Using optimization to re-wire biological networks for improved biofuel production

Jennifer Reed

University of Wisconsin, Madison

Overview of Constraint-Based Modeling Sessions

1. Reconstructing metabolic networks and flux balance analysis
2. Finding alternate solutions and predicting the effects of gene knockout
3. Improving models using optimization
4. Using models for metabolic engineering

Reconstruction of Metabolic Networks

Genome-scale Metabolic Model Reconstruction

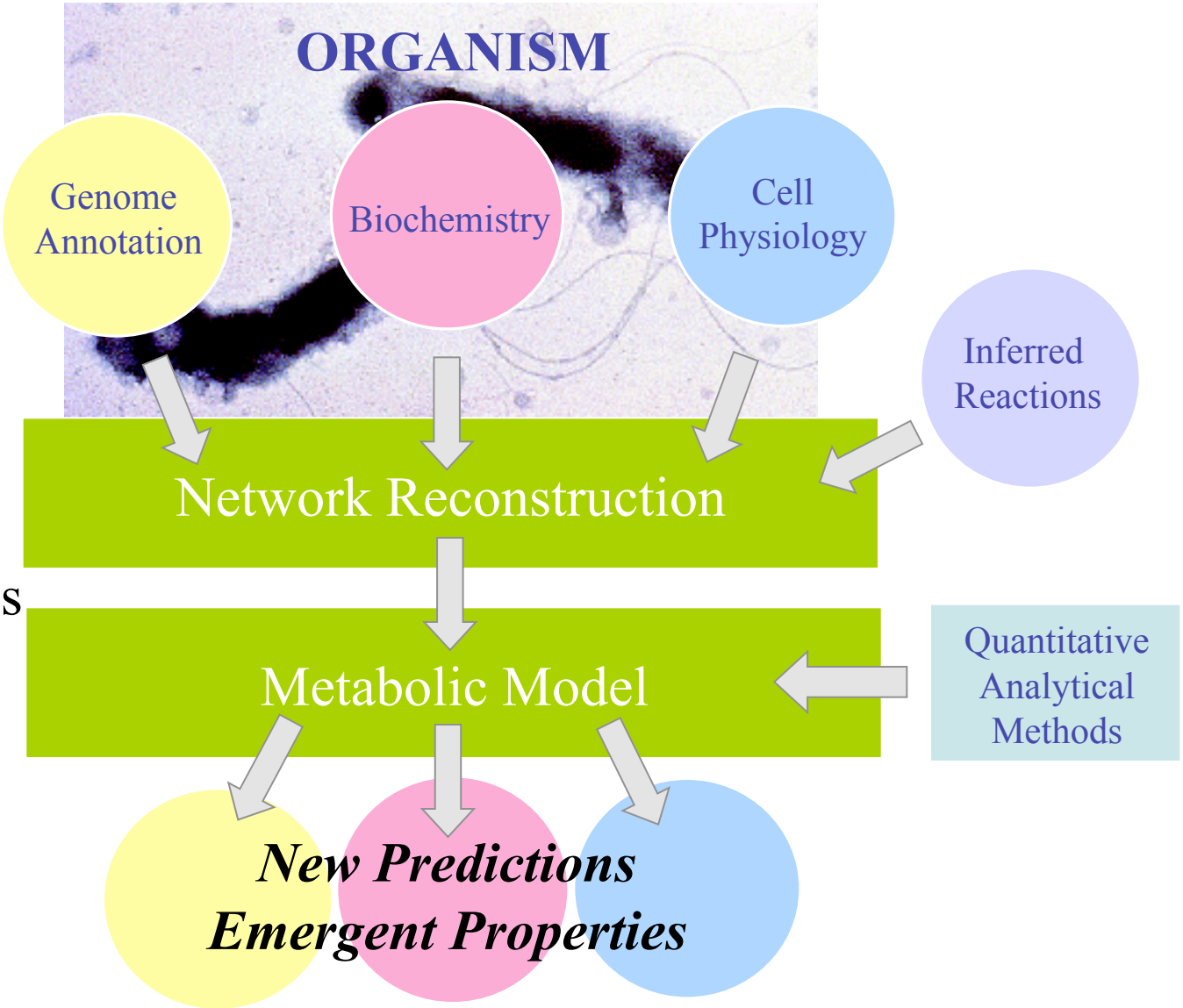
Genome Annotation
- by homology, location

Biochemical Data
- protein characterized

Physiological Data
- indirect, pathway known

Inferred Reactions
- indirect, inferred from biomass requirements

Quantitative Analysis
- simulate cell behavior
- drive experimental studies

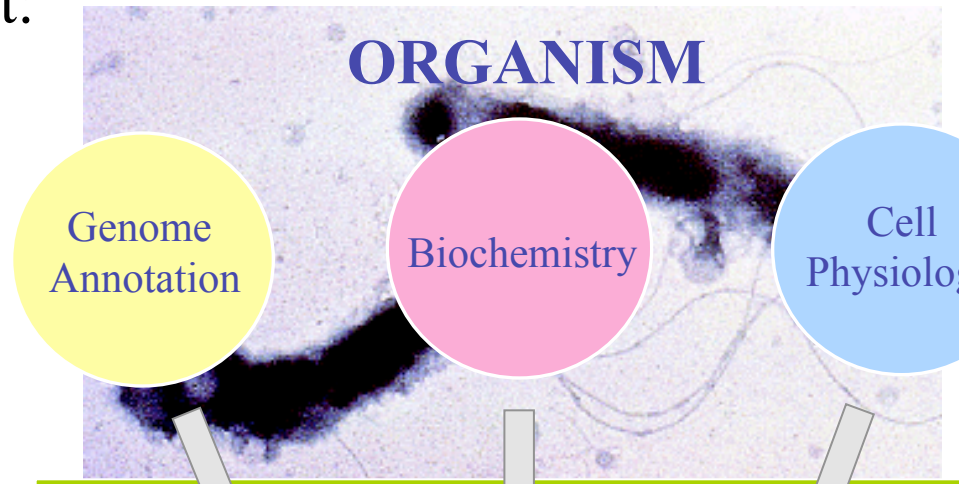


Genome-scale Metabolic Model Reconstruction

Model Development:
an iterative process

- Biochemical data
- Revised ORF assignments

Computational,
Biochemical
Investigation



Genome Annotation

Biochemistry

Cell Physiology

Inferred Reactions

Network Reconstruction

Metabolic Model

Quantitative Analytical Methods

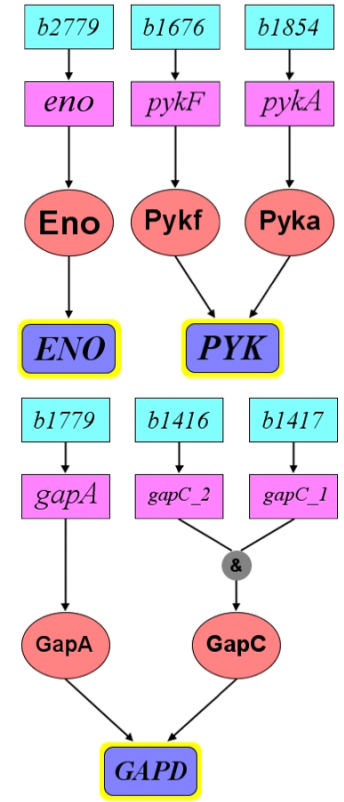
New Predictions
Emergent Properties

Network Assembly and Representation

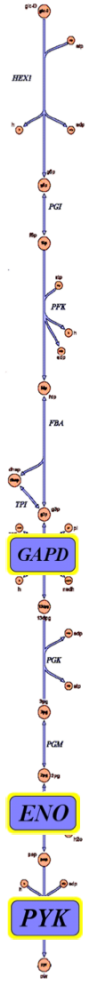
Reconstruction of Glycolytic Pathway

Abbr.	Glycolytic Reactions	Genes
HEX1	[c]glc +atp ⇌ g6p + adp	glk
PGI	[c]g6p ⇌ f6p	pgi
PFK	[c]atp + f6p ⇌ adp + fdp + h	pfkA,pfkB
FBA	[c]fdp ⇌ dhap + g3p	fbaA,fbaB
TPI	[c]dhap ⇌ g3p	tpiA
GAPD	[c]g3p + nad + pi ⇌ 13dpg + h + nadh	gapA,gapC_1,gapC_2
PGK	[c]13dpg + adp ⇌ 3pg + atp	pgk
PGM	[c]3pg ⇌ 2pg	gpmA,gpmB
ENO	[c]2pg ⇌ h2o + pep	eno
PYK	[c]adp + h + pep ⇌ atp + pyr	pykA,pykF

	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
atp	-1	0	-1	0	0	0	1	0	0	1
glc	-1	0	0	0	0	0	0	0	0	0
adp	1	0	1	0	0	0	-1	0	0	-1
g6p	1	-1	0	0	0	0	0	0	0	0
h	1	0	1	0	0	1	0	0	0	-1
f6p	0	1	-1	0	0	0	0	0	0	0
fdp	0	0	1	-1	0	0	0	0	0	0
dhap	0	0	0	1	-1	0	0	0	0	0
g3p	0	0	0	1	1	-1	0	0	0	0
nad	0	0	0	0	0	-1	0	0	0	0
pi	0	0	0	0	0	-1	0	0	0	0
13dpg	0	0	0	0	0	1	-1	0	0	0
nadh	0	0	0	0	0	1	0	0	0	0
3pg	0	0	0	0	0	0	1	-1	0	0
2pg	0	0	0	0	0	0	0	1	-1	0
pep	0	0	0	0	0	0	0	0	1	-1
h2o	0	0	0	0	0	0	0	0	1	0
pyr	0	0	0	0	0	0	0	0	0	1

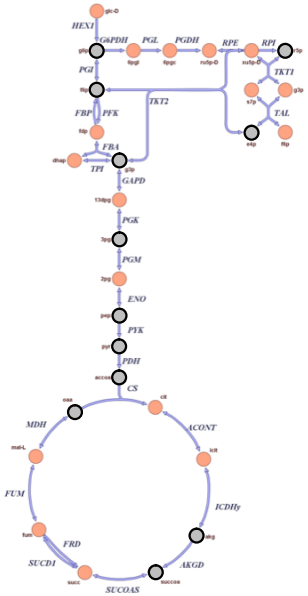


PYK: IF *pykA* OR *pykF*
ENO: IF *eno*
GAPD: IF *gapA* OR (*gapC_1* AND *gapC_2*)



Network Evaluation

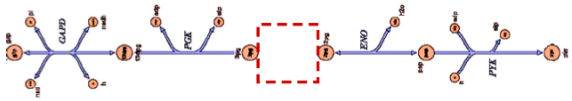
- **Precursor Metabolite Formation**



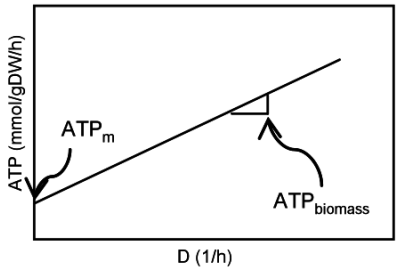
- **Incorporating Biomass Composition**

D = 0.1	%(w/w)
Proteins	
Amino Acids	45.0
Free Amino Acids	1.1
Carbohydrates	
Monosaccharides	-
Disaccharides	
Trehalose	0.8
Oligosaccharides	-
Polysaccharides	
Glycogen	8.4
Mannan	13.1
Other Carbohydrates	18.4
Nucleotides	
RNA	6.3
DNA	0.4
Lipids	2.9
Ash	5.0
Total	101.4

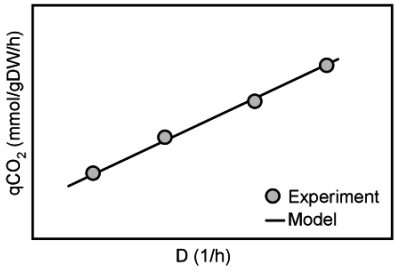
- **Filling Network Gaps**

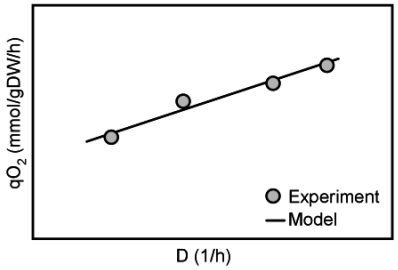


- **ATP Maintenance Calculation**

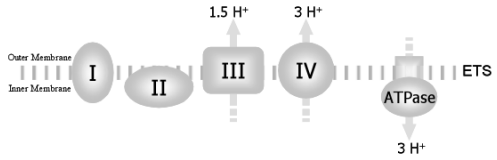


- **Physiological Data Comparison**

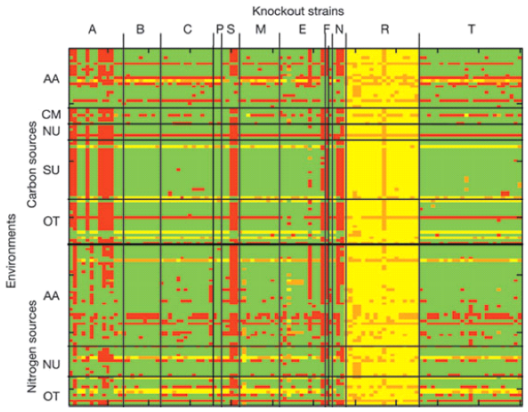




- **P/O Ratio Calculation**

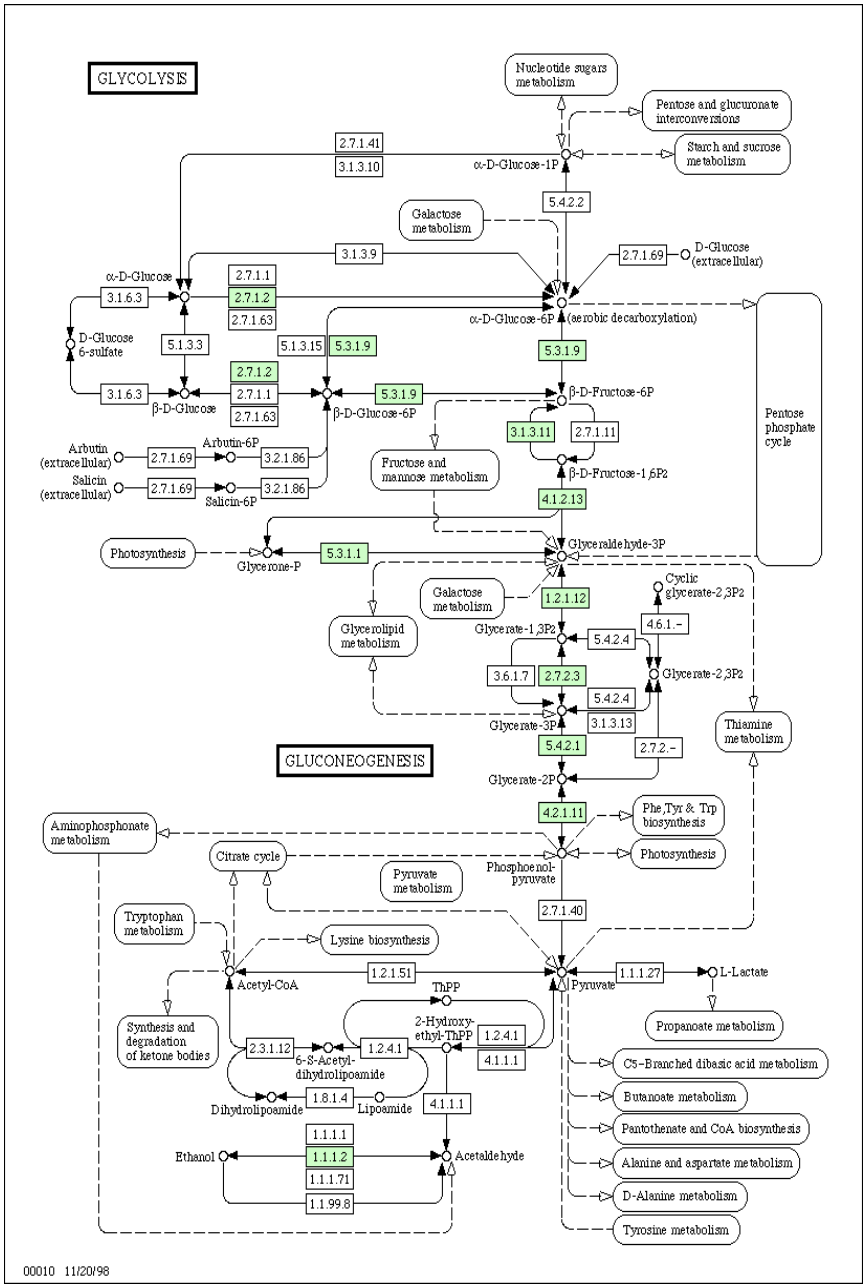


- **Knockout Data Comparison**



III. Physiological Information and Inferred Reactions:

Filling in the Gaps based on indirect evidence



Filling in the Gaps – an Example

- Experiments determine which amino acids are taken up by *H. pylori* vs. which can be produced *in vivo*
- Missing steps of amino acid biosynthesis are added if necessary on the basis of this physiological evidence

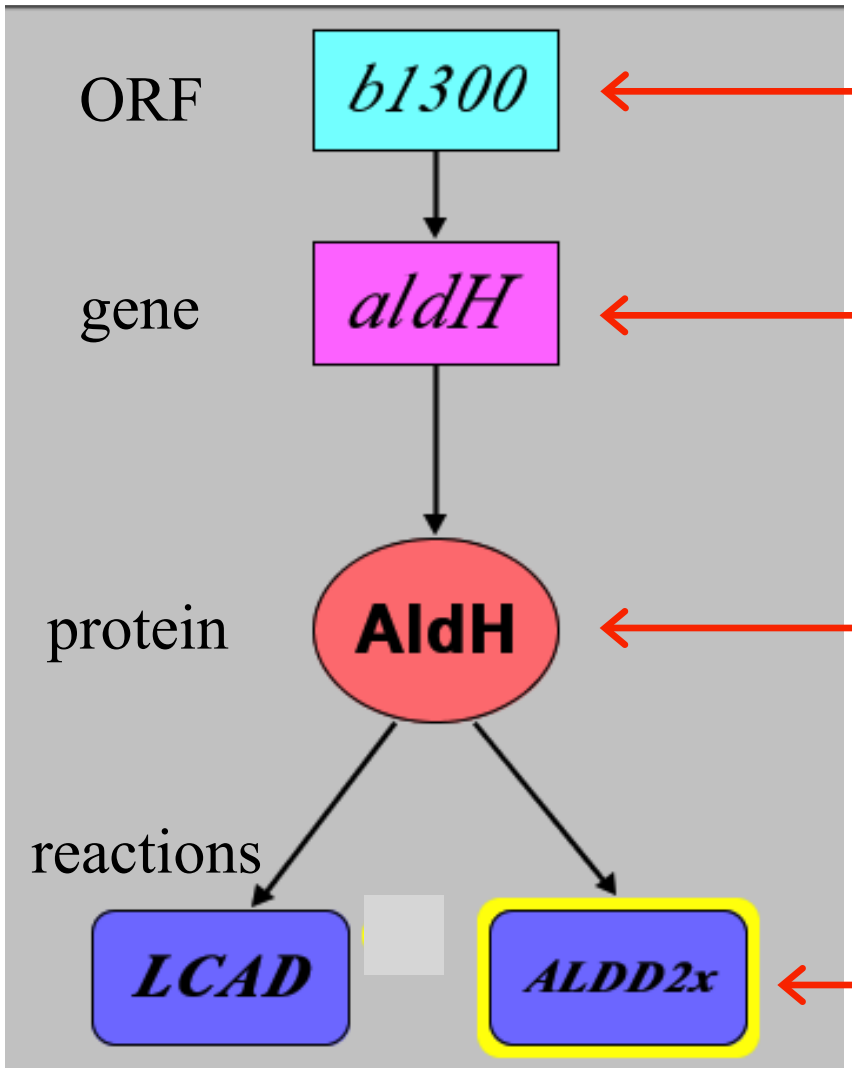
Amino Acid Requirements		
AA	Reynolds	Model
Ala	-	-
Arg	-	-
Asn	+	+
Asp	+	+
Cys	+	+
Gln	+	+
Glu	+	+
Gly	+	+
His	-	-
Ile	-	-
Leu	-	-
Lys	+	+
Met	-	-
Phe	-	-
Pro	+	+
Ser	+	+
Thr	+	+
Trp	+	+
Tyr	+	+
Val	-	-

in vivo in silico♪

Inferred Reactions

- Some reactions are included based on indirect physiological evidence (by inference)
  - Assumption: the cell must be able to produce all biomass components to grow
  - Reactions are added if necessary
  - Generally transporters, etc.
  - Most tentative; should be examined more carefully

Integrating "-omics" Data



Genomics
ORF annotation

Transcriptomics
mRNA levels

Proteomics
protein levels

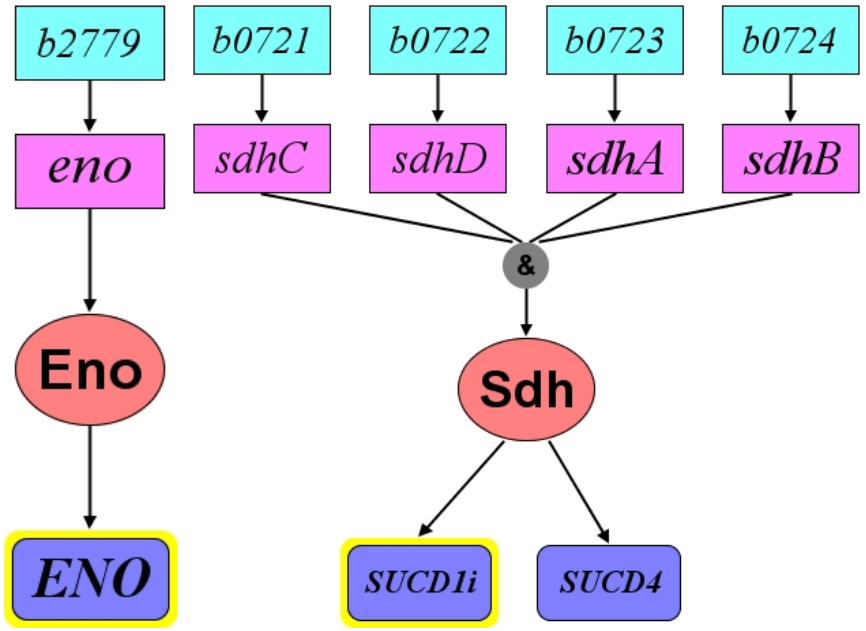
"Fluxomics"
flux measurements

http://gcrg.ucsd.edu/organisms/ecoli_GPR.html

Enolase

1 gene
1 reaction

(342/904 Genes are one-to-one)



Succinate Dehydrogenase

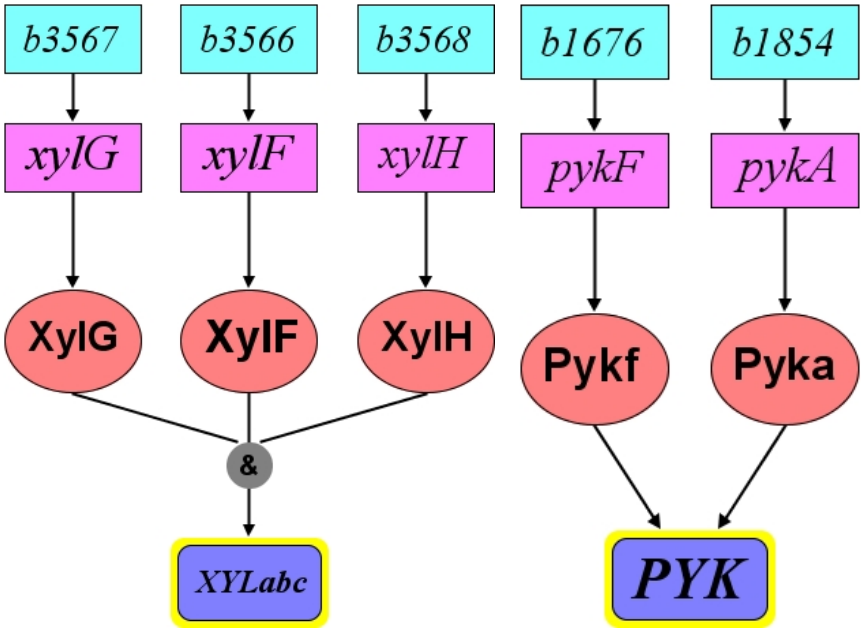
4 subunits
2 reactions

(135/904 Genes interact as Subunits)

Xylose ABC Transporter

Protein complex
(3 proteins)
1 reaction

(153/904 Genes interact as PCs)

