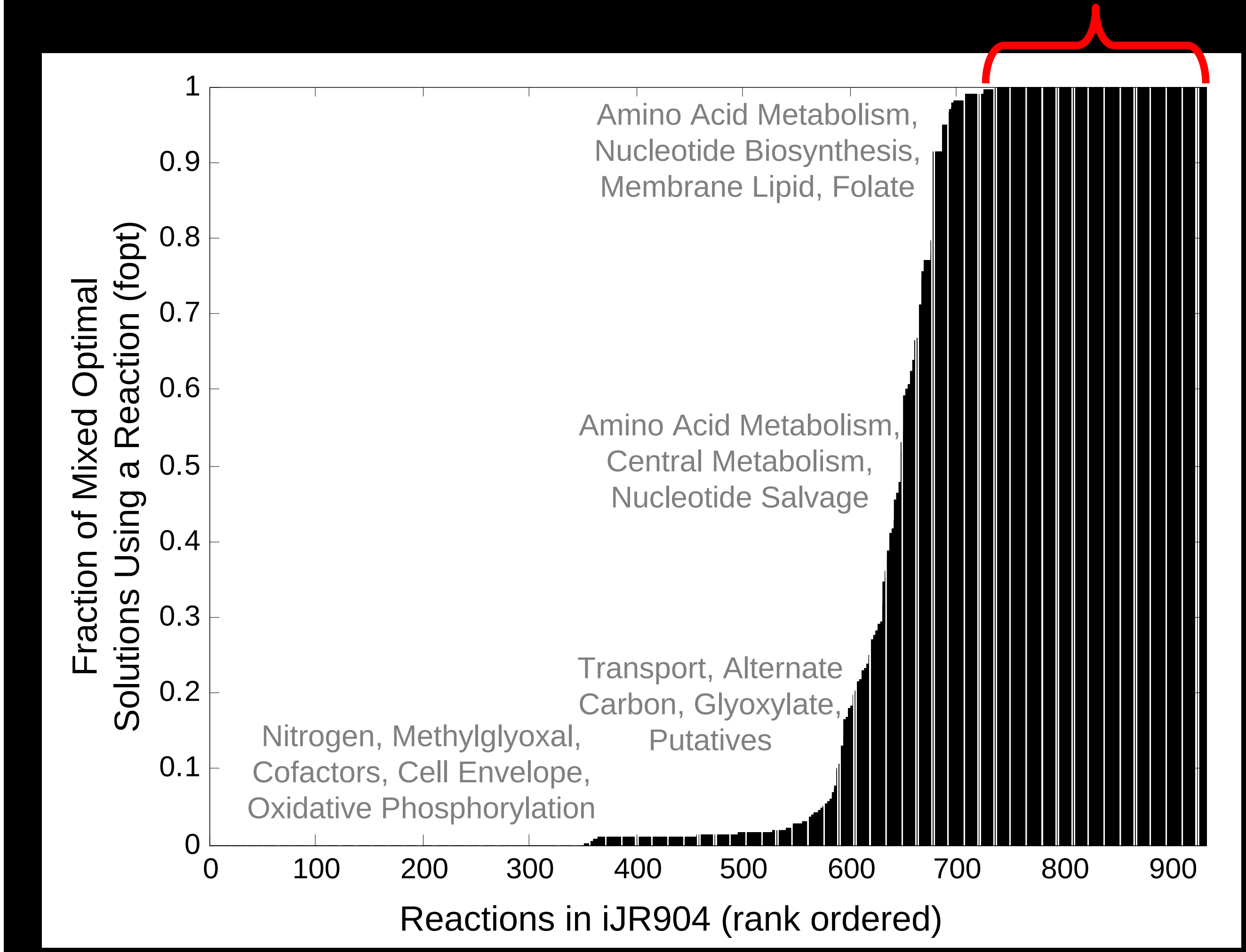
Only 100 Alt. Optima are needed to find all variable fluxes

More are needed to find full range of flux values

Figure 1 Comparisons of properties for sampled optima with all optima. The number of variable fluxes and the allowable ranges for these fluxes across all optima were calculated using a flux variability analysis. Each line is for one of the 88 carbon sources capable of supporting aerobic growth. (*A*) shows that as the number of calculated optima increases, the number of variable fluxes found in these sampled optimal solutions approaches the total number of variable fluxes. (*B*) shows how the magnitude of the flux variations is represented by the sampled optima relative to the actual flux variability across all optima.