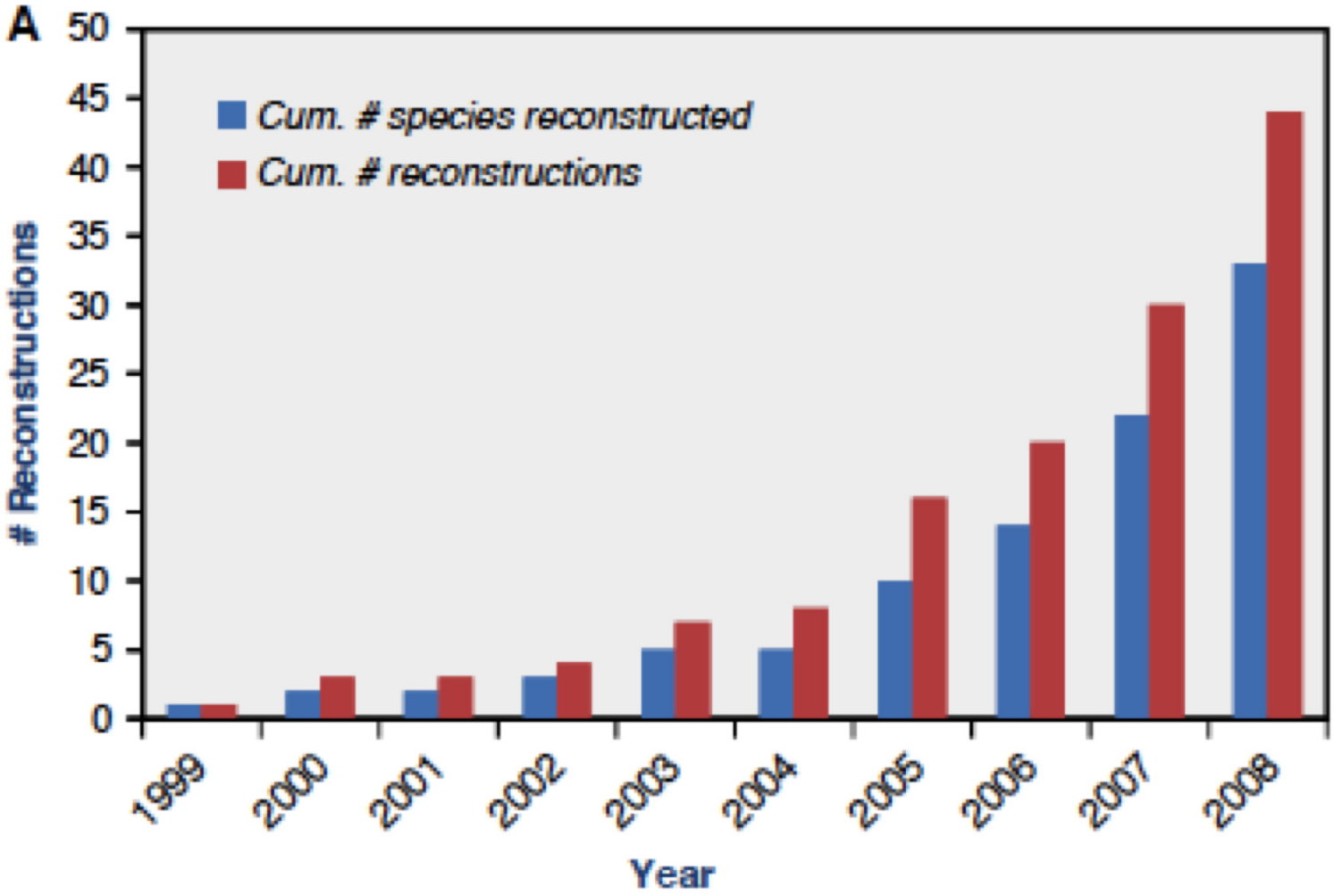


Pyruvate Kinase

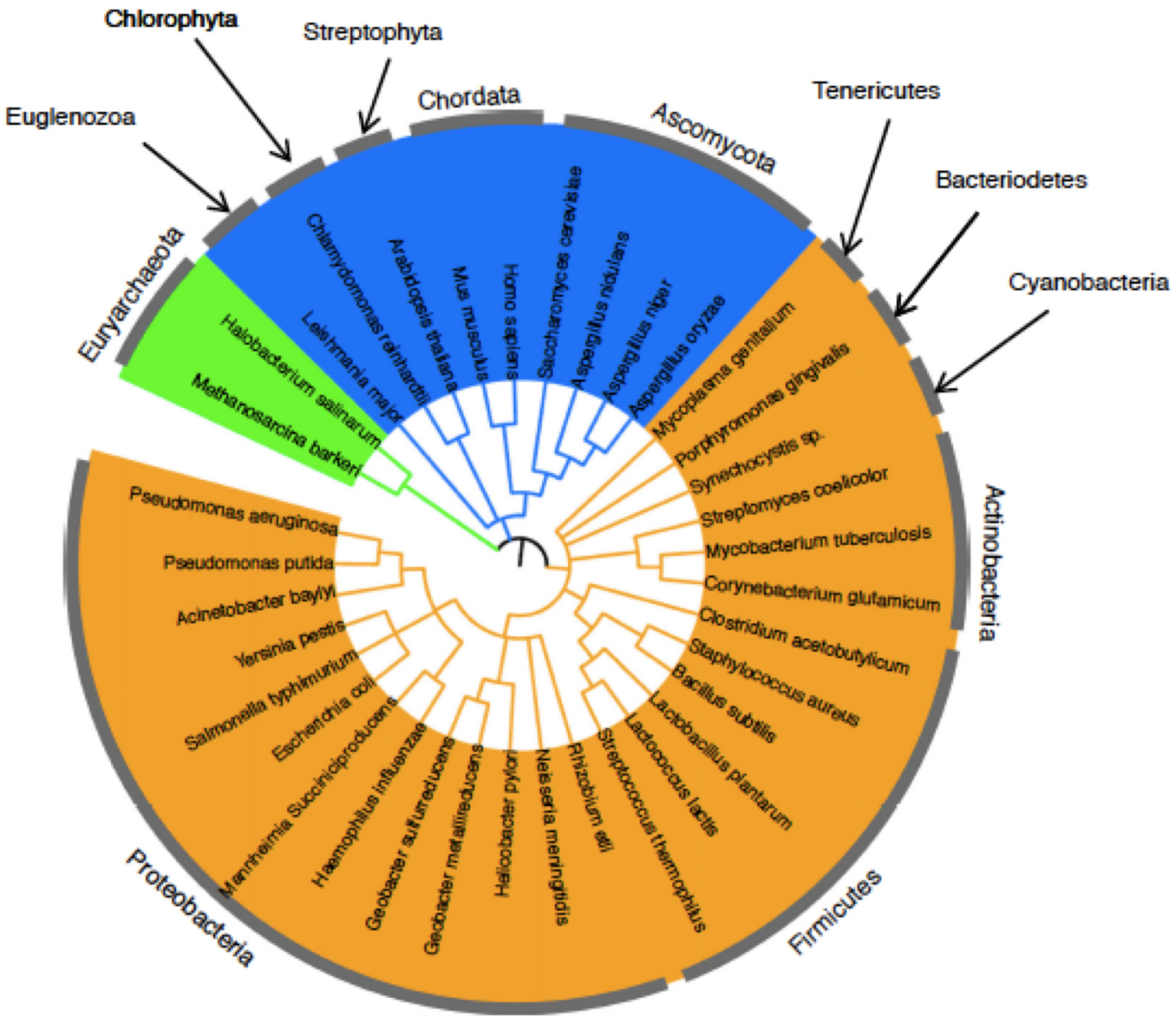
2 isozymes
1 reaction

(149/931 Reactions have multiple isozymes)

Availability of Metabolic Reconstructions

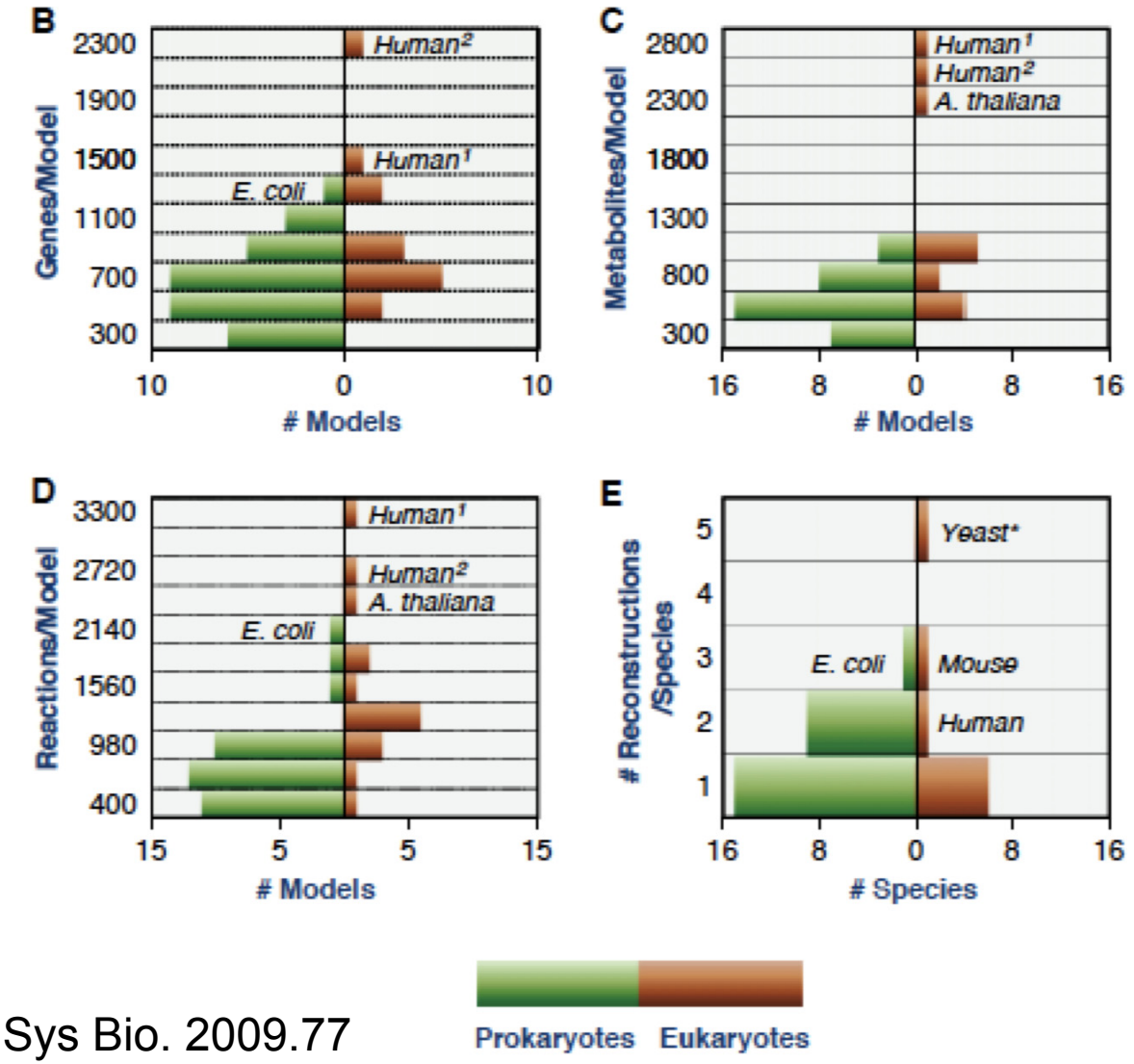


Oberhardt et al. Mol Sys Bio. 2009.77



Oberhardt et al. Mol Sys Bio. 2009.77

Size and Content of Available Reconstructions

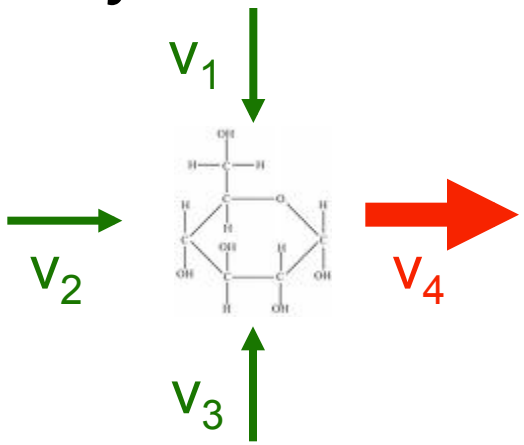


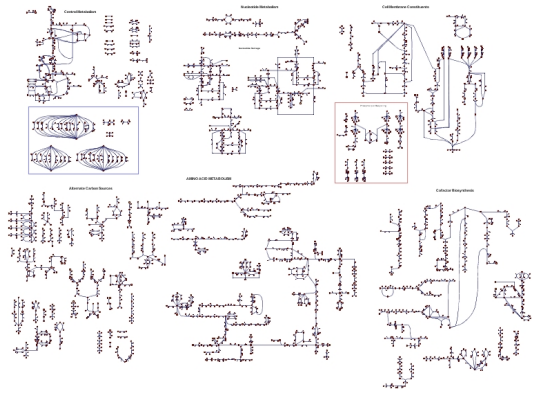
Oberhardt et al. Mol Sys Bio. 2009.77

Constraint-Based Models

Constraints on Metabolic Networks

1. *Steady-State Mass Balance Constraints*





For each metabolite:

$\sum s_{ij} \cdot v_{produce} = \sum -s_{ij} \cdot v_{consume}$

For all metabolites:

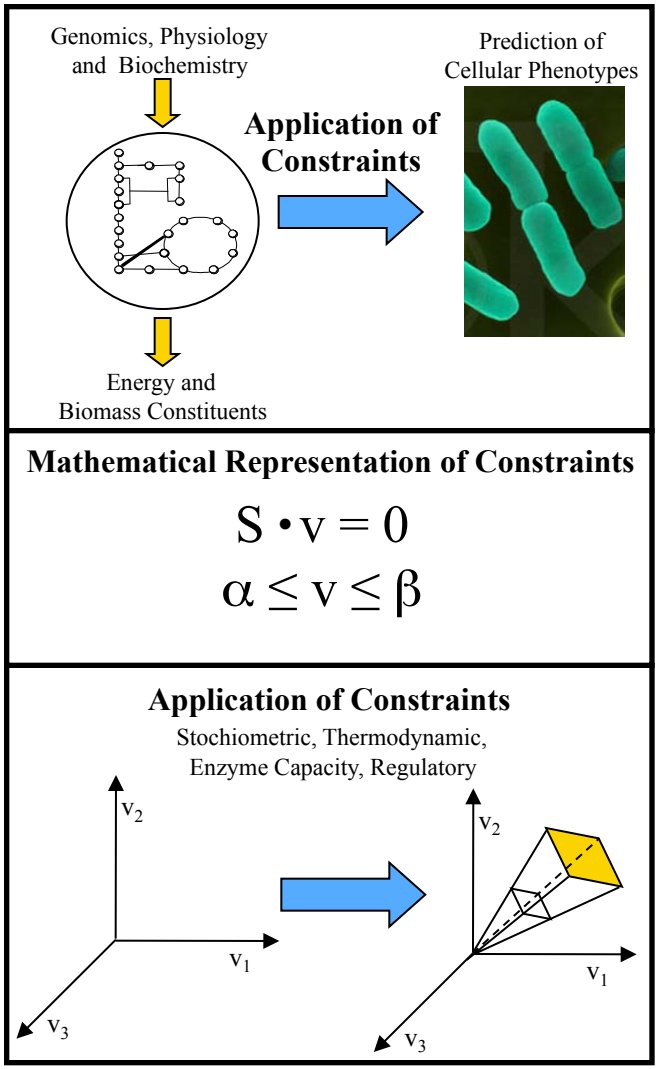
$\mathbf{S} \cdot v = 0$

2. *Enzyme Capacity Constraints:* $\alpha \leq v_j \leq \beta$
3. *Thermodynamic Constraints:* $v_j \geq 0$
4. *Regulatory Constraints*: $\alpha, \beta = 0$ if associated genes are un-expressed

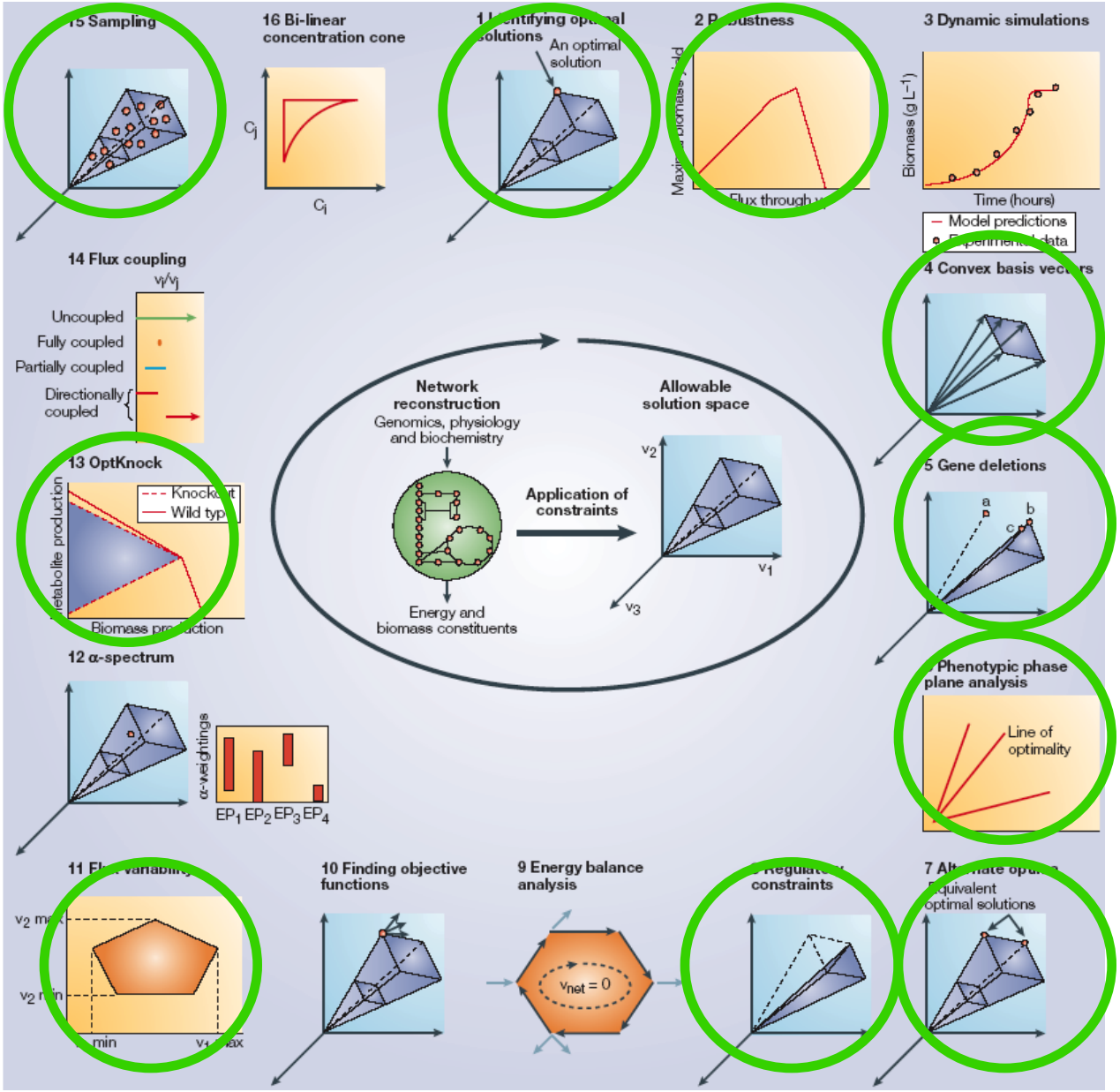
Constraint-Based Analysis

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

–Sherlock Holmes, A Study in Scarlet



Constraint-Based Methods



Price, Reed, and Palsson Nat. Reviews Microbiol. 2004

- Optimal Solutions
  1. FBA
  2. Flux Variability
- Flux Dependencies
  1. Robustness
  2. Phase Planes
  3. Flux Coupling
- All Allowable Solutions
  1. Extreme Pathways
  2. Elementary Modes
  3. Sampling
- Altering Phenotypes
  1. Genetic Mutations
  2. Strain Design
- Application of Additional Constraints
  1. Regulation
  2. Energy Balance

Flux Balance Analysis

FBA Optimization Problem Statement

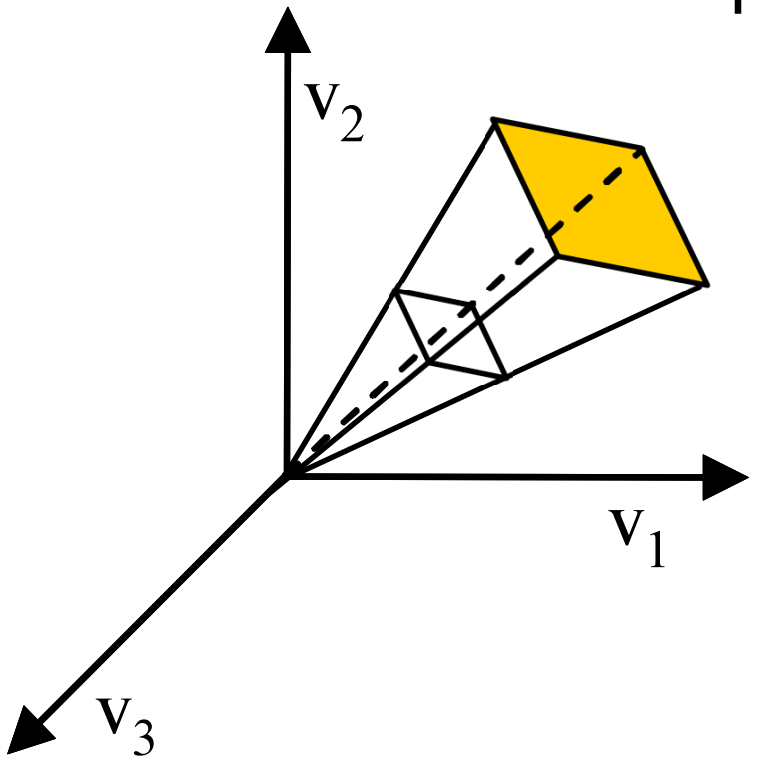
- Objective Function: A function that is maximized or minimized to identify optimal solutions

- Constraints: Place limits on the allowable values the solutions can take on.

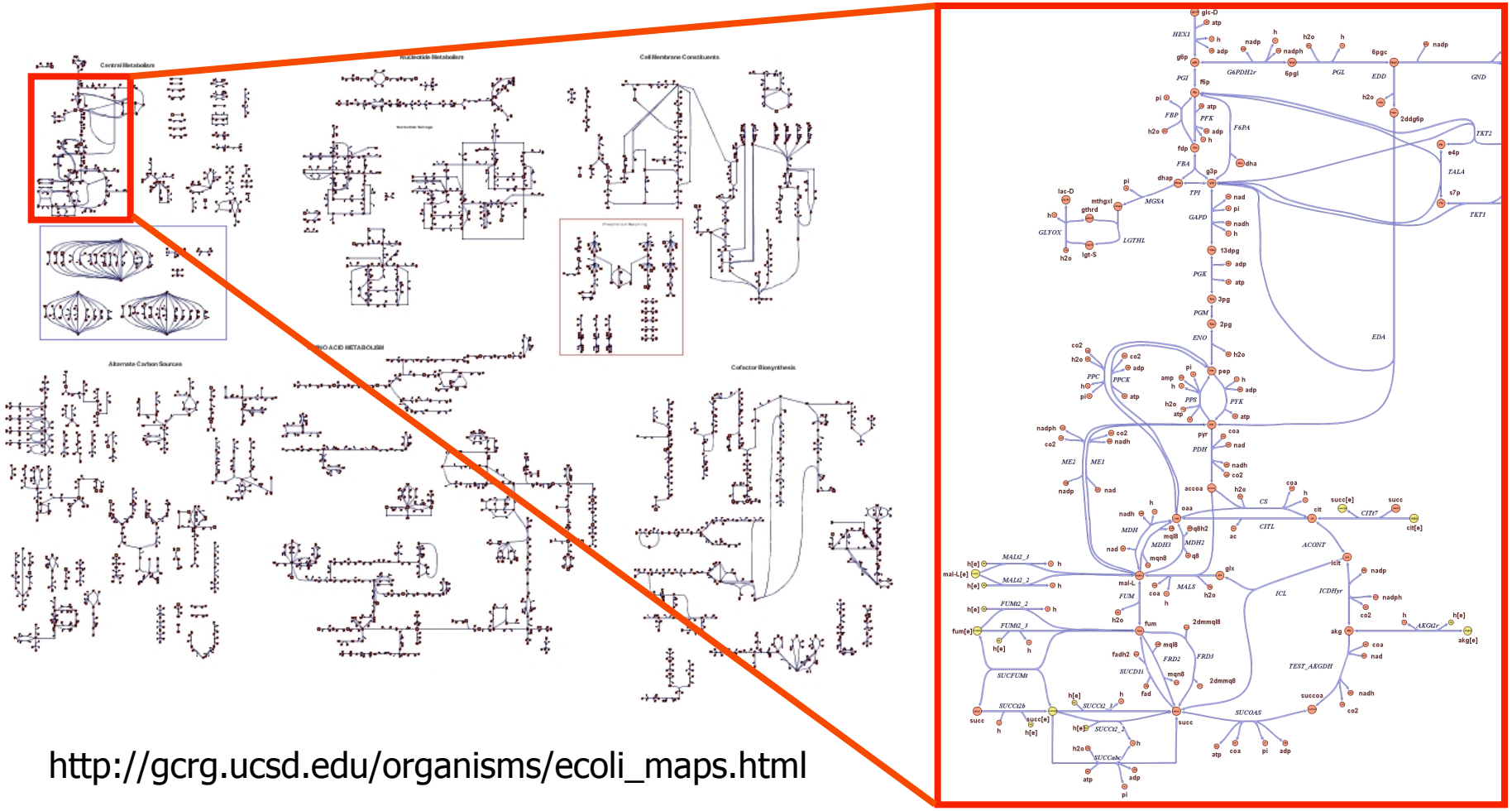
Maximize: $c \cdot v$

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

$\alpha \leq v \leq \beta$



Escherichia coli Metabolism

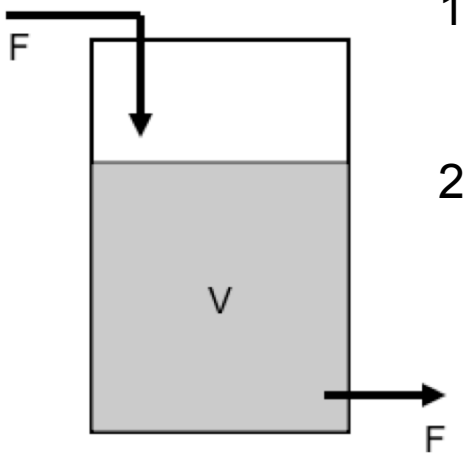


http://gcrg.ucsd.edu/organisms/ecoli_maps.html

Growth (GAR) and Non-Growth (NGAR) Associated ATP Requirements

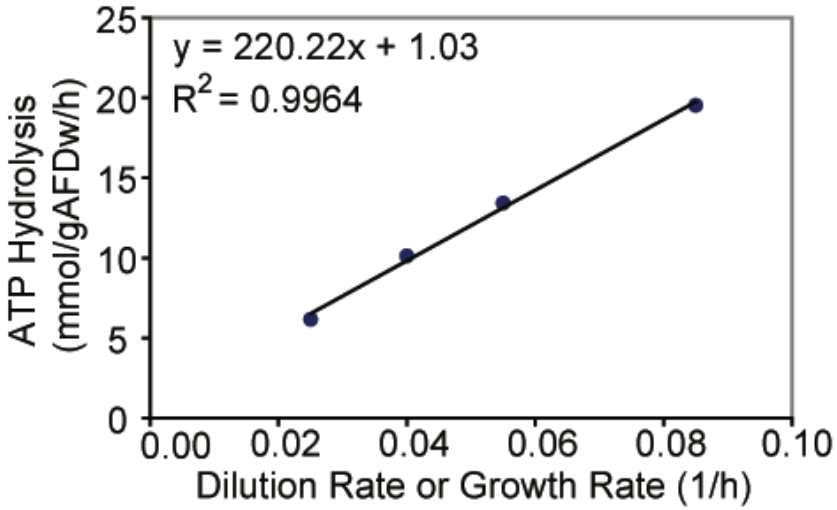
In Chemostat:

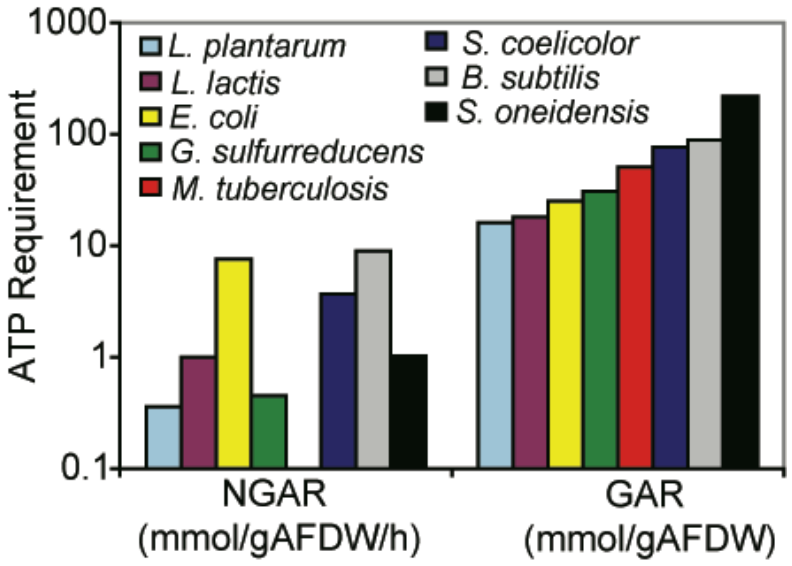
D = F/V = cell growth rate @ SS



1. Fix growth rate and substrate uptake rate
2. Maximize the amount of excess ATP that can be made (i.e. hydrolysis of ATP)

S. oneidensis MR-1



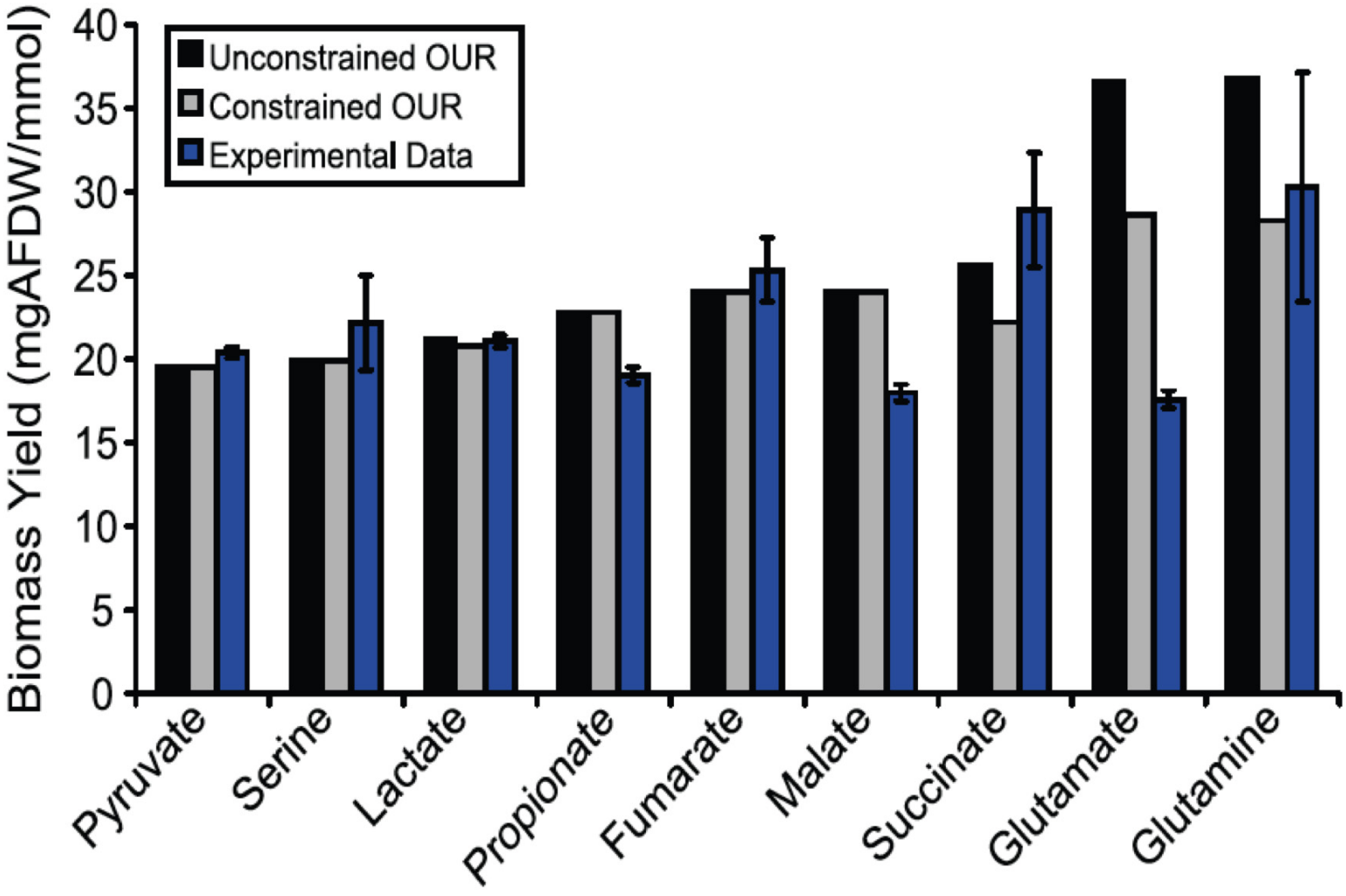


GAR = 220.2
(mmol ATP /gAFDW)

NGAR = 1.03
(mmol ATP /gAFDW/hr)

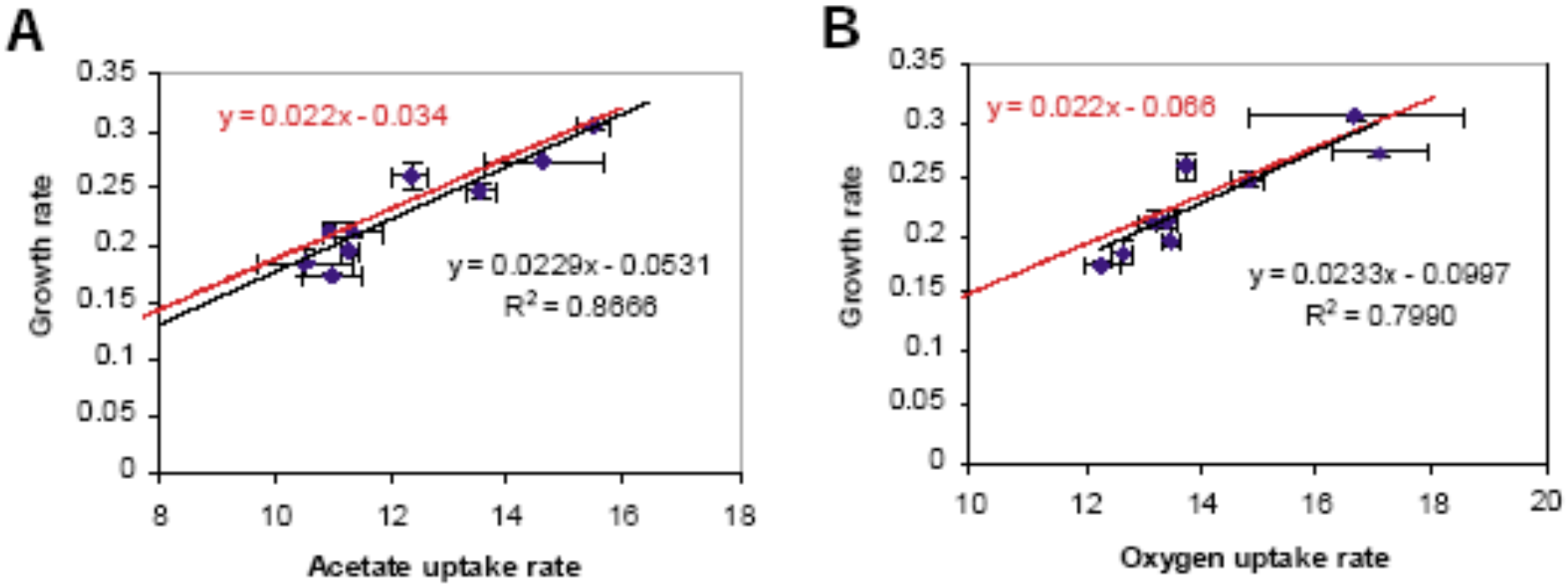
Pinchuk et al. PLoS Comp Biol. 6(6) (2010)

Predicted vs. Experimental Biomass Yields (*Shewanella oneidensis*)



Pinchuk et al. PLoS Comp Biol. 6(6) (2010)

FBA Predicted Maximal Growth Rates vs. Experimental Growth Rates in Batch Culture



Edwards, Ibarra, Palsson. Nature Biotechnology. 19:125-130 (2001).

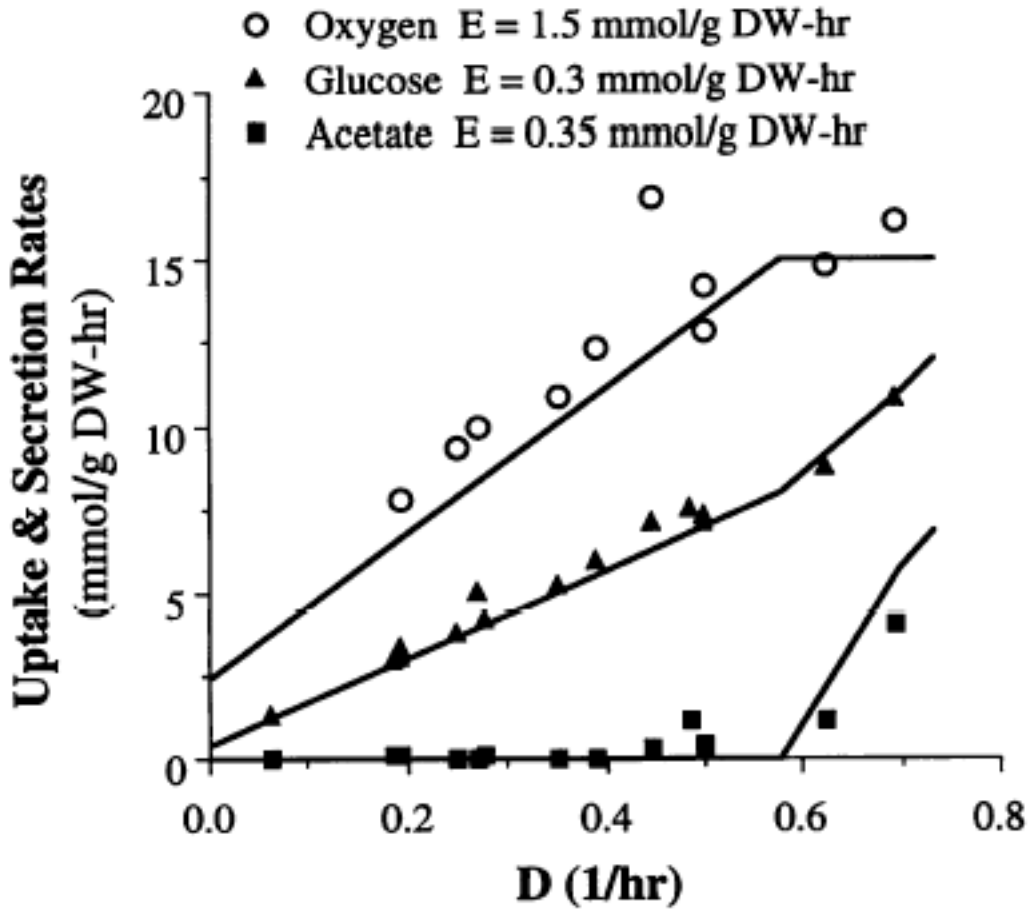
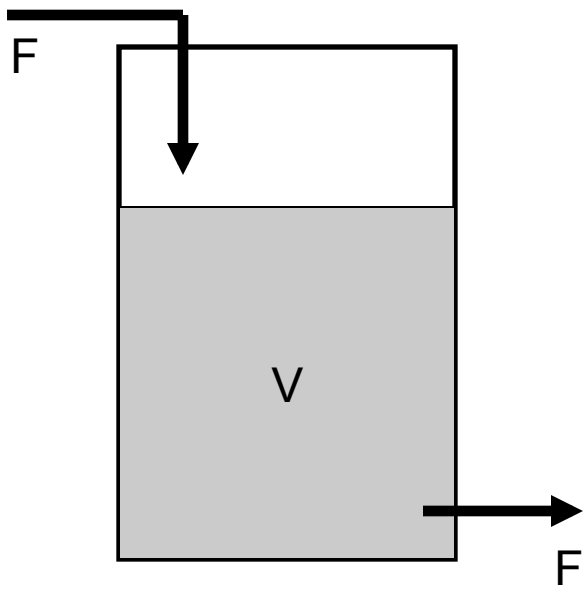


FIG. 6. Analysis of aerobic chemostat culture showing the glucose and oxygen uptake rates and the acetate secretion rate as functions of the dilution or growth rate. The chemostat was not limited for minerals. The solid lines represent the flux balance model simulations. E, average deviation between predictions of the model and experimental measurements; DW, dry weight.

In Chemostat:

D = F/V = cell growth rate @ SS



Varma and Palsson, App. Environ. Microbiol. 60(10): 3723-3731 (1994)

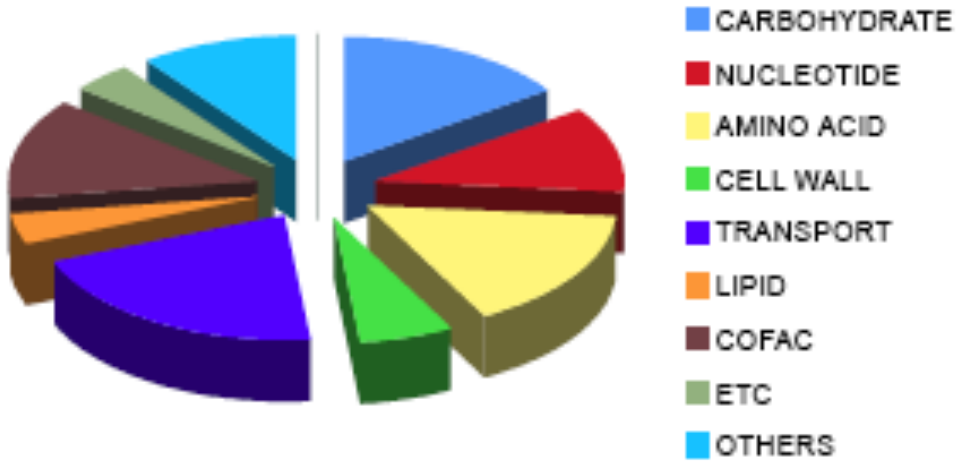
Approach:

1. Genome Comparison
   - MG1655 vs LT2
   - Reciprocal best hits
   - Identity > 70%
2. Draft Reconstruction
   - Fill in Gaps
   - Organism Specific Pathways
   - Biomass Components
3. Generate Model & Compare Predicted Growth Phenotypes w/ Experimental Data
   - Carbon Sources
   - Gene deletions
4. Refine Metabolic Reconstruction

S. typhimurium Genome Size	4,857,432 bp
Open Reading Frames	4553

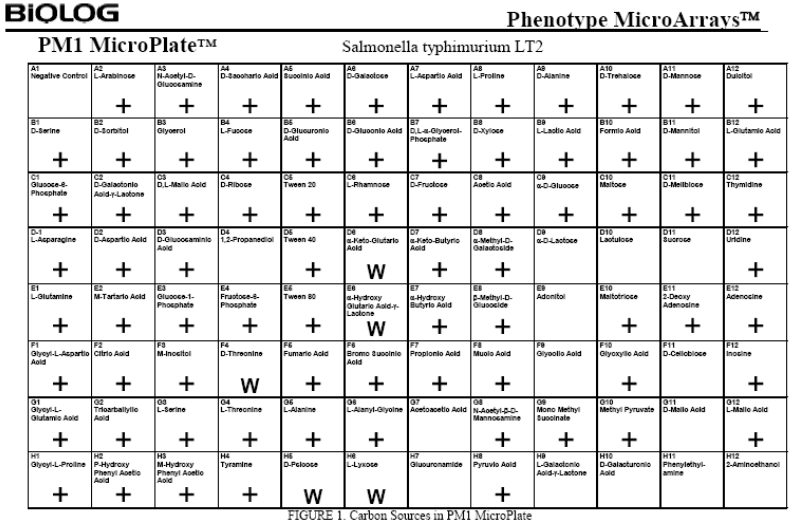
*i*RR1083 *in silico S. typhimurium* characteristics

Genes	1083
Proteins	973
Reactions	1087
Gene associated	1018
Non-Gene associated	69
Intracellular Metabolites	744



A. Raghunathan, et al. BMC Systems Biology, 3:38 (2009).

Qualitative Growth Phenotypes



BIOLOG — Phenotype MicroArrays™

PM1 MicroPlate™ — Salmonella typhimurium LT2

A1 Negative Control	A2 L-Arabinose +	A3 N-Acetyl-D-Glucosamine +	A4 D-Saccharic Acid +	A5 Succinic Acid +	A6 D-Galactose +	A7 L-Aspartic Acid +	A8 L-Proline +	A9 D-Alanine +	A10 D-Trehalose +	A11 D-Mannose +	A12 Dulcitol +
B1 D-Serine +	B2 D-Sorbitol +	B3 Glycerol +	B4 L-Fucose +	B5 D-Glucuronic Acid +	B6 D-Gluconic Acid +	B7 D,L-α-Glycerol-Phosphate +	B8 D-Xylose +	B9 L-Lactic Acid +	B10 Formic Acid +	B11 D-Mannitol +	B12 L-Glutamic Acid +
C1 Glucose-6-Phosphate +	C2 D-Galactonic Acid-γ-Lactone +	C3 D,L-Malic Acid +	C4 D-Ribose +	C5 Tween 20 +	C6 L-Rhamnose +	C7 D-Fructose +	C8 Acetic Acid +	C9 α-D-Glucose +	C10 Maltose +	C11 D-Melibiose +	C12 Thymidine +
D1 L-Asparagine +	D2 D-Aspartic Acid +	D3 D-Glucosaminic Acid +	D4 1,2-Propanediol +	D5 Tween 40 +	D6 α-Keto-Glutaric Acid W	D7 α-Keto-Butyric Acid +	D8 α-Methyl-D-Galactoside +	D9 α-D-Lactose	D10 Lactulose +	D11 Sucrose +	D12 Uridine +
E1 L-Glutamine +	E2 M-Tartaric Acid +	E3 Glucose-1-Phosphate +	E4 Fructose-6-Phosphate +	E5 Tween 80 +	E6 α-Hydroxy Glutaric Acid-γ-Lactone W	E7 α-Hydroxy Butyric Acid +	E8 β-Methyl-D-Glucoside +	E9 Adonitol	E10 Maltotriose +	E11 2-Deoxy Adenosine +	E12 Adenosine +
F1 Glycyl-L-Aspartic Acid +	F2 Citric Acid +	F3 M-Inositol +	F4 D-Threonine W	F5 Fumaric Acid +	F6 Bromo Succinic Acid +	F7 Propionic Acid +	F8 Mucic Acid +	F9 Glycolic Acid	F10 Glyoxylic Acid +	F11 D-Cellobiose	F12 Inosine +
G1 Glycyl-L-Glutamic Acid +	G2 Tricarballylic Acid +	G3 L-Serine +	G4 L-Threonine +	G5 L-Alanine +	G6 L-Alanyl-Glycine +	G7 Acetoacetic Acid	G8 N-Acetyl-β-D-Mannosamine +	G9 Mono Methyl Succinate +	G10 Methyl Pyruvate +	G11 D-Malic Acid	G12 L-Malic Acid +
H1 Glycyl-L-Proline +	H2 P-Hydroxy Phenyl Acetic Acid +	H3 M-Hydroxy Phenyl Acetic Acid +	H4 Tyramine +	H5 D-Psicose W	H6 L-Lyxose W	H7 Glucuronamide	H8 Pyruvic Acid +	H9 L-Galactonic Acid-γ-Lactone	H10 D-Galacturonic Acid	H11 Phenylethyl-amine	H12 2-Aminoethanol

FIGURE 1. Carbon Sources in PM1 MicroPlate

JOURNAL OF BACTERIOLOGY, Oct. 1969, p. 215–219 Vol. 100, No. 1

Compounds Which Serve as the Sole Source of Carbon or Nitrogen for *Salmonella typhimurium* LT-2

DAVID GUTNICK,[1] JOSEPH M. CALVO, TADEUSZ KLOPOTOWSKI, AND BRUCE N. AMES

National Institutes of Health, Bethesda, Maryland 20014, Department of Biochemistry and Molecular Biology, Cornell University, Ithaca, New York 14850, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw 12, Poland, and Department of Biochemistry, University of California, Berkeley, California 94720

Received for publication 7 August 1969

About 600 compounds were screened as possible carbon or nitrogen sources for *Salmonella typhimurium* LT-2. About 100 utilizable compounds were found.

Model Predictions	**Exp. Growth**	**Exp. No Growth**	**No Data**	**Total**
Growth (Carbon)	75	21 (FP)	17	113
No Growth (Carbon)	1 (FN)	28	21	50
Growth (Nitrogen)	37	5 (FP)	8	50
No Growth (Nitrogen)	9 (FN)	23	16	48

- **Overall Accuracy = 82% and Untested = 24%**
- **False Positives (missing regulation) > False Negatives (missing reactions)**
- **13 of the 21 false positive carbon sources can serve as nitrogen sources**

A. Raghunathan, et al. BMC Systems Biology, 3:38 (2009).

Interpretting the Dual: Reduced Costs & Shadow Prices

Value	Shadow Price	Reduced Cost
Positive	Removing metabolite increases objective.	Increasing flux will increase objective. (usually occurs when fluxes are at their '.up' bounds).
Negative	Adding metabolite increases objective	Increasing flux will decrease objective.
Zero	Adding/removing metabolite does not change objective	Changing flux will not change objective

Maximal Production of Metabolites Under Glucose Aerobic Conditions

TABLE 2

Maximum stoichiometric yields of biosynthetic precursors on glucose for an aerobic non-growing cell

Metabolite	Yield	Carbon conversion	ATP shadow price	Constraint
3PG	2	100%	0	None
PEP	2	100%	0	
Pyr	2	100%	0	
OA	2	133·3%	0	
G6P	0·908	90·8%	0·046	Energy
F6P	0·908	90·8%	0·046	
R5P	1·08	90%	0·055	
E4P	1·33	88·7%	0·068	
T3P	1·73	86·5%	0·088	
AcCoA	2	66·7%	0	Stoichiometry
αKG	1	83·3%	0	
SuccCoA	1	66·7%	0	

Table shadow prices are multiplied by -1 to be consistent with our definition of shadow prices

Varma and Palsson, J Theoretical Biology. 165:477-502 (1993)

Evaluate Shadow Prices for Model Corrections

- You can use shadow prices and reduced costs to evaluate your results.

- For example: If you maximize growth rate and find zero growth, you can identify metabolites which are needed in order to grow (those with a negative shadow price).
  - This is useful if you are debugging a network.

FBA Using GAMS

Metabolic Network Example

Reaction List

v_1: A → B

v_2: A → C

b_1: → A

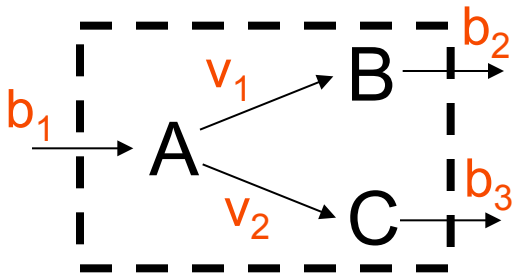
b_2: B →

b_3: C →

S Matrix

	v_1	v_2	b_1	b_2	b_3
A	-1	-1	1	0	0
B	1	0	0	-1	0
C	0	1	0	0	-1

Metabolic Map



Maximize $Z = c \cdot v = v_1$

Such that $S \cdot v = 0$

$0 \leq v \leq 10$

```
sets
metabolites /A,B,C,D/
reactions /v1,v2,b1,b2,b3/;

parameters
c(reactions)
S(metabolites,reactions)
/A.v1 -1
B.v1 1
A.v2 -1
C.v2 1
A.b1 1
B.b2 -1
C.b3 -1/;

variables
v(reactions)
Z;

equations
massbalance(metabolites)
objectivefunction;

massbalance(metabolites).. sum(reactions, S(metabolites,reactions)*v(reactions))=e=0;
objectivefunction.. Z=e=sum(reactions,c(reactions)*v(reactions));

v.lo(reactions)=0;
v.up(reactions)=10;

c('v1')=1;

model FBA /all/;
solve FBA using lp maximizing Z;
```

Define S

Define Equations (Z=c·v and S·v=0)

Apply Variable Bounds: Upper (.up)
Lower (.lo)

Define C, ie. pick flux to optimize

Constraint and Variable Values

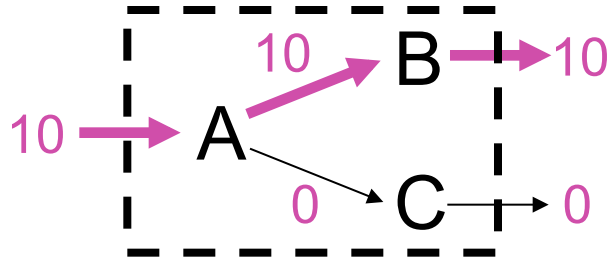
```
---- EQU massbalance

     LOWER      LEVEL      UPPER     MARGINAL

A      .          .          .        -1.000
B      .          .          .          EPS
C      .          .          .          EPS

                          LOWER      LEVEL      UPPER     MARGINAL

---- EQU objective~         .          .          .         1.000
```

S·v = level

Z-c·v = level

```
---- VAR v

      LOWER      LEVEL      UPPER     MARGINAL

v1     .         10.000     10.000       .
v2     .           .        10.000     -1.000
b1     .         10.000     10.000      1.000
b2     .         10.000     10.000       .
b3     .           .        10.000       .

                          LOWER      LEVEL      UPPER     MARGINAL

---- VAR Z                 -INF      10.000      +INF         .
```

Flux Map



Review of Shadow Prices & Reduced Costs

- Shadow Prices (SP):
  - One for each constraint or metabolite
  - dZ/db_i
  - SP<0 means adding metabolite (ie. change b=0 to b<0) would increase Z.
  - SP>0 means removing metabolite (ie. change b=0 to b>0) would increase Z.
- Reduced Costs (RC):
  - One for each variable or flux.
  - dZ/dv_j (for zero fluxes)
  - RC < 0 means increasing flux (v_j) would reduce Z.