Reed and Palsson. Genome Research (2004). 14:1797–1805

Reaction Usage Across 136 Different Environmental Conditions



<u>201 Reactions Were Always Used:</u>

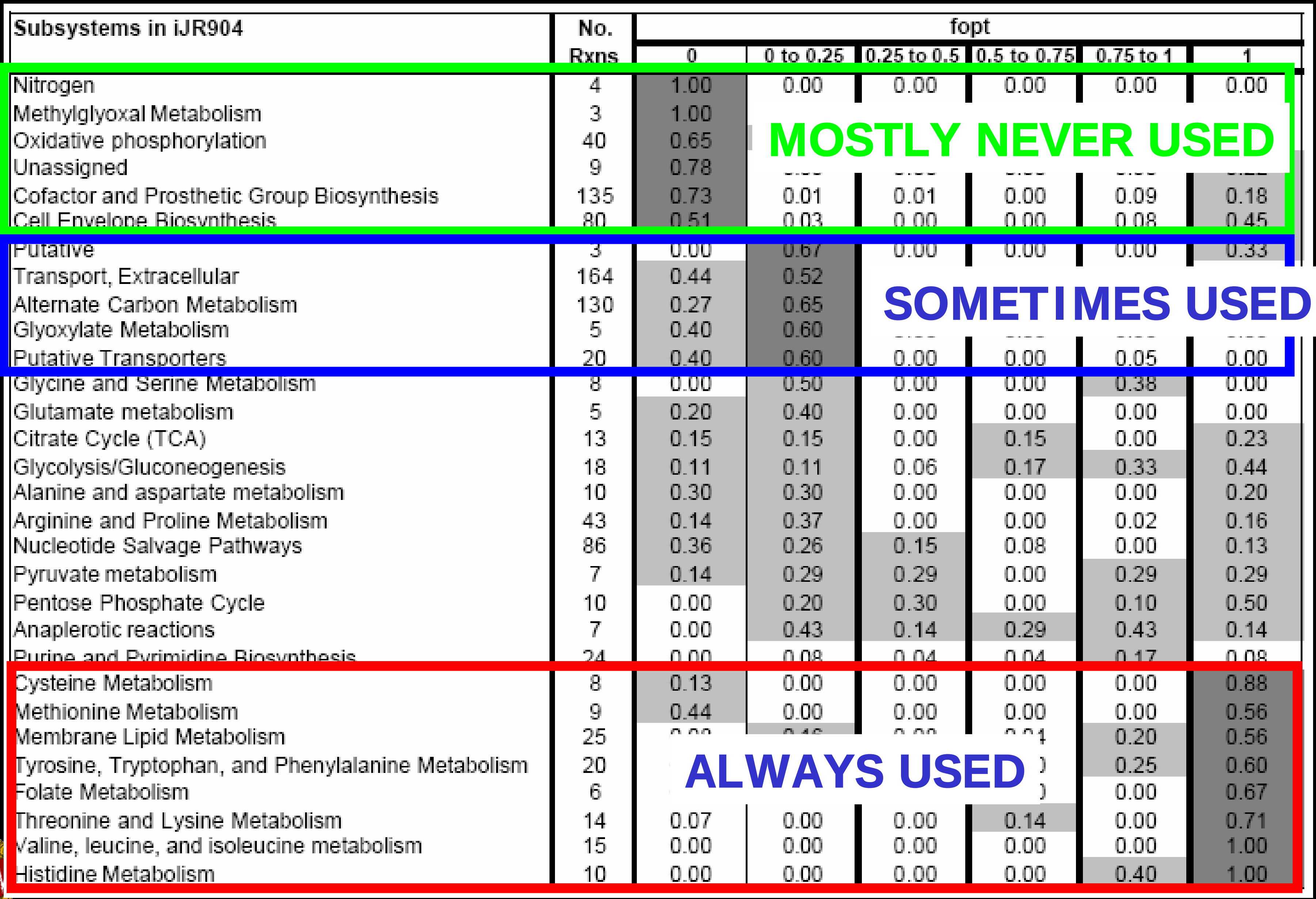
81 Reactions Associated With Essential Genes in Rich Media