Shadow Prices (1 per metabolite)

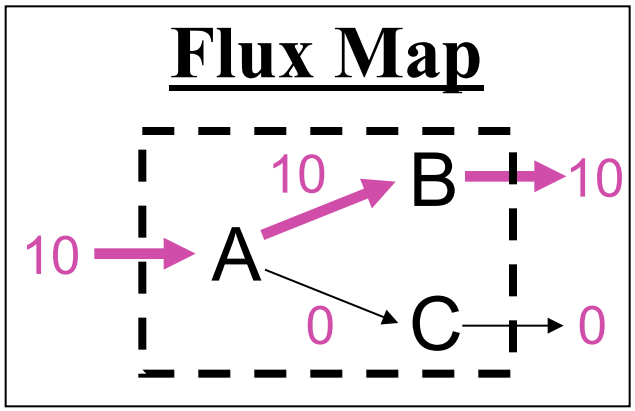
$SP_A = -1$ →

•If we change b_A from zero to 1: we are saying the production of A has to be higher than the consumption of A by 1 unit (remember S·v=production-consumption).

•A lower consumption of A means that the flux through v1 will have to go down by 1 unit. Hence, $dZ/db_A = -1$.

•For example, if $b_A = 1$ then Z = 9.

$SP_C = SP_B$ = EPS (~0) → This is because if you added B or C to the network they wouldn't allow for higher flux through v_1.



---- EQU massbalance

	LOWER	LEVEL	UPPER	MARGINAL
A	.	.	.	-1.000
B	.	.	.	EPS
C	.	.	.	EPS

Central Metabolic Network (ie. CoreTextbookModel.gms)

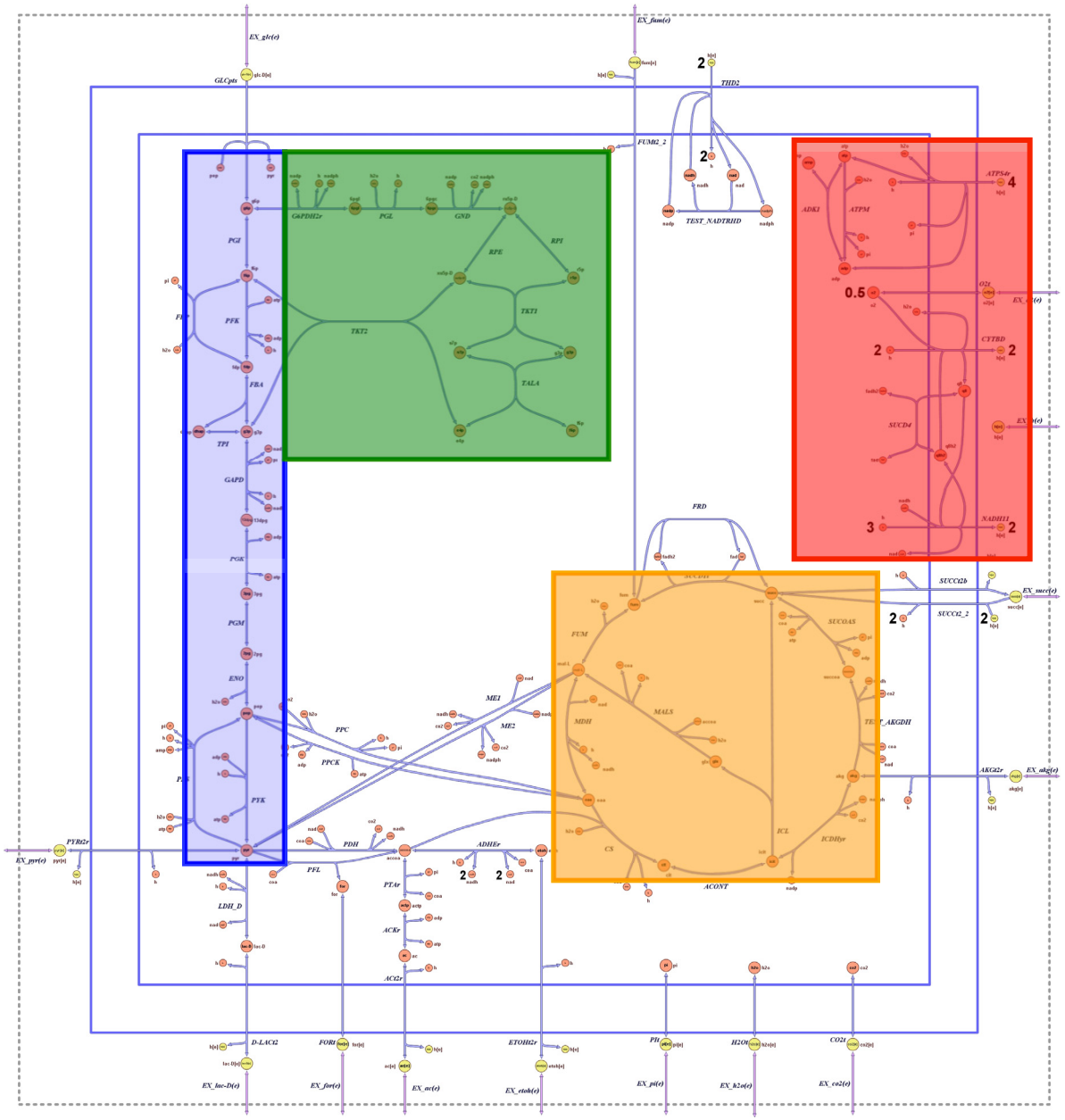
Core E. coli Metabolic Network

Glycolysis

Pentose Phosphate Pathway

TCA Cycle

Oxidative Phosphorylation



101 Genes
63 Metabolites
62 Reactions

Common Flux Abbreviations

- Biomass: this is a drain of biomass components in their appropriate ratios.
- Exchange fluxes:
  - (+) values secretion
  - (-) values means uptake

Exchange Flux	Metabolite
EX_glc_e	Glucose
EX_ac_e	Acetate
EX_succ_e	Succinate
EX_for_e	Formate
EX_etoh_e	Ethanol
EX_o2_e	Oxygen

Running FBA

```
**THIS CODE WAS WRITTEN FOR CBE 782 J.REED (3/2011)
$onecho > cplex.opt
eprhs 0
epopt 0
epint 0
$offecho

*Read in the appropriate S matrix
$include EcoliCoreTextbookModel.gms

*Place limits on the exchange fluxes based on the minimal media for a negative flux through the exchange reactions implies that
*the metabolites are being taken up or consumed by the cell. By default the upperlimits of the exchange fluxes are all set to
*the Vmax, indicating that the cell can secrete any of the extracellular metabolites.

UpperLimits(j)=Vmax;
LowerLimits(exch)=0;

*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_o2_e')=0;

Parameter
c(j) used to define the objective function for FBA /Biomass 1/;
```

Tolerances

Define Carbon Source and Uptake Rate

Define LowerLimits on Oxygen Uptake Rate

Define Objective Function

Running FBA

```
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 FBA /massbalance, calcobj/;
FBA.optfile=1;

v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);

solve FBA using lp maximizing Obj;
```

Mass Balance: $S \bullet v = 0$

Objective: $Obj = C \bullet v$

Flux Limits: $\alpha \leq v \leq \beta$

GAMS Results: LST File

Non-Zero Fluxes

```
----    376 VARIABLE v.L  flux values through reaction in network

ACONT      1.917,    AKGDH      1.389,    ATPS4r    15.198,    Biomass    0.490,    CO2t      -9.160,    CS         1.917
CYTBD     17.188,    ENO        6.798,    EX_co2_e   9.160,    EX_glc_e  -5.000,    EX_h2o_e  10.057,    EX_h_e     5.181
EX_o2_e   -8.594,    EX_pi_e   -1.802,    FBA        3.345,    FUM        1.389,    G6PDH2r    3.504,    GAPD       7.530
GLCpts     5.000,    GND        3.504,    H2Ot     -10.057,    ICDHyr     1.917,    MDH        1.389,    NADH11    15.798
O2t        8.594,    PDH        3.753,    PFK        3.345,    PGI        1.396,    PGK       -7.530,    PGL        3.504
PGM       -6.798,    PIt       -1.802,    PPC        1.403,    PYK        0.140,    RPE        1.984,    RPI       -1.520
SUCD1i     1.389,    SUCD4      1.389,    SUCOAS    -1.389,    TALA       1.080,    TKT1       1.080,    TKT2       0.904
TPI        3.345
```

Growth Rate

```
----    376 VARIABLE Obj.L                 =        0.490  this
                                                          s
```

ive function for the FBA solution

Non-Zero Shadow Prices

```
----    376 EQUATION massbalance.M  mass balance equations for each metabolite

13dpg_c  -0.049,    2pg_c    -0.044,    3pg_c    -0.044,    6pgc_c   -0.098,    6pgl_c   -0.096,    ac_c     -0.025
ac_e     -0.024,    accoa_c  -0.030,    actp_c   -0.030,    adp_c     0.005,    akg_c    -0.068,    akg_e    -0.066
amp_c     0.009,    cit_c    -0.078,    co2_c       EPS,    co2_e       EPS,    dhap_c   -0.055,    e4p_c    -0.072
etoh_c   -0.043,    etoh_e   -0.041,    f6p_c    -0.104,    fadh2_c  -0.003,    fdp_c    -0.110,    for_c       EPS
for_e       EPS,    fum_c    -0.052,    fum_e    -0.049,    g3p_c    -0.055,    g6p_c    -0.104,    glc_D_e  -0.098
glx_c    -0.022,    h2o_c       EPS,    h2o_e       EPS,    h_c       0.002,    h_e         EPS,    icit_c   -0.078
lac_D_c  -0.044,    lac_D_e  -0.043,    mal_L_c  -0.052,    nad_c     0.008,    nadph_c  -0.010,    o2_c        EPS
o2_e        EPS,    oaa_c    -0.046,    pep_c    -0.044,    pi_c        EPS,    pi_e        EPS,    pyr_c    -0.038
pyr_e    -0.036,    q8h2_c   -0.003,    r5p_c    -0.088,    ru5p_D_c -0.088,    s7p_c    -0.121,    succ_c   -0.055
succ_e   -0.052,    succoa_c -0.060,    xu5p_D_c -0.088
```

FBA Calculations: Using Glucose

1. What is the maximum growth rate for glucose aerobic growth (max. glucose uptake rate of 5)?

2. What is the maximum growth rate for glucose anaerobic (no oxygen uptake) growth (max. glucose uptake rate of 5)?

3. What are the by-products that are secreted during maximal glucose anaerobic growth?

4. What are the aerobic and anaerobic biomass yields (gDW/g glucose)? Hint: Your flux units are mmol/gDW/h for exchanges and 1/h for biomass.

5. What is the molar yield for ethanol under anaerobic conditions (mmol ethanol/mmol glucose)

FBA Calculations: Using Glucose

1. What is the maximum growth rate for glucose aerobic growth (max. glucose uptake rate of 5)?
   - **0.49 1/hr**
2. What is the maximum growth rate for glucose anaerobic (no oxygen uptake) growth (max. glucose uptake rate of 5)?
   - **0.20 1/hr**
3. What are the by-products that are secreted during glucose anaerobic growth?
   - **acetate, ethanol, formate**
4. What are the aerobic and anaerobic biomass yields (gDW/g glucose)? Hint: Your flux units are mmol/gDW/h for exchanges and 1/h for biomass.
   - **Aerobic = (0.49 1/h)/(5 mmol glc/gDW-h)/(0.180 g glc/mmol glc)**
   - **Aerobic = 0.54 gDW/g glucose**
   - **Anaerobic = (0.20 1/h)/(5 mmol glc/gDW-h)/(0.180 g glc/ mmol glc)**
   - **Anaerobic = 0.22 gDW/g glucose**
5. What is the molar yield for ethanol under anaerobic conditions (mmol ethanol/mmol glucose)
   - **Ethanol Yield = (EX_etoh_e)/(EX_glc_e) = 3.4/5**
   - **Yield = 0.68 mmol ethanol/mmol glucose**

FBA Calculations: Other Conditions

6. Can *E. coli* grow anaerobically on acetate ? *(hint: to get a feasible solution set lowerlimit to -5 for EX_ac_e and upperlimit to 0)*

7. Looking at the shadow prices for oxygen (o2) do you think that the cells could grow with acetate aerobically?

8. Looking at the reduced costs for the exchange fluxes, what compounds if added would allow for growth?

FBA Calculations: Other Conditions

6. Can *E. coli* grow anaerobically on acetate ?
   - **No**
7. Looking at the shadow prices for oxygen (o2) do you think that the cells could grow with acetate aerobically?
   - **Yes, since the shadow price for o2 is negative it means that if you added it growth would increase.**
8. Looking at the reduced costs for the exchange fluxes, what compounds if added would allow for growth?
   - **akg, fum, glc, lacD, o2, pyr**
   - **Each EX_ flux has a negative shadow price meaning if flux decreased (uptake of nutrient) then growth would increase.**

Overview of Constraint-Based Modeling Sessions

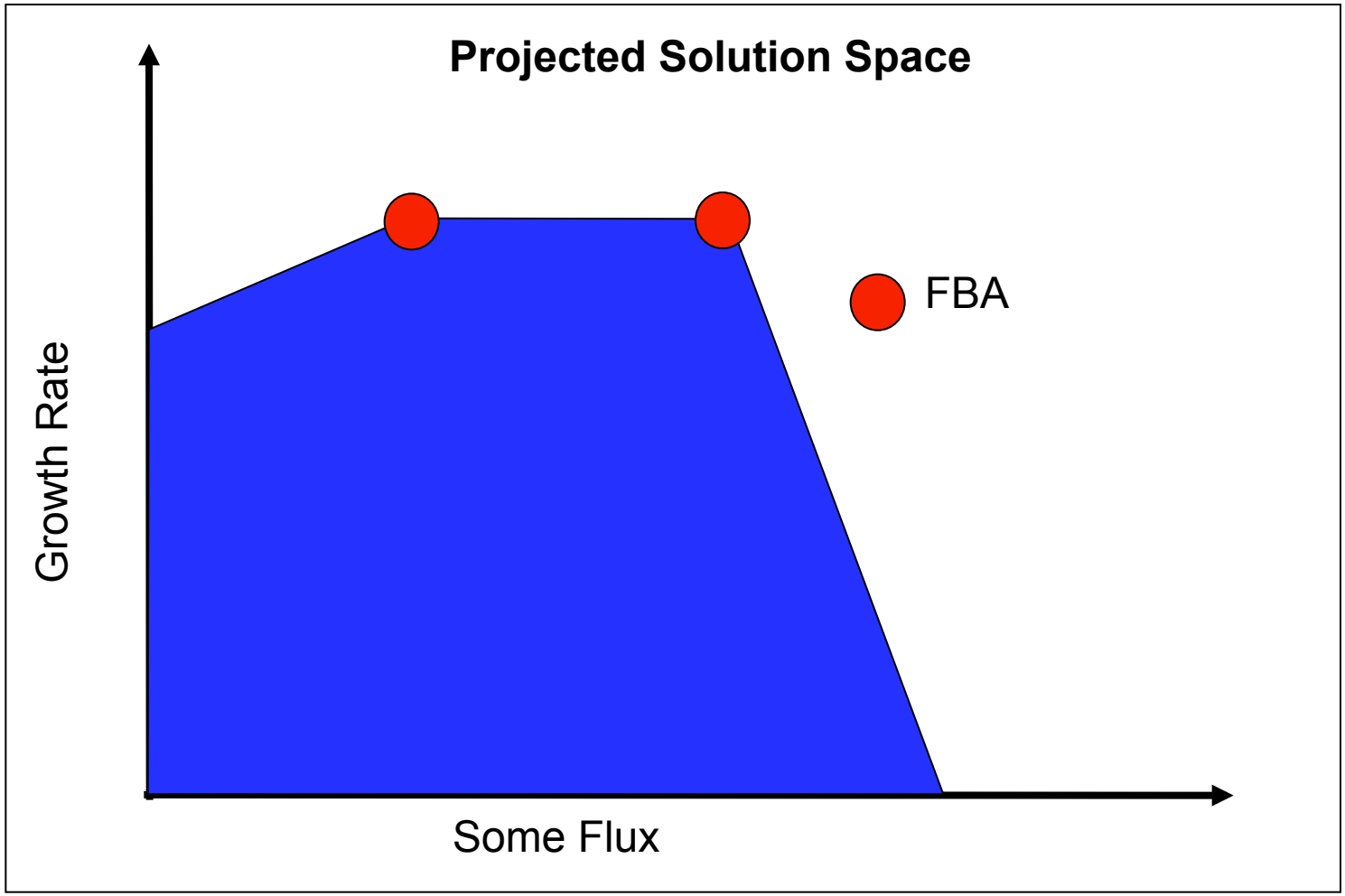
1. Reconstructing metabolic networks and flux balance analysis
2. Finding alternate solutions and predicting the effects of gene knockout
3. Improving models using optimization
4. Using models for metabolic engineering

1. Alternate Solutions

Flux Variability Analysis

Corner Point Solutions

Equivalent Optimal Solutions Exist:



How many solutions are there?

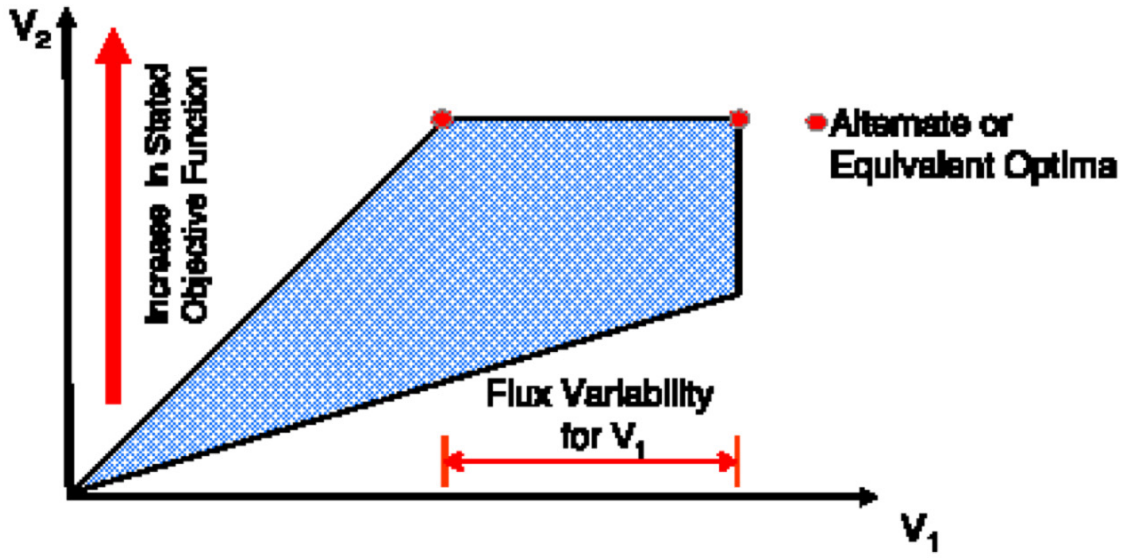
- Most FBA solutions in genome-scale networks are not unique.
  - The value of the objective function is unique.
  - The set of fluxes giving rise to the objective function are often not unique.

- 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

Distribution of Blocked Reactions

	H. pylori 389 rxns	*E. coli* 740 rxns	*S. cerevisiae* 1173 rxns
	number of blocked reactions		
Complex Media (Aerobic)	38	103	338
Glucose (Aerobic)	66	207	460
Glucose (Anaerobic)		210	515
Optimal Glucose (Aerobic)	77	408	774
Optimal Glucose (Anaerobic)		407	791

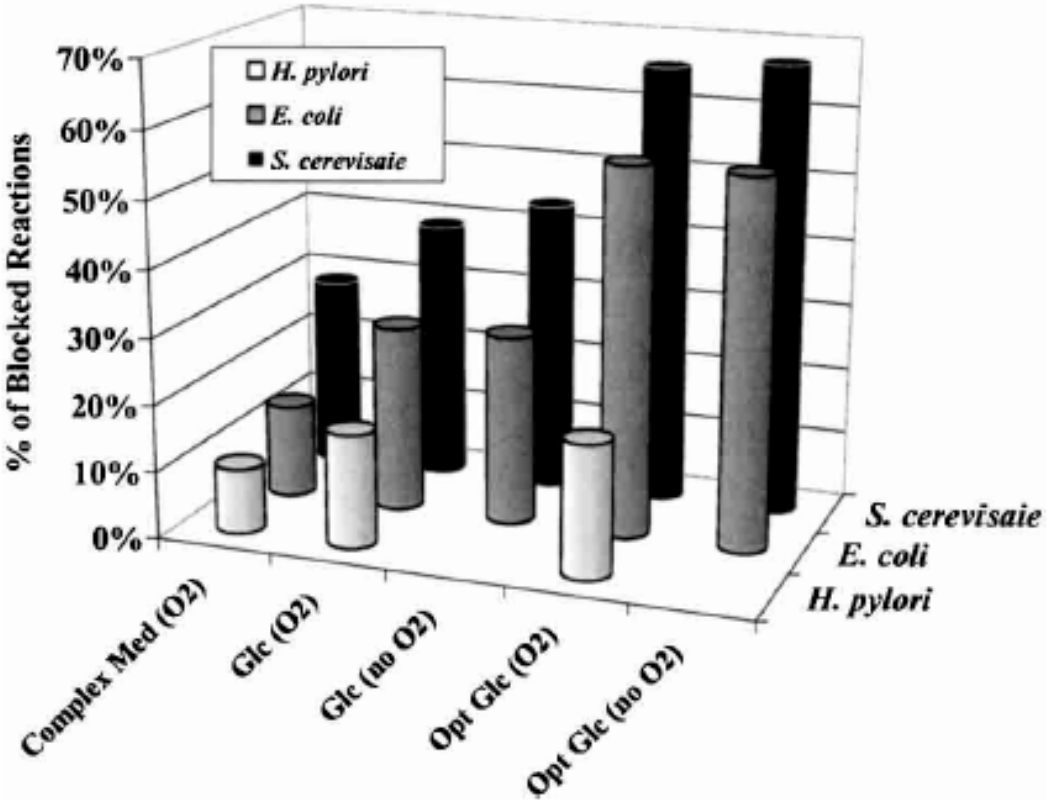


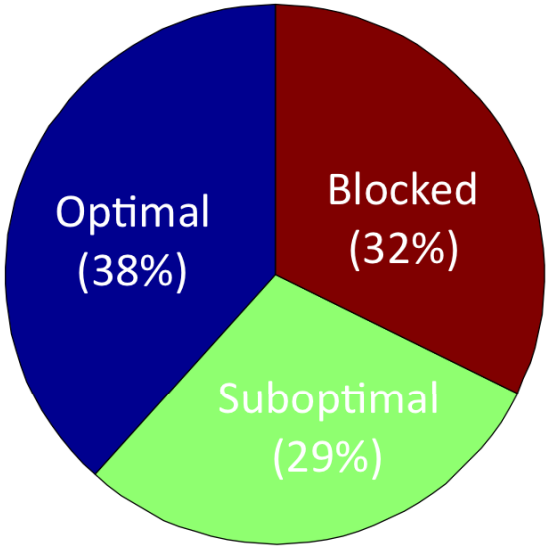
Figure 4 Total numbers and percentages of blocked reactions for the three networks under different growth conditions.

The number of blocked reactions (those which can not carry flux), depend on:

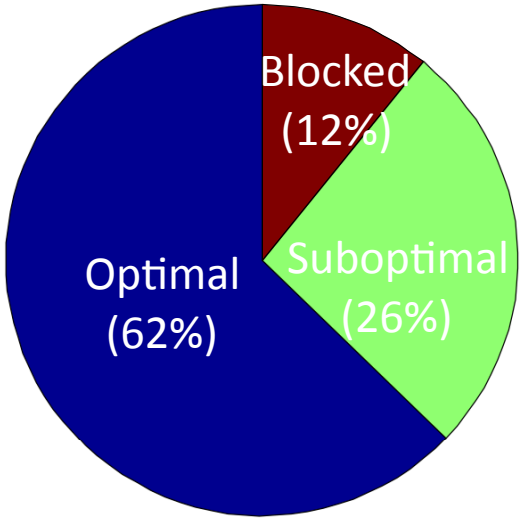
1) Network
2) Growth Condition

Burgard, AP, et al. Genome Research. 14(2):301-12 (2004).