20 Reactions Lack Associated Genes

20 Reactions Have Multiple Isozymes

Reed and Palsson. Genome Research (2004). 14:1797–1805

Usage by Metabolic Subsytem



Subsystems in iJR904	No. Rxns	fopt					
		0	0 to 0.25	0.25 to 0.5	0.5 to 0.75	0.75 to 1	1
Nitrogen	4	1.00	0.00	0.00	0.00	0.00	0.00
Methylglyoxal Metabolism	3	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Oxidative phosphorylation	40	0.65	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Unassigned	9	0.78	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Cofactor and Prosthetic Group Biosynthesis	135	0.73	0.01	0.01	0.00	0.09	0.18
Cell Envelope Biosynthesis	80	0.51	0.03	0.00	0.00	0.08	0.45
Putative	3	0.00	0.67	0.00	0.00	0.00	0.33
Transport, Extracellular	164	0.44	0.52	[illegible]	[illegible]	[illegible]	[illegible]
Alternate Carbon Metabolism	130	0.27	0.65	[illegible]	[illegible]	[illegible]	[illegible]
Glyoxylate Metabolism	5	0.40	0.60	[illegible]	[illegible]	[illegible]	[illegible]
Putative Transporters	20	0.40	0.60	0.00	0.00	0.05	0.00
Glycine and Serine Metabolism	8	0.00	0.50	0.00	0.00	0.38	0.00
Glutamate metabolism	5	0.20	0.40	0.00	0.00	0.00	0.00
Citrate Cycle (TCA)	13	0.15	0.15	0.00	0.15	0.00	0.23
Glycolysis/Gluconeogenesis	18	0.11	0.11	0.06	0.17	0.33	0.44
Alanine and aspartate metabolism	10	0.30	0.30	0.00	0.00	0.00	0.20
Arginine and Proline Metabolism	43	0.14	0.37	0.00	0.00	0.02	0.16
Nucleotide Salvage Pathways	86	0.36	0.26	0.15	0.08	0.00	0.13
Pyruvate metabolism	7	0.14	0.29	0.29	0.00	0.29	0.29
Pentose Phosphate Cycle	10	0.00	0.20	0.30	0.00	0.10	0.50
Anaplerotic reactions	7	0.00	0.43	0.14	0.29	0.43	0.14
Purine and Pyrimidine Biosynthesis	24	0.00	0.08	0.04	0.04	0.17	0.08
Cysteine Metabolism	8	0.13	0.00	0.00	0.00	0.00	0.88
Methionine Metabolism	9	0.44	0.00	0.00	0.00	0.00	0.56
Membrane Lipid Metabolism	25	[illegible]	[illegible]	[illegible]	[illegible]	0.20	0.56
Tyrosine, Tryptophan, and Phenylalanine Metabolism	20	[illegible]	[illegible]	[illegible]	[illegible]	0.25	0.60
Folate Metabolism	6	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.67
Threonine and Lysine Metabolism	14	0.07	0.00	0.00	0.14	0.00	0.71
Valine, leucine, and isoleucine metabolism	15	0.00	0.00	0.00	0.00	0.00	1.00
Histidine Metabolism	10	0.00	0.00	0.00	0.00	0.40	1.00