Reaction Distribution



Detected Protein Distribution



Blocked Reactions (15 out of 129)

Superoxide Dismutase

Amino Acid tRNA Synthetases

Cofactor Biosynthesis

- Heme
- Ubiquinone

Suboptimal Reactions (34 out of 129)

Peroxidases

Respiration

- Cytochrome bd oxidase
- DMSO reductase

Fermentation

- Lactate Dehydrogenase
- Pyruvate Formate Lyase

Futile Cycles

- Phosphoenolpyruvate Synthase & Pyruvate Kinase
- Fructose Bisphosphatase & Phosphofructokinase

Purine Biosynthesis (4)

Amino Acid Biosynthesis (9):

- Threonine, Cysteine, Arginine, Asparagine

Raghunathan et al. BMC Systems Biology, 3:38 (2009)

```
$onecho > cplex.opt
eprhs 1e-9
epopt 1e-9
epint 1e-9
$offecho
```

Makes a file called cplex.opt with the following lines in it.
This changes the default tolerances for how exact the equations are (eprhs), how close to the optimal solution we are (epopt), and how close to integer values we are (epint).

```
LowerLimits('EX_glc_e')=-5;
UpperLimits('EX_glc_e')=0;
*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')=0;
LowerLimits('EX_pi_e')=-Vmax;
```

Define Media Inputs
(negative lower limits)

```
Set objective(j) /Biomass/;
```

Pick Flux to Optimize (here as a subset)

```
c(objective)=1;
solve fluxvariability using lp maximizing Obj;
v.fx(objective)=v.l(objective);
c(j)=0;

loop(duplicate_j,c(duplicate_j)=1;
         solve fluxvariability using lp maximizing Obj;
         store_maxs(duplicate_j)=Obj.l;
         solve fluxvariability using lp minimizing Obj;
         store_mins(duplicate_j)=Obj.l;
         c(duplicate_j)=0;  );
```

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

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: Instead of fixing flux at optimal value change line to be:

    v.lo(objective)=0.9*v.l(objective);

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 maximal 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 (i.e. biomass flux)?
  - ANS: 70

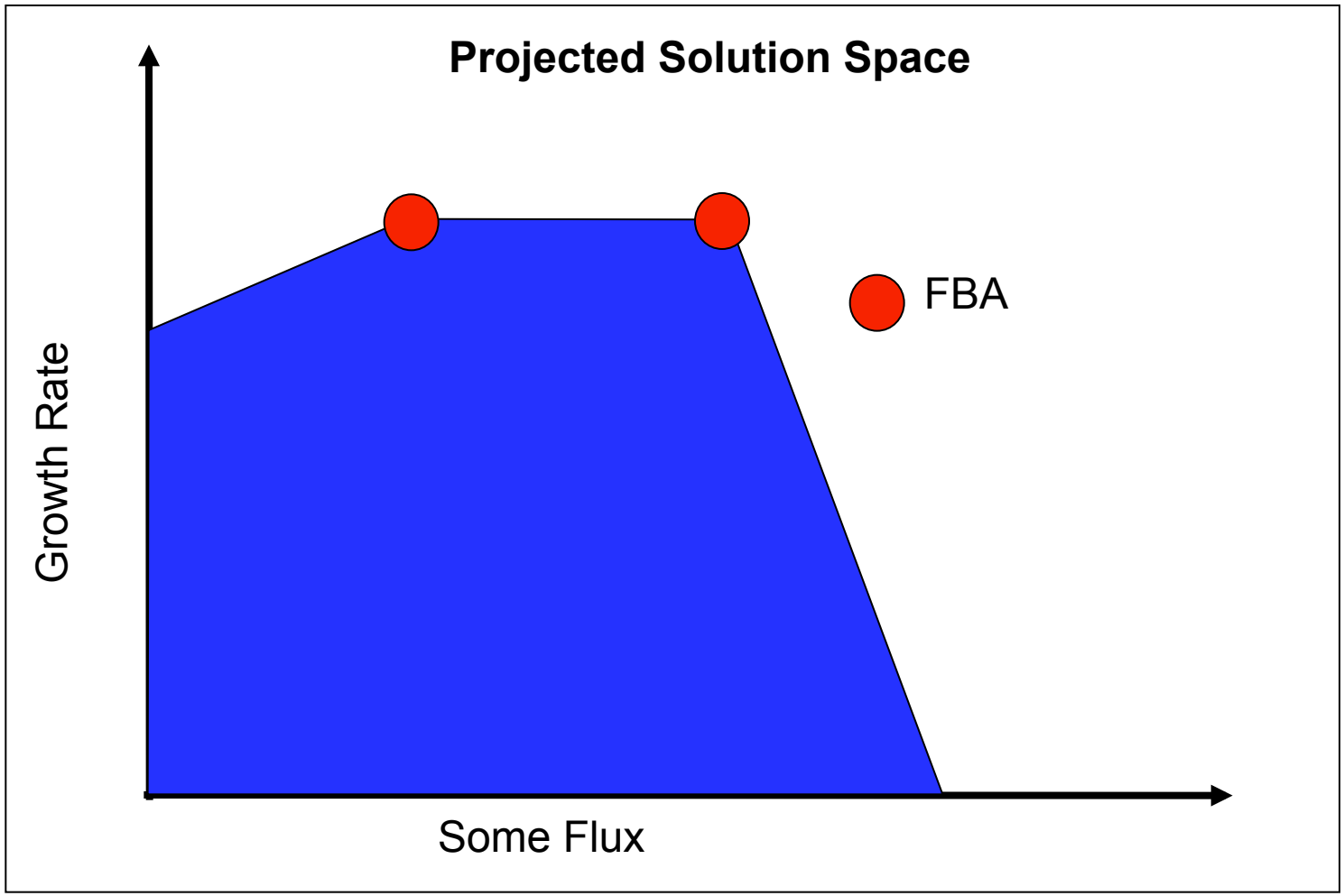
Flux Variability Calculations: Max Ethanol Production

- 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 maximal 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 Corner Point Solutions

Using Integer Cuts

Equivalent Optimal Solutions Exist:



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.

Enumeration Using Integer Cuts

$$\max \quad c \cdot v$$

$$\text{such that} \quad S \cdot v = 0$$

$$y_j \alpha_j \le v_j \le y_j \beta_j$$

$$\sum_{j \in NZ^k} y_j \le \left| NZ^k \right| - 1 \quad \text{for k} = 1, 2 \ldots \text{n} - 1$$

$$y_j \in \{0,1\}$$

NZ^k is the subset of fluxes that were non-zero in previous iteration k

$|NZ^k|$ is the number of fluxes that were non-zero in previous iteration k

Flux Variability vs. Alternate Optima

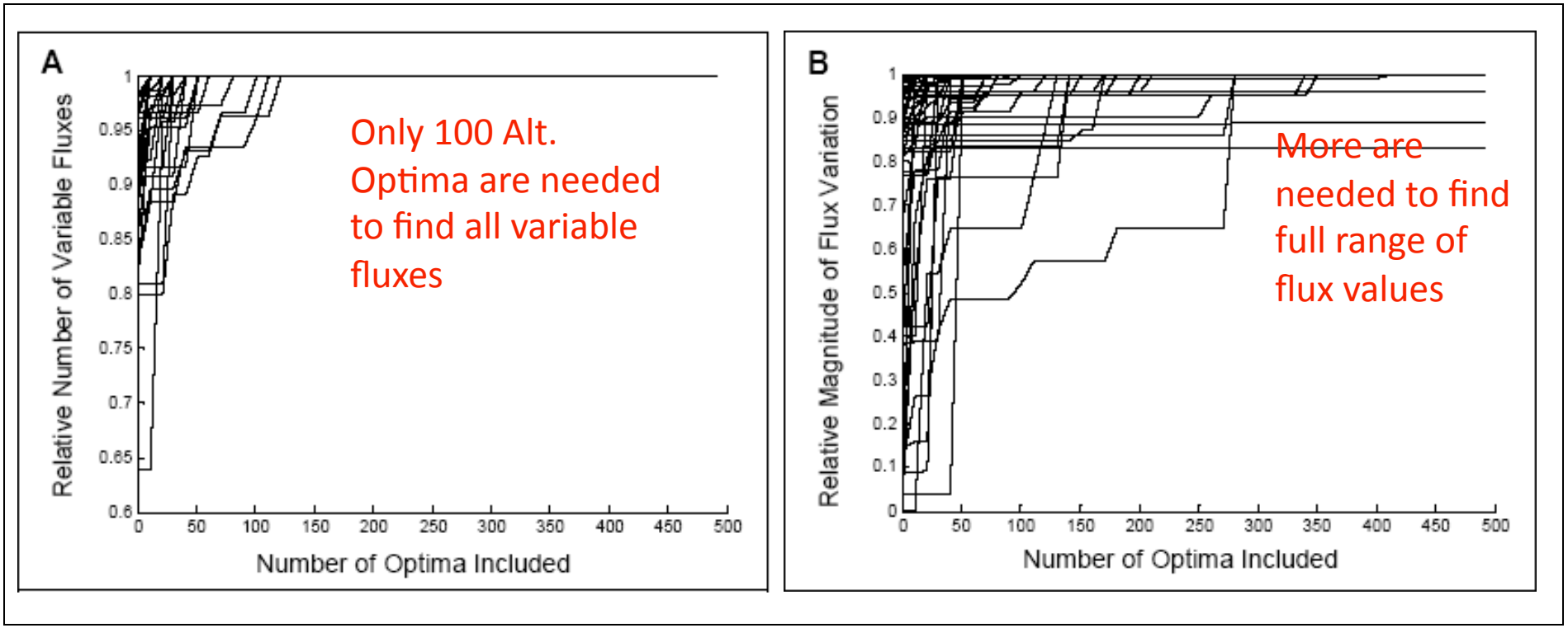
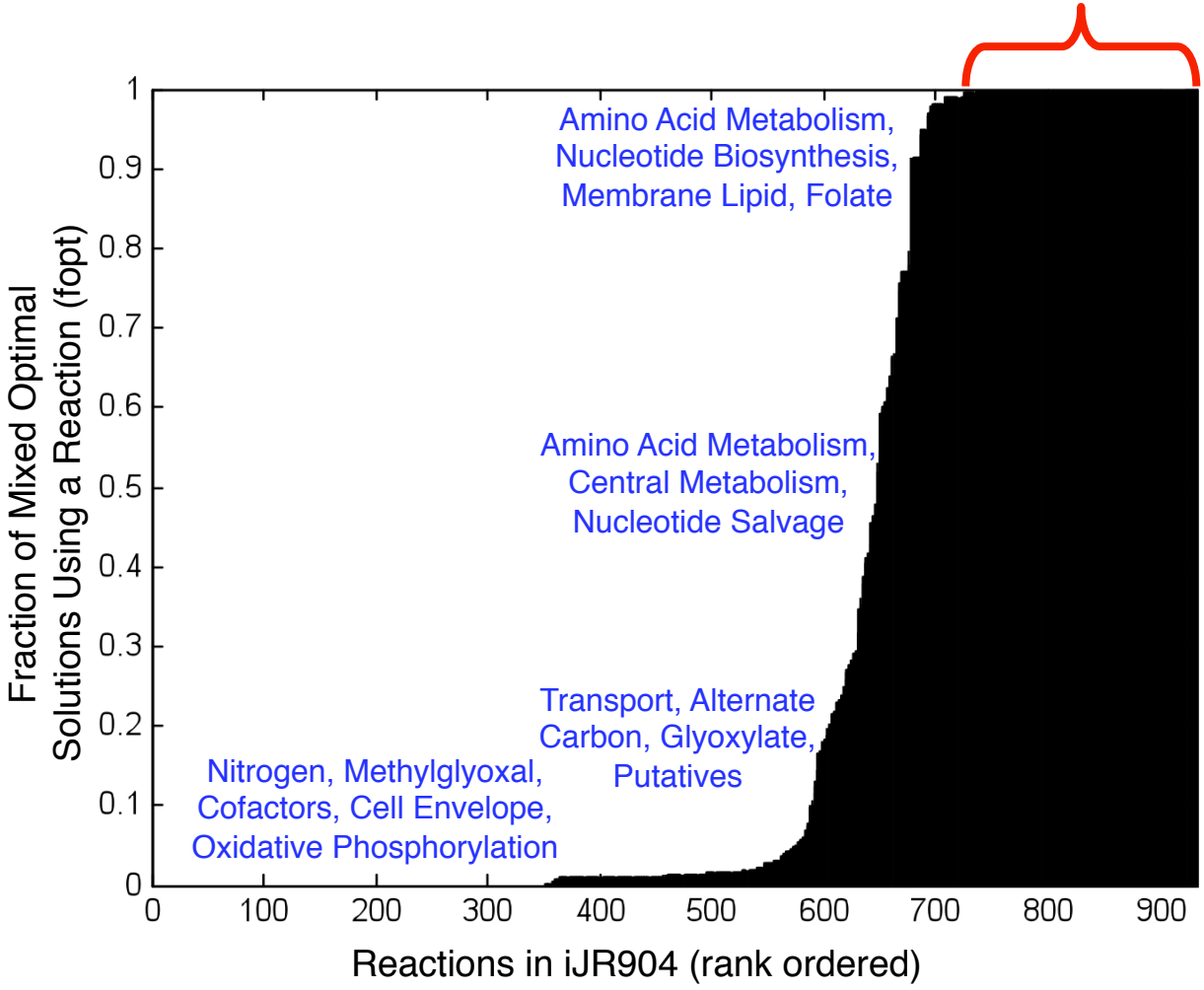


Figure 1 Comparisons of properties for sampled optima with all optima. The number of variable fluxes and the allowable ranges for these fluxes across all optima were calculated using a flux variability analysis. Each line is for one of the 88 carbon sources capable of supporting aerobic growth. (*A*) 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					
Oxidative phosphorylation	40	0.65					
Unassigned	9	0.78					
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				
Alternate Carbon Metabolism	130	0.27	0.65				
Glyoxylate Metabolism	5	0.40	0.60				
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					0.20	0.56
Tyrosine, Tryptophan, and Phenylalanine Metabolism	20					0.25	0.60
Folate Metabolism	6					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

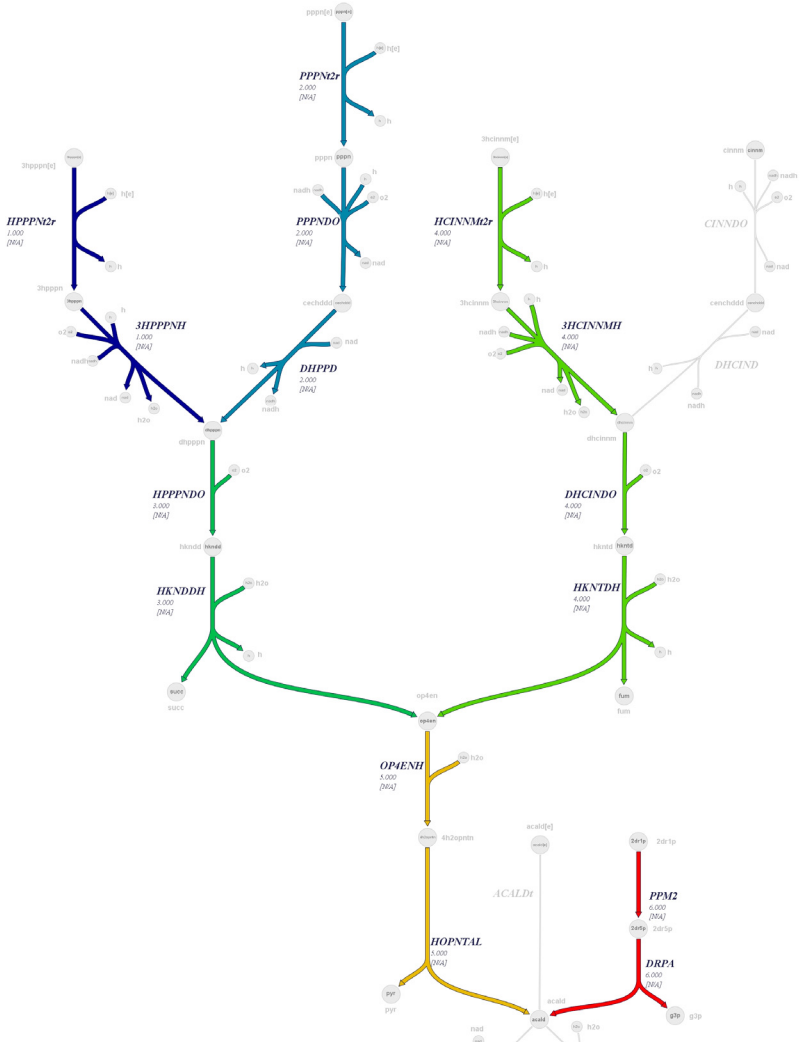
MOSTLY NEVER USED

SOMETIMES USED

ALWAYS USED

Correlated Reaction Sets in *E. coli*

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



Set of Rxns that Distinguish Alt Optima

```
*Define reactions that are used in distinguishing between alternate optimal solutions
*if you want all reactions to be used just use "Sets subj(j)" without the list of reactions
Sets subj(j) /ACKr,ACONT,ADHEr,ADK1,AKGDH,ATPS4r,CS,CYTBD,ENO,FBA,FBP,FRD,FUM,G6PDH2r,GAPD,GND,
ICDHyr,ICL,LDH_D,MALS,MDH,ME1,ME2,NADH11,NADTRHD,PDH,PFK,PFL,PGI,PGK,PGL,PGM,PPC,PPCK,PPS,PTAr,
PYK,RPE,RPI,SUCD1i,SUCD4,SUCOAS,TALA,THD2,TKT1,TKT2,TPI/
k how many alternate solutions to look for /alternate1*alternate20/;
```

How Many to Search For

```
Parameter
c(j) selects which fluxes are maximized in FBA /Biomass 1/
Objcrit stores the optimal value for the FBA objective function
PreviousNZ(subj,k) fluxes that are non-zero in previous solutions
EquivOptima(j,k) saves the flux distributions from previous solutions
PreviousSum(k) stores how many non-zero fluxes there were in previous solutions
epsilon /0.000000001/;

*Initialize PreviousSum and PreviousNZ so that future iterations don't constrain
*the current iteration (i.e. PreviousSum is a large number and PreviousNZ is zero)
PreviousSum(k)=card(j)+1;
PreviousNZ(subj,k)=0;

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

v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);

Binary variable y(subj);
```

Set the objective

A flux less than epsilon is considered to be zero

Initialize all $|NZ^k|$ to a large number so future integer cut constraints have no affect.

Initialize all NZ^k to be zero (i.e. empty)

Y is a binary variable

```
Equations
massbalance(i) mass balance equations for each metabolite
calcobj calculates the dot product of the c vector the flux vector
integercut(k) ensures that at least one non-zero flux from previous iterations
upperbound(subj) constrains fluxes by integer variable y
lowerbound(subj) constrains fluxes by integer variable y;

massbalance(i).. sum( j,S(i,j)*v(j) )=e=0;
calcobj.. Obj=e=sum( j,c(j)*v(j) );
integercut(k).. sum(subj, y(subj)*PreviousNZ(subj,k))=l=PreviousSum(k)-1;
upperbound(subj).. v(subj)=l=( y(subj)*UpperLimits(subj) );
lowerbound(subj).. v(subj)=g=( y(subj)*LowerLimits(subj) );

model FBA /massbalance, calcobj/;
model AltOptima /massbalance,calcobj,integercut,lowerbound,upperbound/;
FBA.optfile=1; AltOptima.optfile=1; AltOptima.OptCr=0;
```

All integer cut constraints, for past and future iterations.

For future iteractions PreviousNZ will be zero and PreviousSum will be large.

```
solve FBA using lp maximizing Obj;
Objcrit=Obj.l;
EquivOptima(j,'alternate1')=v.l(j);

alias (k,temp);

*Use AltOptima to find other equivalent solutions
loop(temp,
  if(ord(temp)>1,
    if(abs(Objcrit-Obj.l)<=epsilon,
         PreviousSum(temp-1)=0;
         loop(subj,
               if( abs(v.l(subj))>epsilon, PreviousNZ(subj,temp-1)=1; PreviousSu
              );
         solve AltOptima using mip maximizing Obj;
         EquivOptima(j,temp)=v.l(j););  );
  );
```

Solve for first optimal solution

For iterations greater than 1

If last solution was still an optimal solution

Find NZ and |NZ| from last solution

Find another solution with the integer cut constraint from previous solution modified so that it is now constraining

Alternate Optima Calculations

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

- What if instead you maximize for ethanol production:
  - How many alternate solutions are there?
  - Are the differences in fluxes across solutions because of exchange reactions or due to internal metabolic reactions?
  - Do cells need oxygen to make the maximum amount of ethanol from glucose?

Alternate Optima Calculations

- How many alternate solutions are there for glucose aerobic growth, where you maximize for biomass production?
  - ANS: Only 1 solution

- What if instead you maximize for ethanol production:
  - How many alternate solutions are there?

    ANS: ten solutions
  - Are the differences in fluxes across solutions because of exchange reactions or due to internal metabolic reactions?

    ANS: internal metabolic reactions, exchanges are the same.
  - Do cells need oxygen to make the maximum amount of ethanol from glucose?

    ANS: internal metabolic reactions, exchanges are the same.

2. Gene Deletions

FBA

MOMA

ROOM

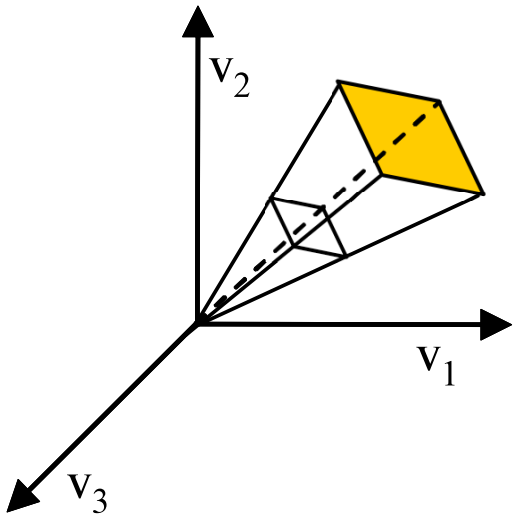
FBA Optimization Problem Statement

- 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 $S \cdot v = b = 0$

$$\alpha \leq v \leq \beta$$

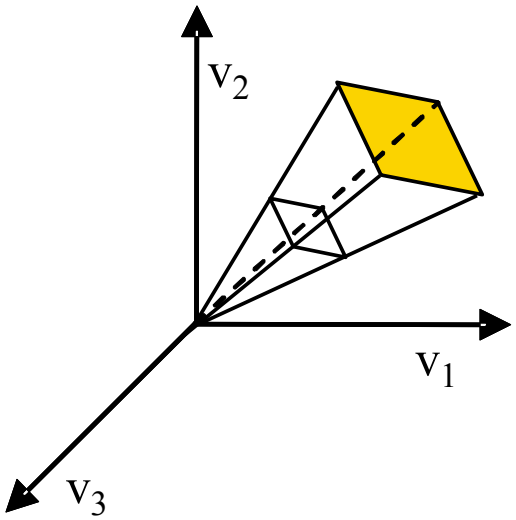


FBA: Wildtype vs. Knockout Mutant

Maximize: $c \cdot v$

Such that $S \cdot v = 0$

$\alpha \leq v \leq \beta$

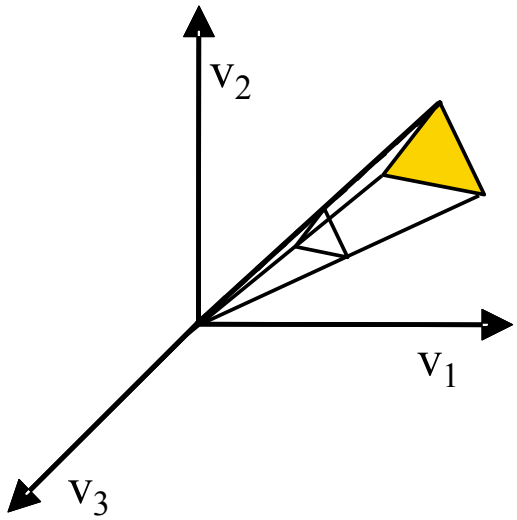


Maximize: $c \cdot v$

Such that $S \cdot v = 0$

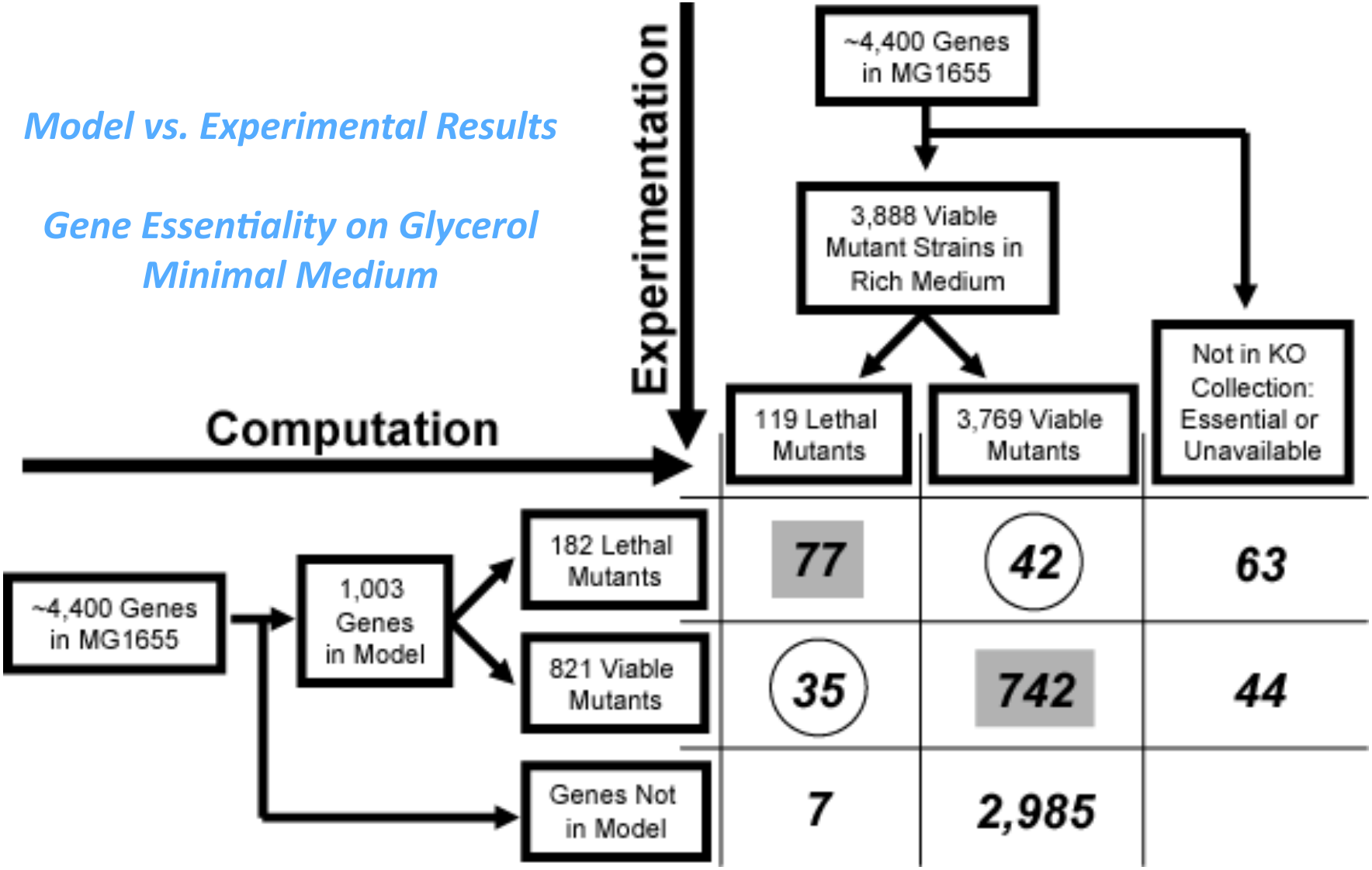
$\alpha \leq v \leq \beta$

$v_k = 0$



Model vs. Experimental Results

Gene Essentiality on Glycerol Minimal Medium



Overall Model is 91 % Accurate

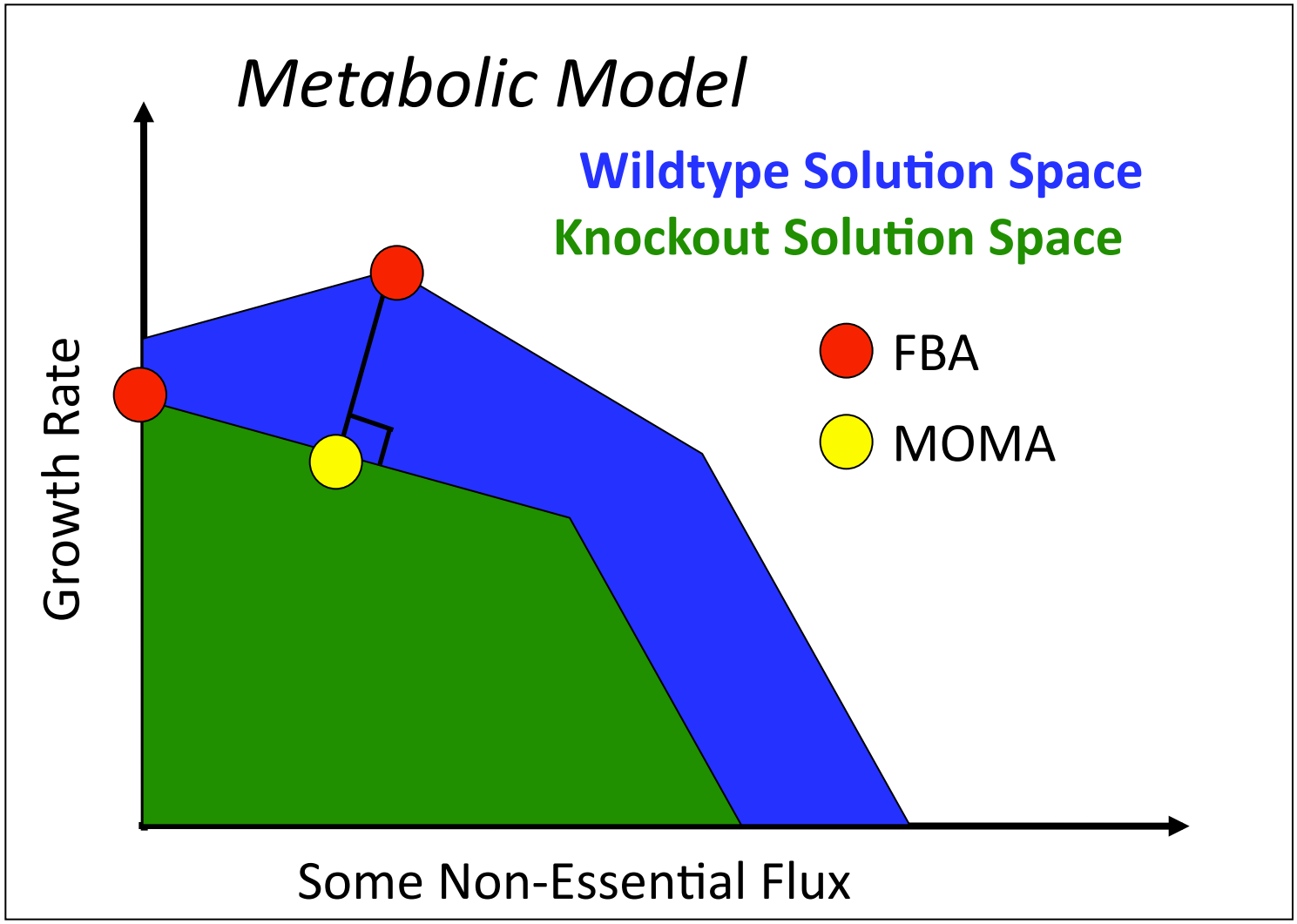
Discrepancies Generate Hypotheses

TABLE 2. Discrepancies between experimental identification and model prediction for essential and nonessential genes[a]

Rationale	Subsystem	Gene(s) (Blattner no.)
False negatives	**Essential Experimentally, Model Predicts Growth**	
Molecules not included in biomass	Ubiquinone biosynthesis	*ubiG* (b2232); *ubiH* (b2907)
	Pyridoxine biosynthesis	*pdxA* (b0052); *pdxB* (b2320); *pdxH* (b1638); *pdxJ* (b2564)
	Thiamine biosynthesis	*iscC* (b2530)
Model includes alternative pathways/ isozymes	Amino acid biosynthesis	*carA* (b0032); *carB* (b0033); *glpD* (b3426); *glyA* (b2551); *proA* (b0243); *proB* (b0242); *thrB* (b0003); *thrC* (b0004)
Model predicts impaired but not lethal phenotype	ATP synthase	*atpA* (b3734); *atpB* (b3738); *atpC* (b3731); *atpF* (b3736); *atpG* (b3733); *atpH* (b3735)
Regulatory effect on *glpK*	PTS/PEP metabolism	*crr* (b2417); *glpK* (b3926); *ppc* (b3956); *ptsI* (b2416); *fruR* (b0080)
False positives	**Non-Essential Experimentally, Model Predicts No Growth**	
Model biomass components which might not be essential components	Fatty acid and lipid biosynthesis	*cls* (b1249); *fabF* (b1095)
	Glycogen	*glgA* (b3429); *glgC* (b3430)
	LPS synthesis	*dgkA* (b4042); *gmhA* (b0222); *gmhB* (b0200); *lpxL* (b1054); *msbB* (b1855); *rfaC* (b3621); *rfaD* (b3619); *rfaE* (b3052); *rfaF* (b3620); *rfaG* (b3631); *rfaI* (b3627); *rfaJ* (b3626); *rfaL* (b3622)
	Spermidine synthesis	*pfs* (b0159); *speD* (b0120); *speE* (b0121)
Unaccounted-for transport mechanisms	Ammonium transport	*amtB* (b0451)
	Glycerol transport	*glpF* (b3927)
	Sulfate transport	*cysW* (b2423)
Unaccounted-for metabolic enzymes	Arginine biosynthesis	*argB* (b3959); *argC* (b3958); *argD* (b3359); *argG* (b3172)
	Aspartate biosynthesis	*aspC* (b0928)
	Branced amino acid biosynthesis	*ilvY* (b3773); *ilvE* (b3770); *lrp* (b0889)
	Central metabolic	*aldA* (b1415)
	Cofactor biosynthesis	*coaA* (b3974); *coaE* (b0103); *pabC* (b1096)
	Glycolytic	*pgi* (b4025)
	Lysine biosynthesis	*dapF* (b3809); *ushA* (b0480); *lysR* (b2839)
	Nucleotide biosynthesis and salvage	*pyrI* (b4244); *trxB* (b0888); *ndk* (b2518)

[a] Twenty-six false-negative cases in which the model incorrectly predicted growth of the gene deletion strain were identified, in addition to 42 false-positive cases in which the model incorrectly predicted that genes were essential. Each case is grouped based on the likely rationale for the discrepancy and the gene functional annotation.

MOMA: Minimize Distance Between Wildtype & Mutant Flux Distributions



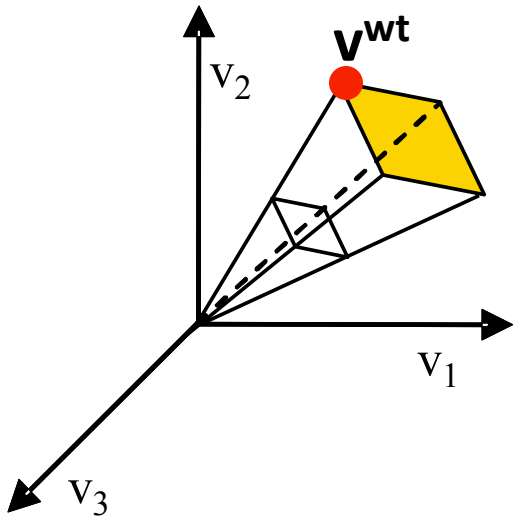
MOMA Prediction Algorithm

Maximize: $c \cdot v$

Such that $S \cdot v = 0$

$\alpha \leq v \leq \beta$

SOLUTION = v^{wt}

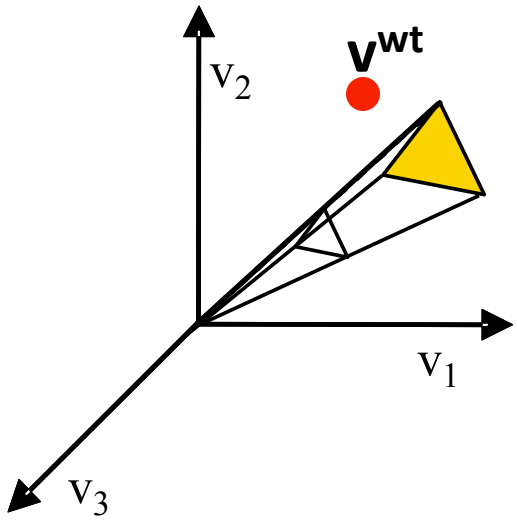


Minimize: $\sum (v_j^{wt} - v_j)^2$

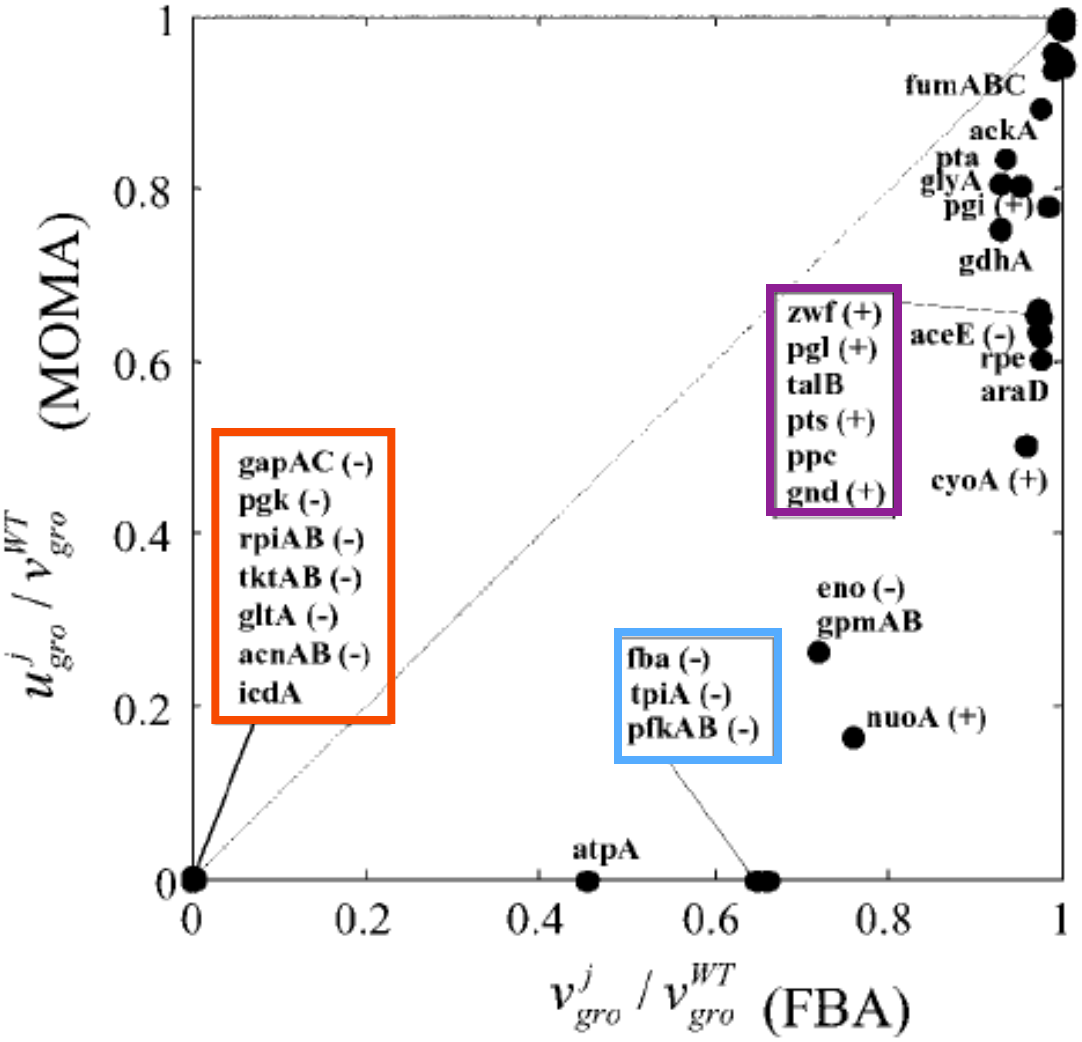
Such that $S \cdot v = 0$

$\alpha \leq v \leq \beta$

$v_k = 0$



FBA vs. MOMA Mutant Growth Rate Predictions

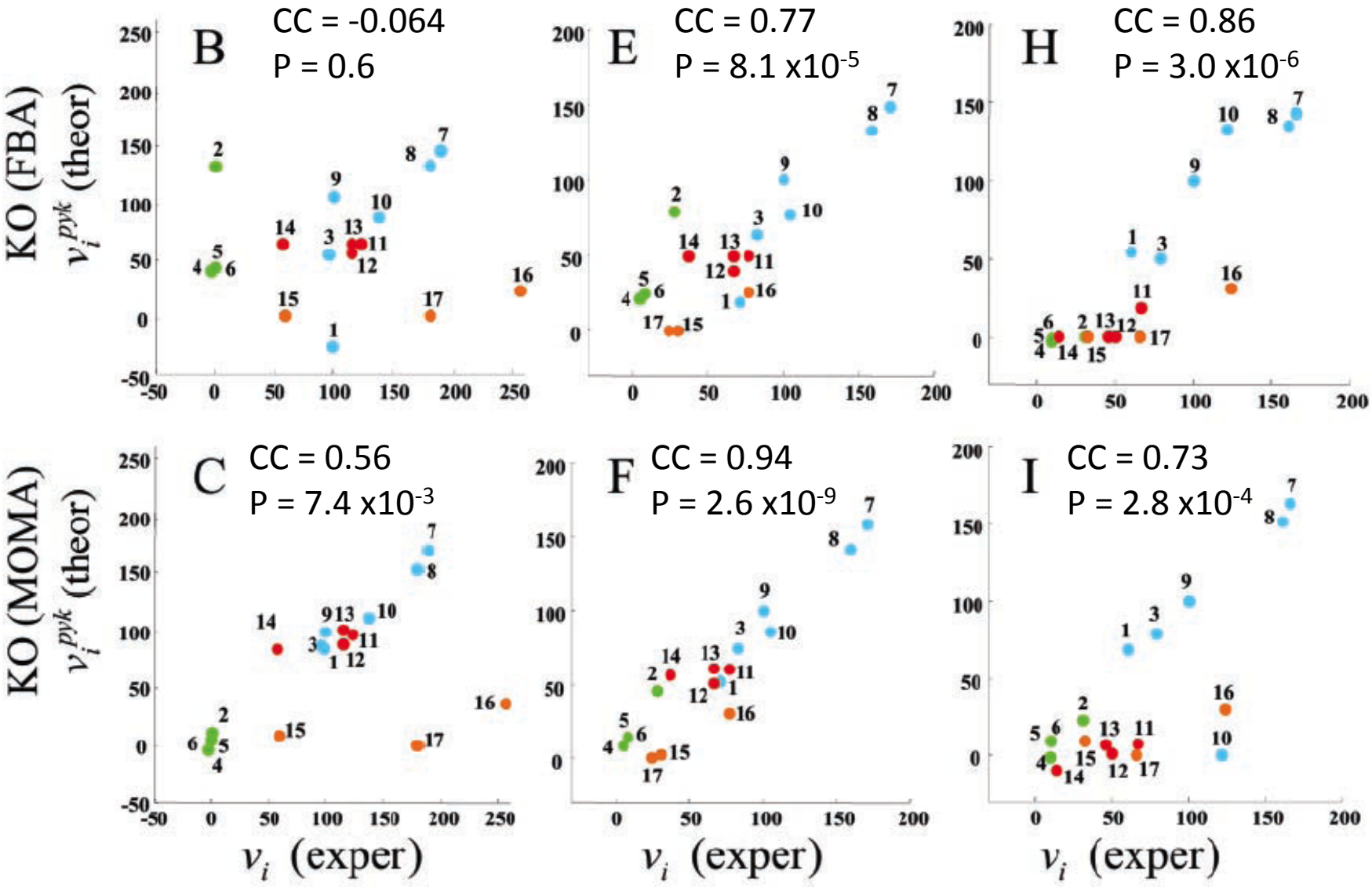


Both FBA and MOMA predict lethal phenotypes, agreeing with experimental data

Both FBA and MOMA predict non-lethal phenotypes, agreeing with experimental data

Only MOMA predicts a lethal phenotype, agreeing with experimental data

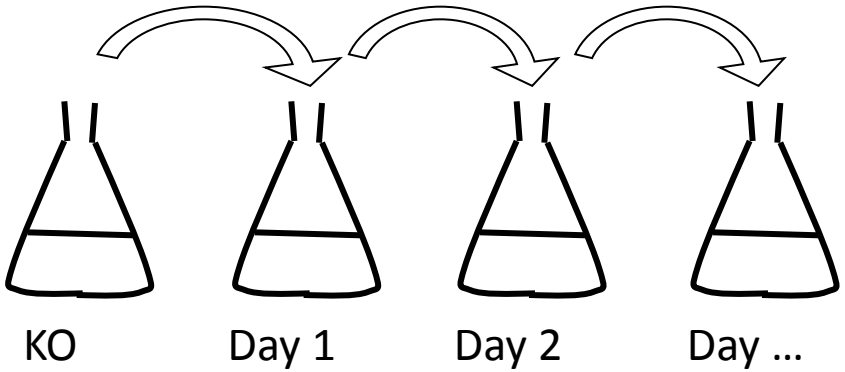
FBA vs. MOMA Flux Level Predictions

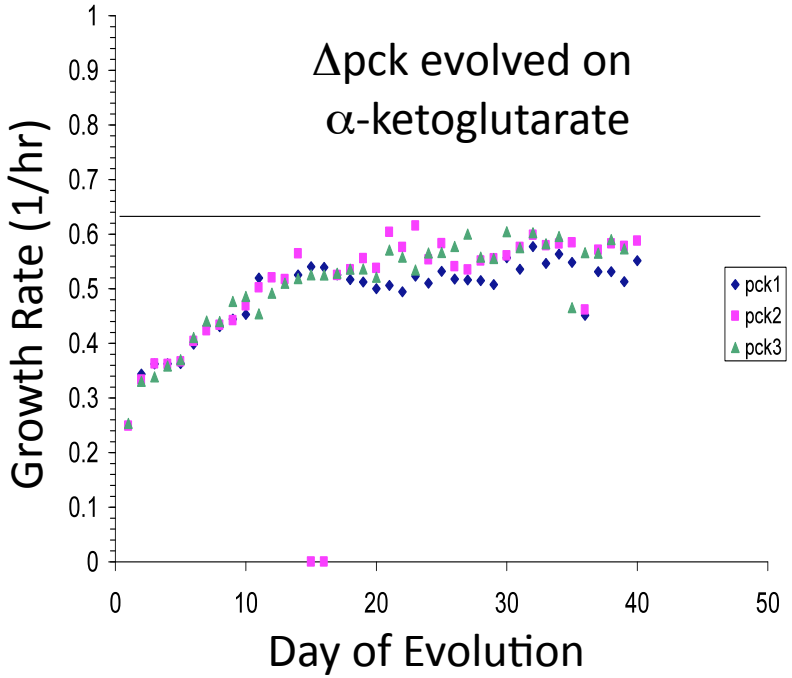


CC = Correlation Coefficient and P = p-value

Segre, et al. PNAS. 99(23): 15112-15117 (2002)

What Happens if Cells Evolve?

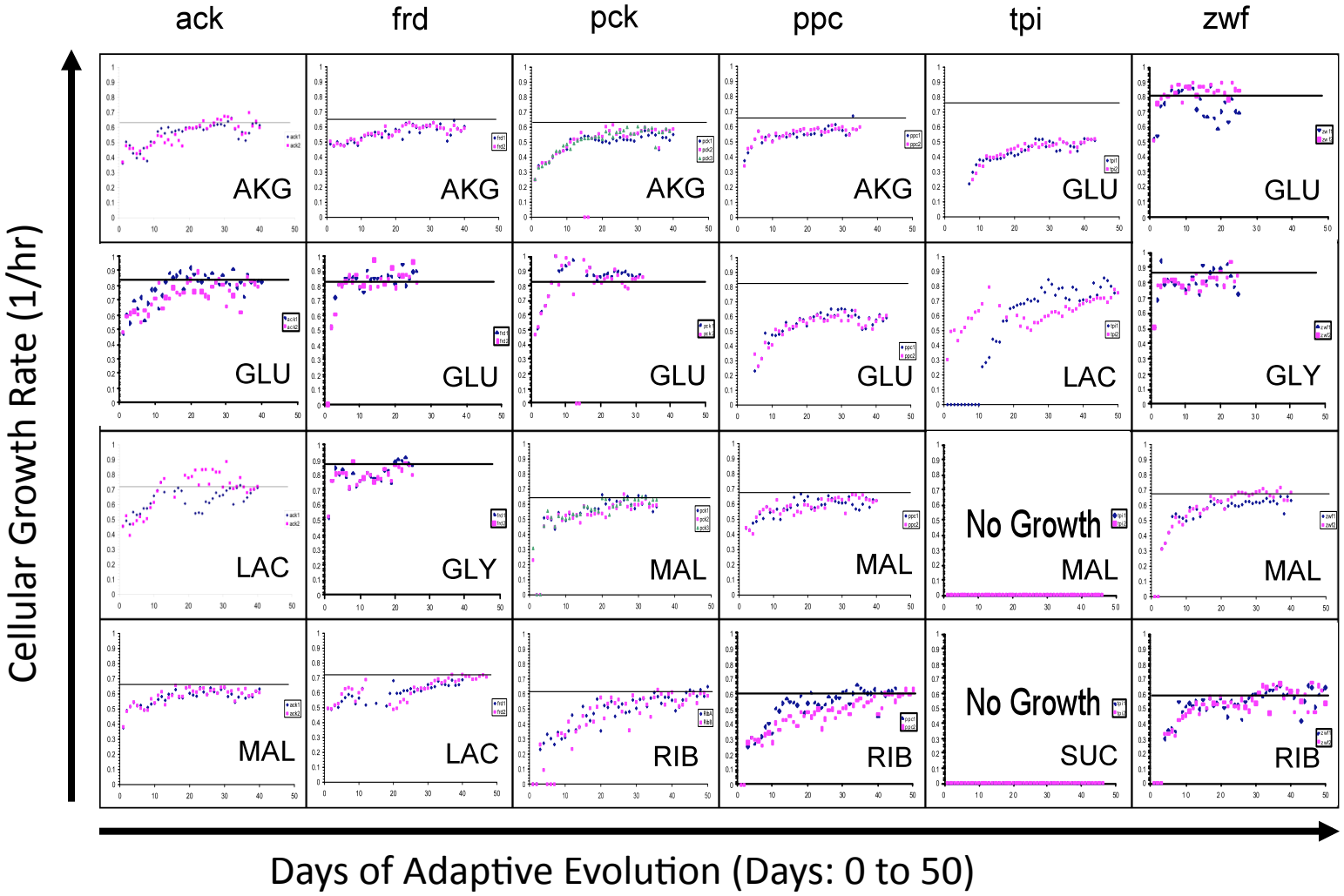




Faster growing cells outcompete others and select for cells with higher growth rates

Fong et al. Nature Genetics. 36(10): 1056-1058 (2004)

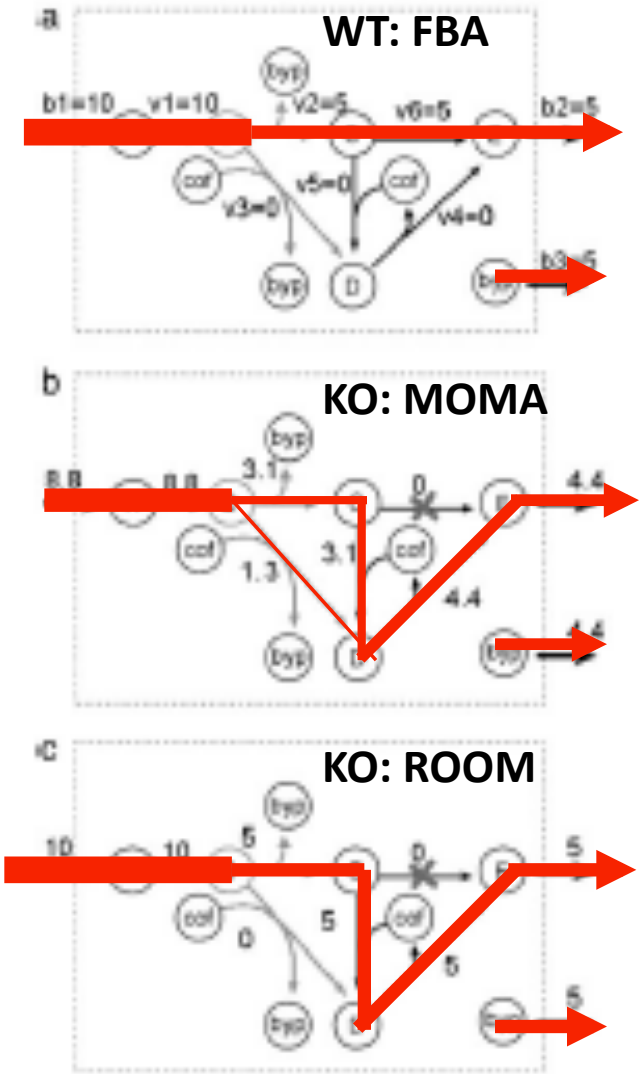
Deletion Strain Evolution



- *39 of 50 cases correctly predicted computationally (within 10%)*
- *Parallel cultures exhibit similar endpoint phenotypes*
- *Average GR increase of 87% observed*

Fong et al. Nature Genetics. 36(10): 1056-1058 (2004)

ROOM: Minimize the Number of Fluxes that Change



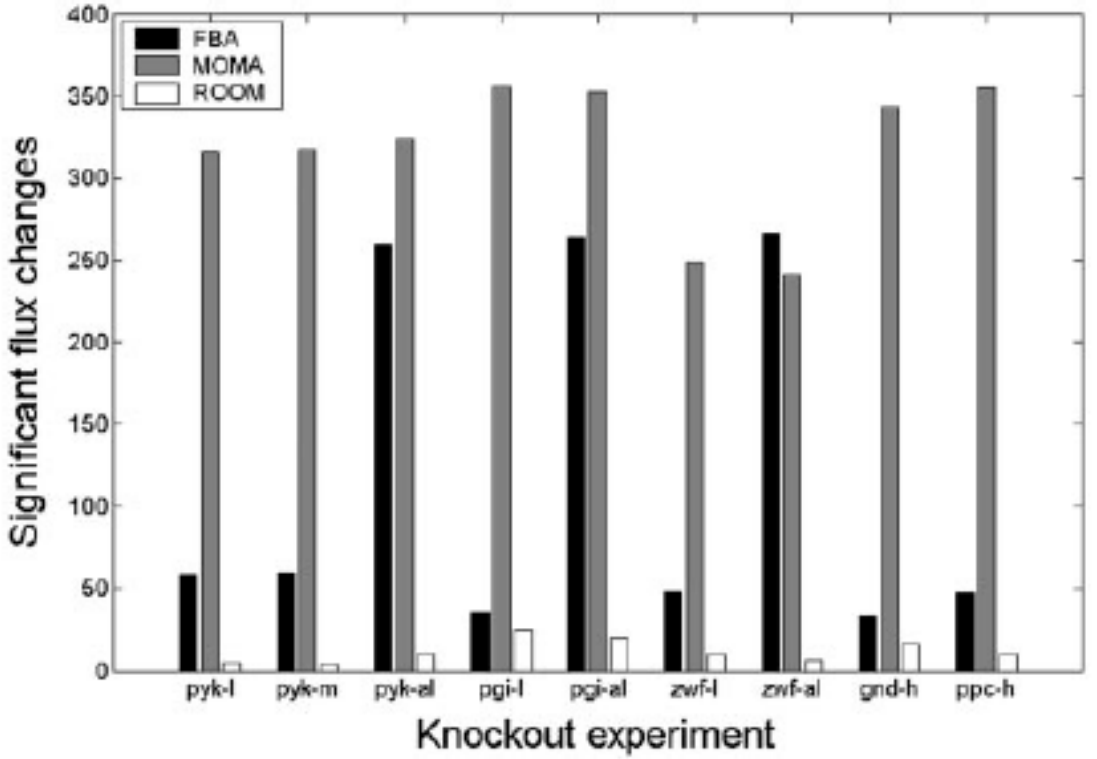


Fig. 4. Number of significant flux changes between the flux distribution of the wild-type strain and the flux distributions predicted by FBA, MOMA, and ROOM for five knocked-out organisms, under different growth conditions. The marking on the *x* axis is explained in the caption of Fig. 3.