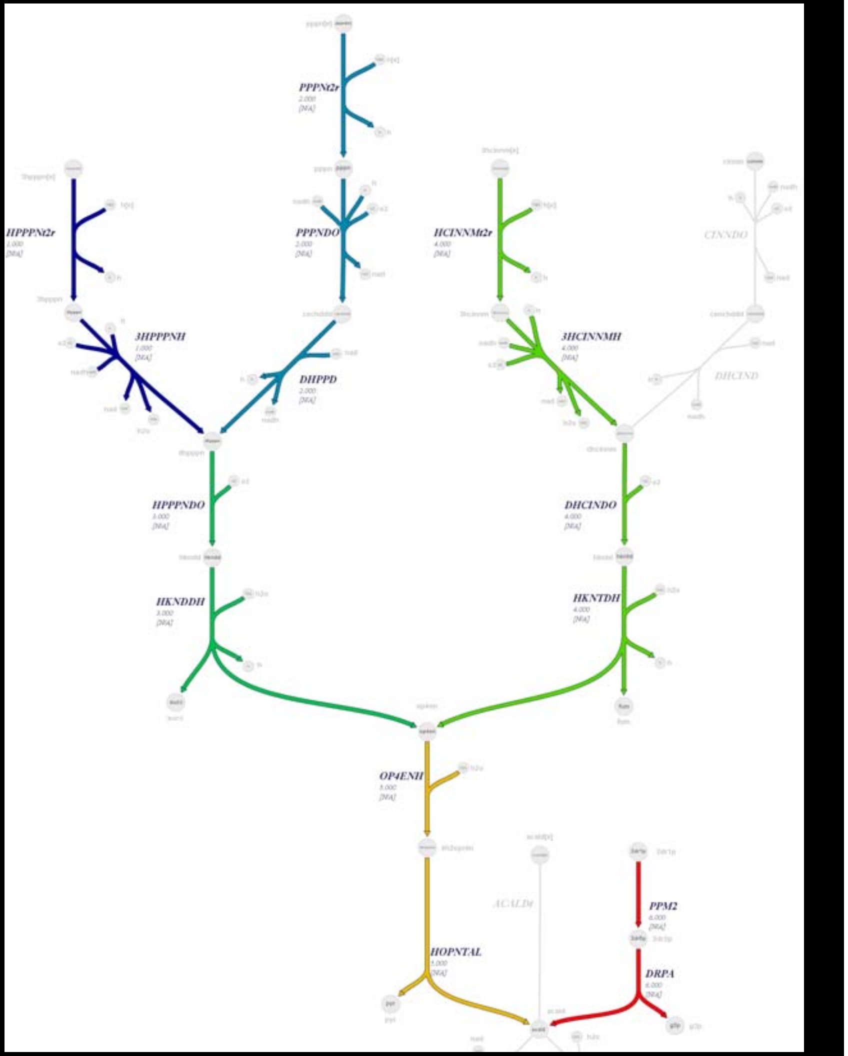
MOSTLY NEVER USED

SOMETIMES USED

ALWAYS USED

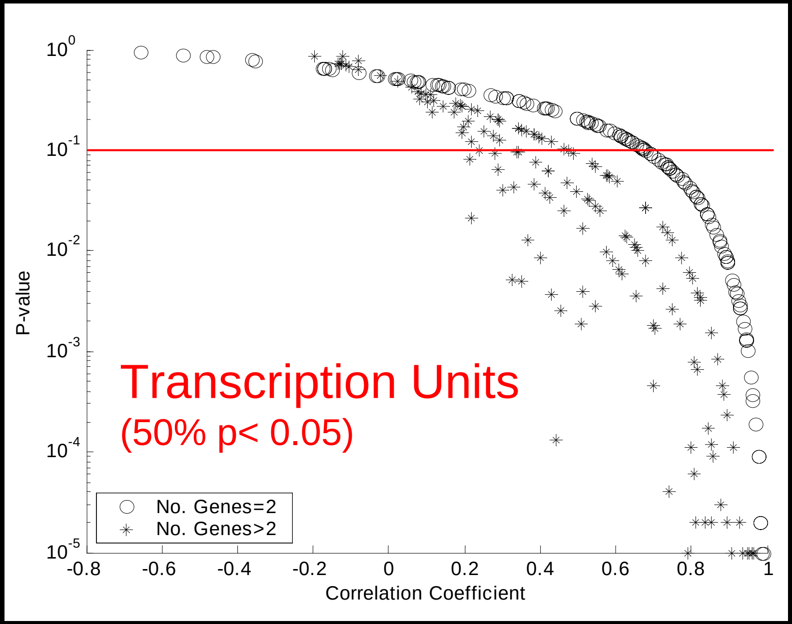
Correlated Reaction Sets in *E. coli*

Correlated Reaction Sets: Reactions where a non-zero flux through one reaction implies a non-zero flux through all other reactions in the set (and vice versa).