Shlomi et al. PNAS (2005). 102(21):7695-700

Method Comparison to Experimental Data

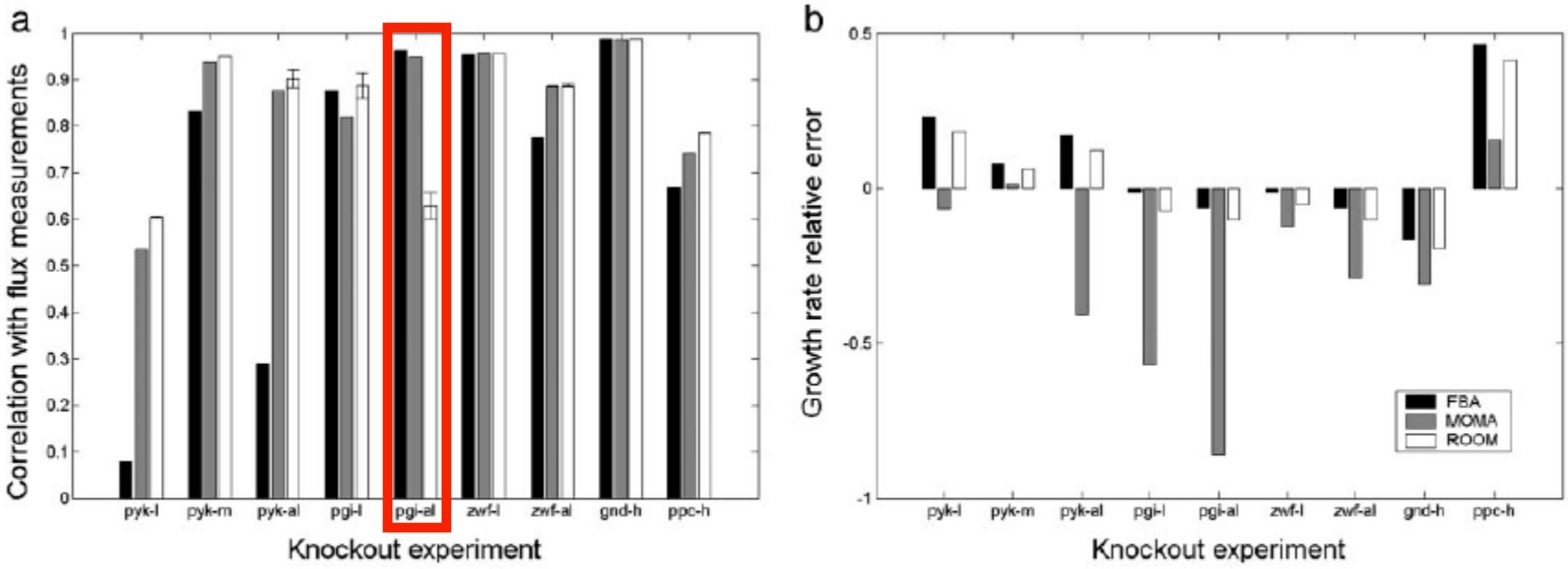


Fig. 3. Flux and growth-rate comparison among FBA, MOMA, and ROOM for five knocked-out organisms, under different growth conditions. The marking *x-y* on the *x* axis denotes knockout of gene *xxx-y* in a mutant strain grown on media *y*. l, m, h, and al stand for glucose-low, glucose-medium, glucose-high, and ammonia-low, respectively. (*a*) Pearson correlations between experimental fluxes and predictions. (*b*) Relative errors in growth rate predictions, calculated by subtracting the experimentally measured growth rate from the predicted growth rate and dividing by the experimentally measured growth rate.

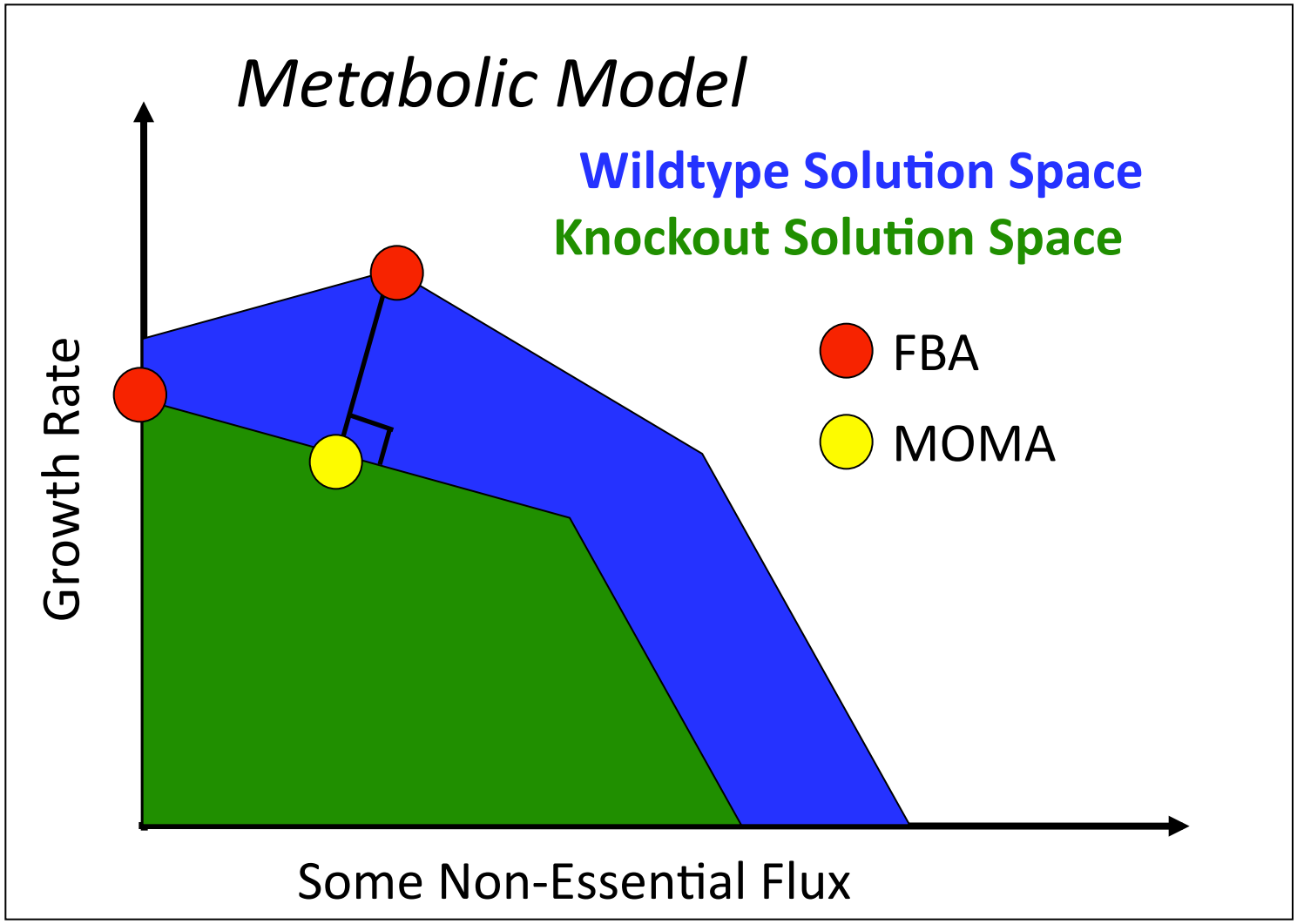
- In 8 out of 9 cases ROOM has better or equal prediction capabilities with respect to flux over MOMA.
- MOMA tends to more significantly underpredict growth rate (quantitative comparison)

Shlomi et al. PNAS (2005). 102(21): 7695-700

Final Points

- FBA will always predict higher (or equal) growth rates as compared to MOMA or ROOM.
- The MOMA solution is unique given a single wildtype flux distribution.
- The ROOM solution is not unique, there are often multiple flux distributions with the same number of altered fluxes.
- FBA better at predicting adaptive evolutionary outcomes.

MOMA: Minimize Distance Between Wildtype & Mutant Flux Distributions



ROOM, MOMA and FBA Predictions for Mutants

```
LowerLimits('EX_glc_e')=-5;
UpperLimits('EX_glc_e')=0;
*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;

*Define reactions that are used in the ROOM objective function if you want to
*consider all reactions then use 'alias(subj,j);' instead.
Set subj(j) /ACKr,ACONT,ADHEr,ADK1,ATPS4r,CS,CYTBD,ENO,FBA,FBP,FRD,FUM,G6PDH2r
GAPD,GND,ICDHyr,ICL,LDH_D,MALS,MDH,ME1,ME2,NADH11,PDH,PFK,PFL,PGI,PGK,PGL,PGM
PPC,PPCK,PPS,PTAr,PYK,PYRt2r,RPE,RPI,SUCD1i,SUCD4,SUCOAS,TALA,AKGDH,NADTRHD
THD2,TKT1,TKT2,TPI/;

Set deletedrxns(j) /TKT1/
WTobj(j) /Biomass/;

Parameter
c(j)            used to define the objective function for FBA
wildtype_v(j)   used to store wildtype FBA fluxes
mutant_room(j)  used to store mutant MOMA fluxes
mutant_fba(j)   used to store mutant FBA fluxes
mutant_moma(j)  used to store mutant MOMA fluxes
delta  used to indicate what flux changes are significant (ROOM)
epsilon used to indicate what flux changes are significant (ROOM)
wL(subj) used to indicate what flux changes are significant (ROOM)
wU(subj) used to indicate what flux changes are significant (ROOM);
```

Define Which Reaction(s) to Delete

Define What to Maximize to get the "Wildtype" Flux Distribution

ROOM, MOMA and FBA Predictions for Mutants

```
*************************************************************************
*This section calculates the FBA solution for maximizing biomass
*for the wildtype strain and stores the fluxes in the wildtype_v parameter
v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);
c(WTObj)=1;
solve FBA using lp maximizing Obj;
wildtype_v(j)=v.l(j);

*Defines allowable variation before becoming significant for ROOM calculations
wU(subj)=wildtype_v(subj)+delta*abs(wildtype_v(subj))+epsilon;
wL(subj)=wildtype_v(subj)-delta*abs(wildtype_v(subj))-epsilon;

v.fx(deletedrxns)=0;
*************************************************************************
*This section calculates the ROOM and MOMA solutions for the appropriate knockout
*indicated by the line v.fx('rxnname')=0;
*It also calculates the FBA solution for this same knockout
solve ROOM using mip minimizing minnumber;
mutant_room(j)=v.l(j);

solve MOMA using nlp minimizing distance;
mutant_moma(j)=v.l(j);
solve FBA using lp maximizing Obj;
mutant_fba(j)=v.l(j);
```

Calculate the WT flux distribution

Fix all fluxes in the deletedrxn set to be 0

Calculate the ROOM Mutant distribution

Calculate the MOMA Mutant distribution

Calculate the FBA Mutant distribution

Knockout Calculations

1. What are the maximum growth rates for the wildtype and mutant strains predicted using: MOMA, ROOM and FBA for the following cases:
   - tpi mutant (glucose aerobic)
   - pgi mutant (glucose aerobic)
   - acnA, acnB double mutant (glucose aerobic)

Knockout Calculations

2. If you delete ACONT, all methods predict a lethal phenotype. Looking at the shadow prices for the Mutant FBA prediction what metabolite can this mutant no longer produce that is needed for biomass production?

BIOMASS REACTION

1.496 3pg + 3.7478 accoa + 1.0789 akg + 55.703 atp + 0.361 e4p + 0.0709 f6p + 0.129 g3p + 0.205 g6p + 55.703 h2o + 3.547 nad + 18.225 nadph + 1.7867 oaa + 0.5191 pep + 2.8328 pyr + 0.8977 r5p

→ 55.703 adp + 3.7478 coa + 41.025 h + 3.547 nadh + 18.225 nadp + 55.703 pi

Knockout Calculations (1. Ans)

- tpi mutant (delete TPI reaction)
  - 0.49(WT FBA)
  - 0.08(Mutant MOMA)
  - 0(Mutant ROOM)*
  - 0.35(Mutant FBA),
- pgi mutant (delete PGI reaction)
  - 0.49(WT FBA)
  - 0.47(Mutant MOMA)
  - 0.262 (Mutant ROOM)*
  - 0.49(Mutant FBA)
- acnA+acnB mutant (delete ACONT reaction)
  - 0.49(WT FBA)
  - All predict methods predict 0

Knockout Calculations (2. Ans)

2. Only a few compounds have negative shadow prices:

akg = a-ketoglutarate

icit = isocitrate

glx = glyoxylate

akg_e = α-ketoglutarate

→ Only akg is part of the biomass equation.

Overview of Constraint-Based Modeling Sessions

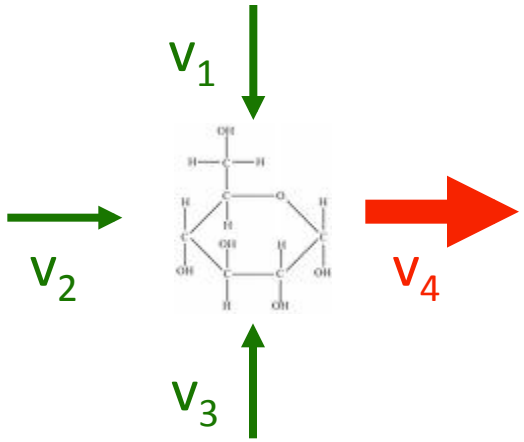
1. Reconstructing metabolic networks and flux balance analysis
2. Finding alternate solutions and predicting the effects of gene knockout
3. Improving models using optimization
4. Using models for metabolic engineering

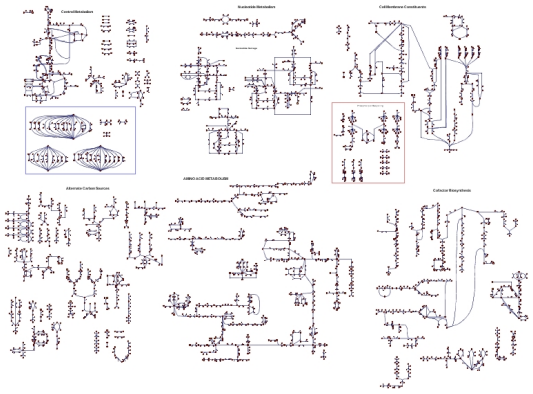
Model Corrections

1. SMILEY (metabolic)
2. GROWMATCH (metabolic)
3. GENEFORCE (metabolic & regulatory)

Constraints on Metabolic Networks

1. *Steady-State Mass Balance Constraints*





For each metabolite:

$\sum s_{ij} \cdot v_{produce} = \sum -s_{ij} \cdot v_{consume}$

For all metabolites:

$\mathbf{S} \cdot v = 0$

2. *Enzyme Capacity Constraints:* $v_{min} \leq v_j \leq v_{max}$
3. *Thermodynamic Constraints:* $v_j \geq 0$
4. *Regulatory Constraints*: v_{min}, $v_{max} = 0$ if associated genes are not expressed

Current Status of *E. coli* Genome

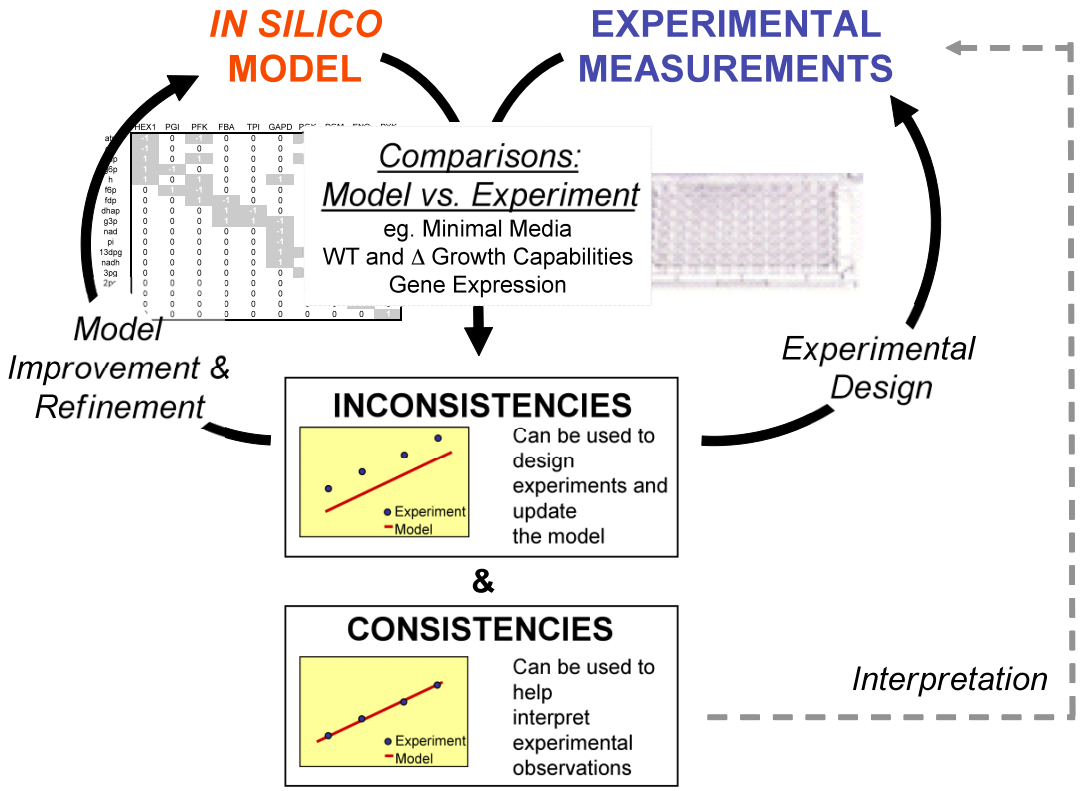
Table 3. Numbers and types of known and predicted gene products of *E.coli* K-12[1]

Code	Gene product type	Number	Percentage[2]
e	Enzyme	1094	33.3
pe	Enzyme, predicted	390	
t	Transporter	[illegible]	13.3
pt	Transporter, predicted	254	
r	Regulator	[illegible]	9.1
pr	Regulator, predicted	164	
m	Membrane	[illegible]	5.7
pm	Membrane, predicted	210	
f	Factor	150	4.7
pf	Factor, predicted	60	
s	Structural component	[illegible]	2.8
ps	Structural component, predicted	37	
c	Carrier	[illegible]	2.7
pc	Carrier, predicted	42	
n	RNA	156	3.5
lp	Lipoprotein	46	1.0
cp	Cell process	56	1.3
l	Leader peptide	11	0.3
su	Pseudogenes in common	74	1.6
i	Site (*oriC*)	1	<0.1
h	Phage/IS in common (including 15 pseudogenes)	304	6.8
d	Partial information	146	3.3
o	Unknown function	471	10.6
Total		4453	100.0

[1]Genes in common to strains MG1655 and W3110.
[2]The percentage is calculated from the sum of known and predicted gene types.

- About 10% of genes have unknown functions.
- Another 26% have "predicted" functions.
- Of these, roughly half might have metabolic roles.

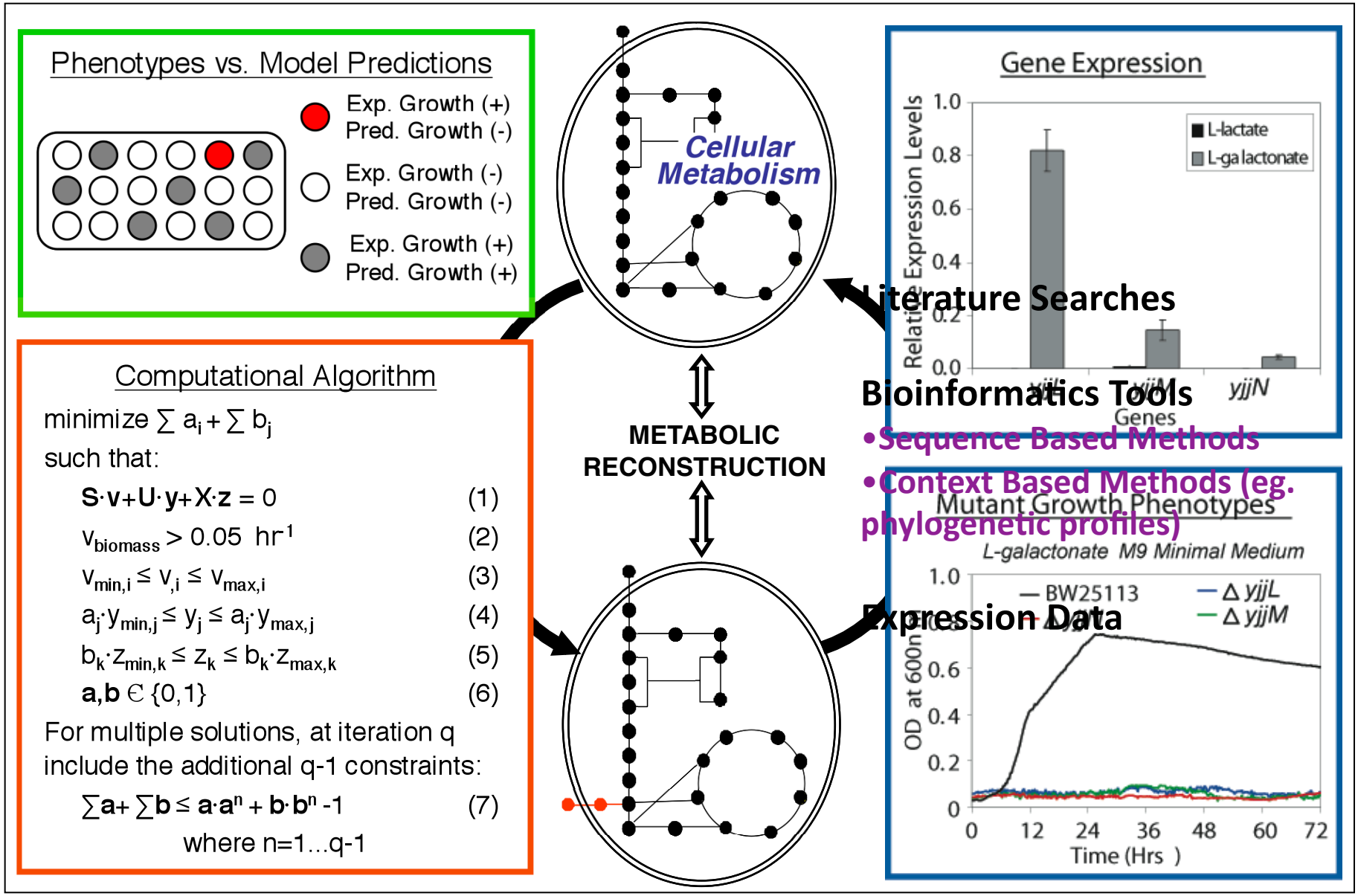
Model Driven Discovery Via High Throughput Testing



Growth Data Comparisons: Two Failure Modes

1. Predicted Growth but NO Experimental Growth
   - Missing regulation or falsely included reactions
2. Experimental Growth but NO Predicted Growth
   - Missing metabolic transport or enzymatic reactions
   - Incorrect regulation

Iterative Methods for Enzyme Identification



Reed et al. PNAS 103`(46):17480-4 (2006)

Computational Algorithm

minimize $\sum a_i + \sum b_j$

such that:

$$\mathbf{S}\cdot\mathbf{v}+\mathbf{U}\cdot\mathbf{y}+\mathbf{X}\cdot\mathbf{z} = 0 \quad (1)$$

$$v_{biomass} > 0.05\ hr^{-1} \quad (2)$$

$$v_{min,i} \le v_{,i} \le v_{max,i} \quad (3)$$

$$a_j \cdot y_{min,j} \le y_j \le a_j \cdot y_{max,j} \quad (4)$$

$$b_k \cdot z_{min,k} \le z_k \le b_k \cdot z_{max,k} \quad (5)$$

$$\mathbf{a},\mathbf{b} \in \{0,1\} \quad (6)$$

For multiple solutions, at iteration q include the additional q-1 constraints:

$$\sum\mathbf{a} + \sum\mathbf{b} \le \mathbf{a}\cdot\mathbf{a}^n + \mathbf{b}\cdot\mathbf{b}^n - 1 \quad (7)$$

where n=1…q-1

a,b are indicator variables of whether a reaction is allowed to occur.