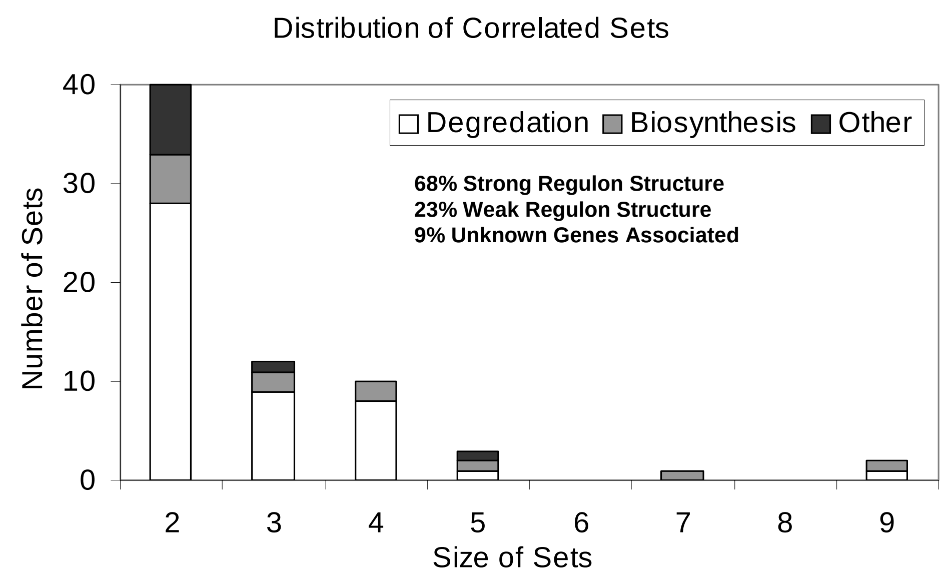
Correlated Reaction Sets in *E. coli*

Correlated Reaction Sets: Reactions where a non-zero flux through one reaction implies a non-zero flux through all other reactions in the set (and vice versa).

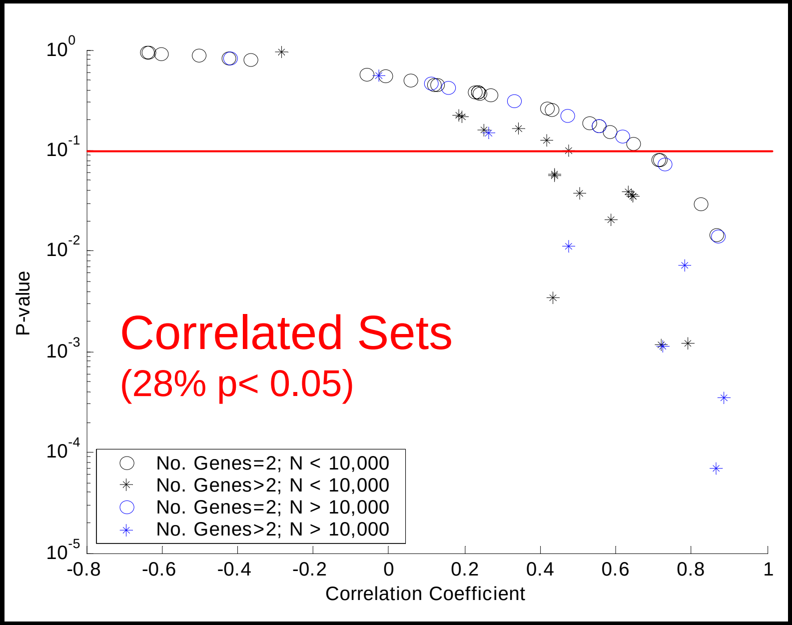


Transcription Units

($50\% \ p < 0.05$)



68% Strong Regulon Structure
23% Weak Regulon Structure
9% Unknown Genes Associated



Correlated Sets

($28\% \ p < 0.05$)