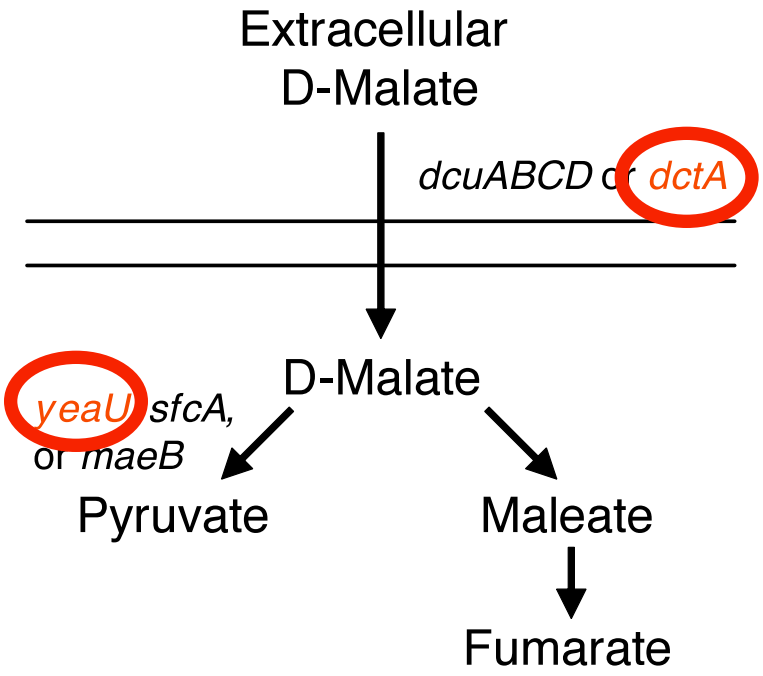
U, X are stoichiometric matrices for KEGG reations and “transport” reactions.

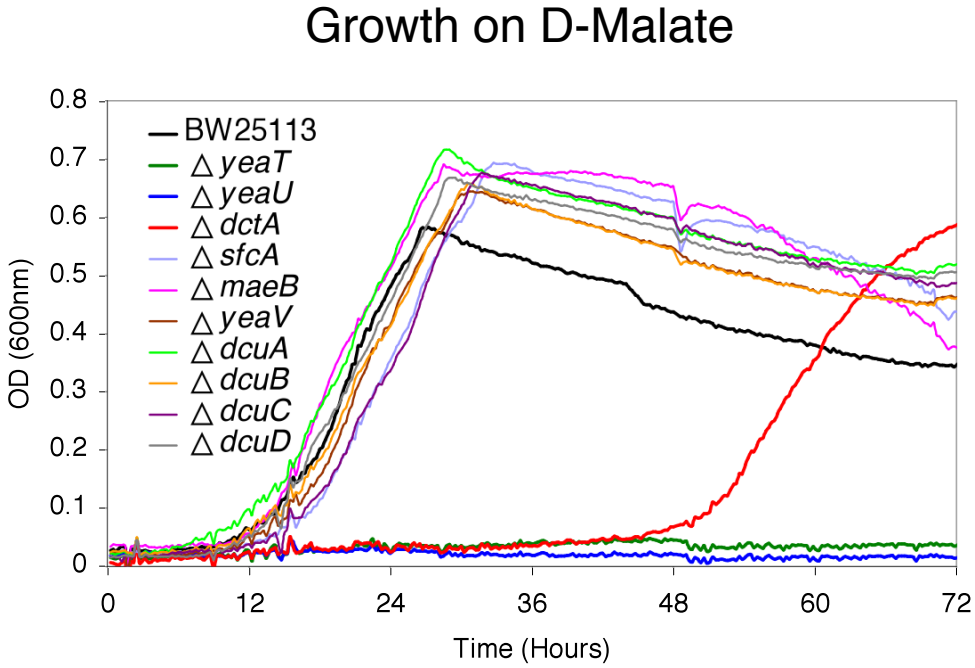
y, z are fluxes through these additional reactions.

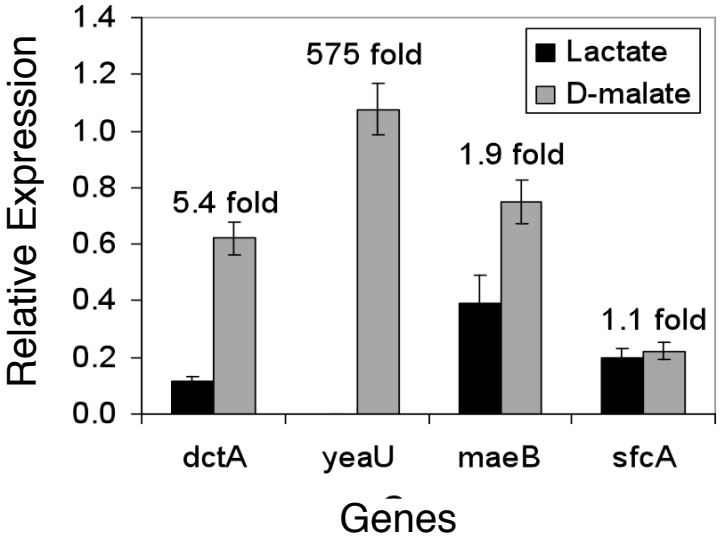
Equation 7 uses integer cuts so that we don’t revisit the same solutions

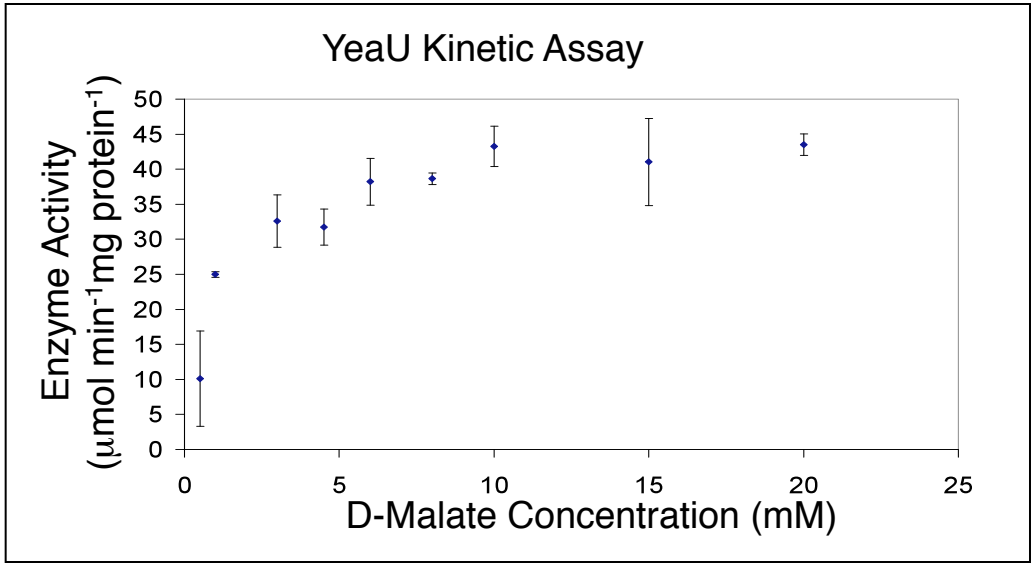
Reed et al. PNAS 103`(46):17480-4 (2006)

Case 1: Growth on D-Malate









GrowMatch: Correcting Under and Over Model Predictions

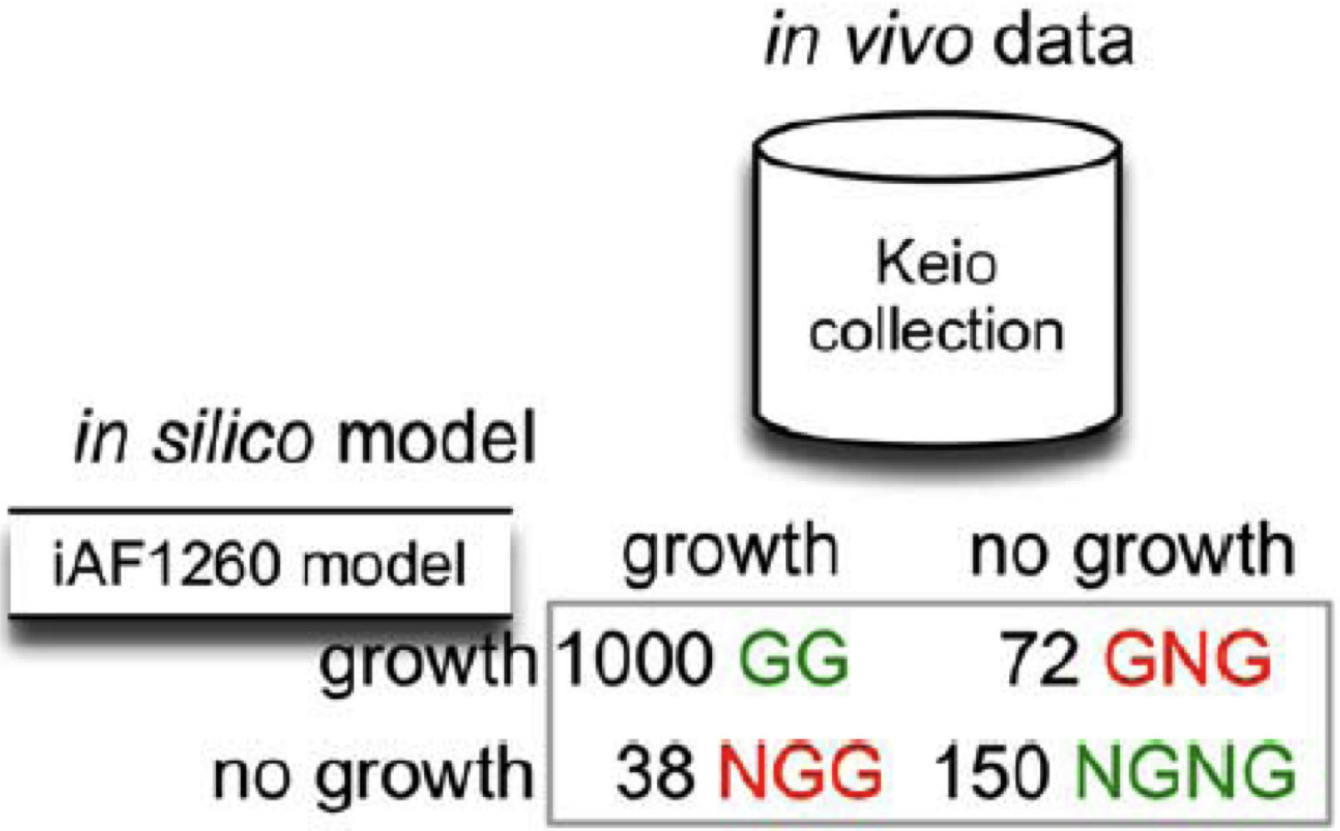
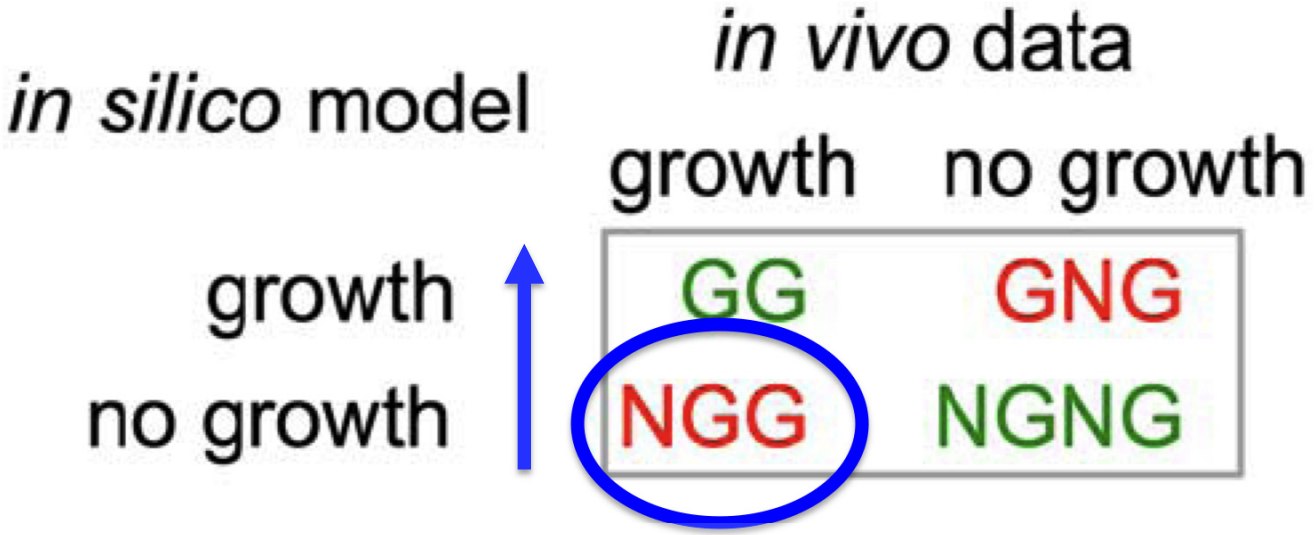


Figure 3. Classification of mutants based on comparison of *in silico* vs. *in vivo* data used in this study.

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

Comparison of Mutant Phenotypes



- Add isozymes
- Add new enzymatic or transport reactions
- Reaction directionality

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

Correcting NGG (model=NG/data=G) Discrepancies

$$Minimize \quad \sum_{j \in Database} y_j$$

$$s.t$$

$$v_j = 0, \qquad \forall \left| G^{nec}{}_{kj} = 1 \& k \in KO^{l^*} \right.$$

$$\sum_j S_{ij} v_j = 0_i, \qquad i = 1 \ldots M$$

$$v_{biomass} > v_{biomass}^{\min}$$

$$v_{atp} = v^{atp}$$

$$v_{uptake} = v^{uptake}$$

$$LB_j \le v_j \le UB_j \qquad \forall j \in Model$$

$$LB_j y_j \le v_j \le UB_j y_j \qquad \forall j \in Database$$

$$y_j = \{0,1\} \qquad \forall j \in Database$$

y are indicator variables of whether a database reaction is allowed to occur.

Database includes:

- KEGG & Metacyc reactions
- Transport reactions
- Reversible version of model reactions

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

38 Instances of NGG

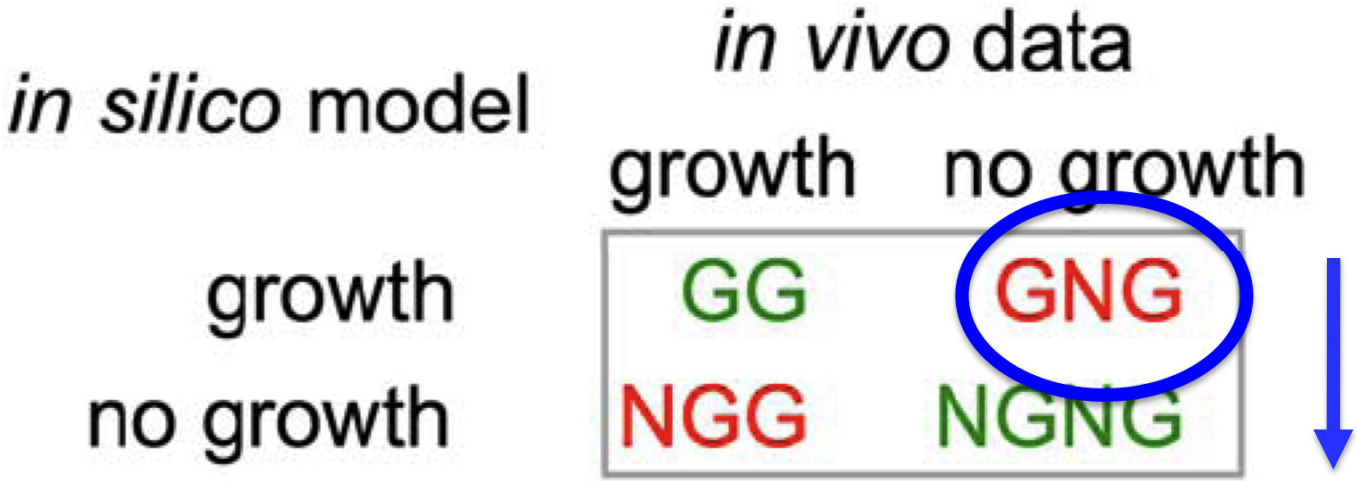
- 8 genes may have other genes that can compensate (e-value by BLAST<$1x10^{-3}$)
- Secretion of products may explain 3 discrepancies
- Addition of reactions may explain 3 discrepancies

Table 4. Resolution of NGG mutants by allowing secretion of metabolites.

NGG	Secreted Metabolite
Mutant	
ΔaldA	glycoaldehyde
ΔluxS	S-Ribosyl-L-homocysteine
ΔfolD	3,4-dihydroxy-2-butanone 4-phosphate

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

Comparison of Mutant Phenotypes



- Remove Reactions
- Remove Isozymes
- Add Metabolites to Biomass

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

Correcting GNG (model=G/data=NG) Discrepancies

$$Minimize \quad v_{biomass}$$

$$s.t \quad Maximize \quad v_{biomass} \quad [\text{Inner}]$$

$$\begin{bmatrix} \sum_{j} S_{ij} v_j = 0 \quad i = 1 \ldots M \\ v_{atp} = v^{atp} \\ v_{uptake} = v^{uptake} \\ LB_j y_j \le v_j \le UB_j y_j \quad \forall j \in Model \end{bmatrix}$$

$$y_j = 0, \qquad \forall j \left| G_{kj}^{nec} = 1 \& k \in KO^{l^*} \right.$$

$$\sum_{j} (1 - y_j) \le n^*$$

$$y_j = \{0,1\} \qquad \forall j \in Model$$

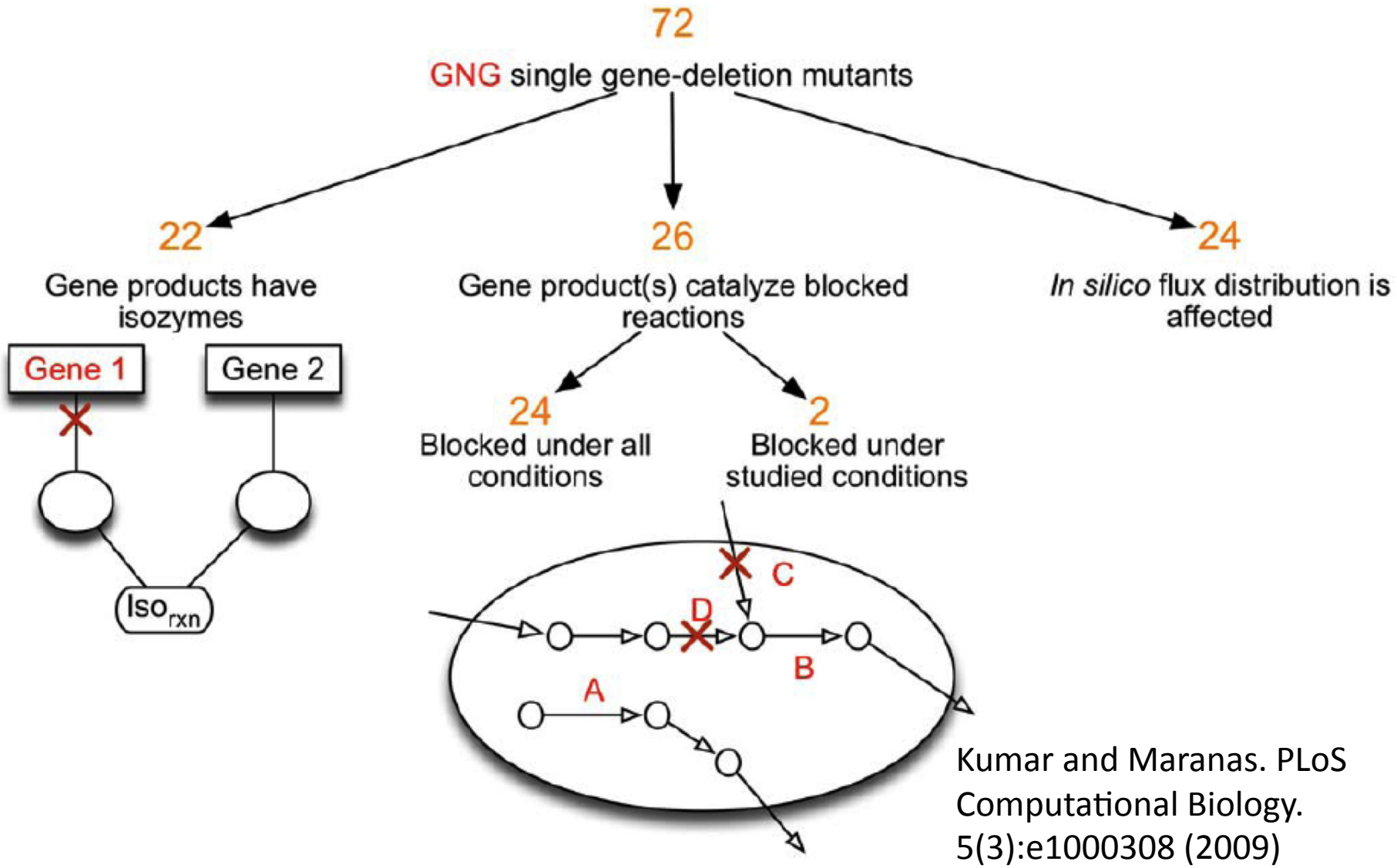
y are indicator variables of a model reaction is deleted (i.e. v_j=0).

Outer problem deletes reactions so that the maximum biomass is the lowest (i.e. growth=0)

Inner problem calculates maximum biomass given deleted reactions chosen by the outer problem

Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

72 GNG Discrepancies

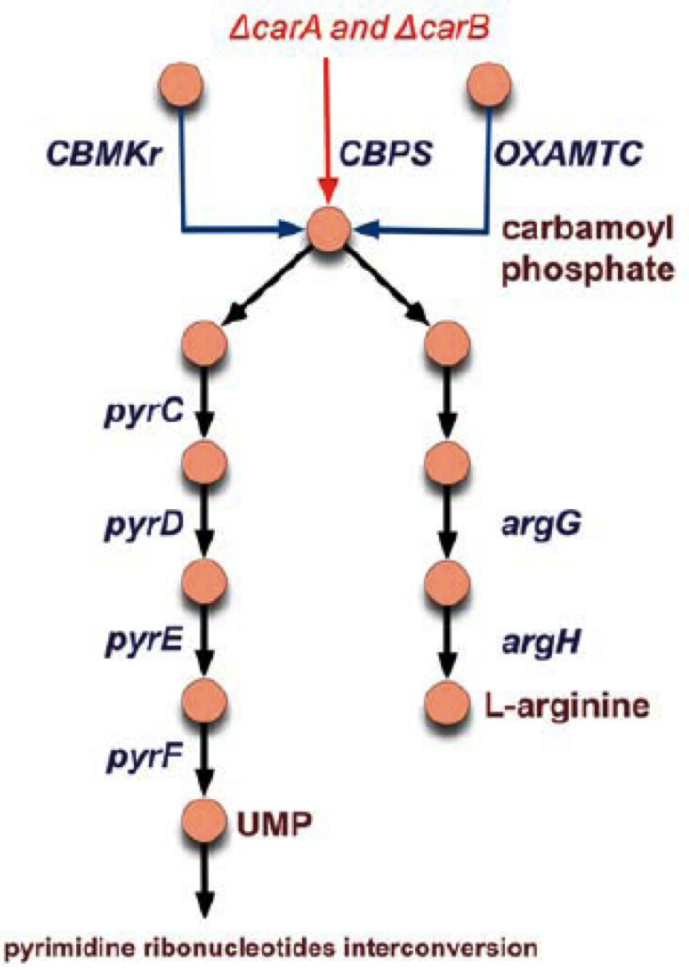


Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

72 GNG Discrepancies

Table 2. Resolution of GNG mutants in which deleted genes encoding for isozymes.

GNG Mutant	Associated Essential Reaction (Pathway)
ΔaroE	SHK3Dr (Tyrosine, Tryptophan and Phenylalanine metabolism)
Δcan	HCO3E (Unassigned)
ΔddlB	ALAAlAr (Cell Envelope Biosynthesis)
ΔfabZ	12 reactions (Cell Envelope Biosynthesis)
ΔfolA	DHFR (Cofactor and Prosthetic Group Biosynthesis)
ΔftsI	MCTP1App (Murein Biosynthesis)
ΔglnA	GLNS (Glutamate metabolism)
ΔilvA	THRD_L (Valine, Leucine and Isoleucine metabolism)
ΔmetC	CYSTL (Methionine Metabolism)
ΔmetE	METS (Methionine metabolism)
ΔmetL	ASPK or HSDY (Threonine and Lysine metabolism)
ΔmrdA	MCTP1App (Murein Biosynthesis)
ΔthrA	ASPK or HSDY (Threonine and Lysine metabolism)
ΔubiD	OPHBDC (Cofactor and Prosthetic Group Biosynthesis)
ΔyshA	H2Otex (Transport, Outer Membrane)



Kumar and Maranas. PLoS Computational Biology. 5(3):e1000308 (2009)

Model Refinement

Adding Reactions to Expand Model

Working with SEED Models

Slightly Modified Version of SMILEY

$$\text{minimize} \sum a_k + \sum b_i$$

such that

$$S \cdot v + U \cdot y - z = 0$$

$$\text{LowerLimit}_j \le v_j \le \text{UpperLimit}_j$$

$$a_k \cdot y_{\min,k} \le y_k \le a_k \cdot y_{\max,k}$$

$$b_i \cdot z_{\min,i} \le z_i \le b_i \cdot z_{\max,i}$$

$$v_{Biomass} \ge 0.05$$

a,b are indicator variables of whether a flux is allowed to occur.

U is a matrix of reactions.
y are fluxes these additional reactions.

z is a flux representing the removal of a metabolite from the system (pos value=removal and neg value=addition).

Note S and U must have the same number of rows which must be aligned!

SMILEY.gms

```
Variables
v(j) flux values through reaction in existing network
y(k) flux values through reaction in the database
z(i) flux values through transport reactions
Obj  number of needed reactions;

Binary Variable
a(k) binary variables
b(i) binary variables;
```

Binary variables indicating whether a reaction in genome-scale model is added (a=1) or if you need to uptake/secrete a metabolite (b=1)

```
Parameters
z_max(i) universal transport reaction maximum fluxes
z_min(i) universal transport reaction maximum fluxes;

z_max(i)=Vmax;
z_min(i)=0;
y_max(k)=Vmax;
```

Set Lower and Upper limits for uptake/secretion reactions. Zmin=0 means the metabolite can only be secreted.

Set Upper Limits for fluxes in Genome-scale model (lower limits y_min are defined in the EcoliMatrices.gms file)

```
* Define Compounds In Media That Aren't In Current Model
z_min('nh4_e')=-Vmax;
z_min('so4_e')=-Vmax;
```

Some metabolites are present in the media but don't have exchange fluxes in the smaller model. So we allow them to be taken up by setting z_min to be -Vmax

Constraints in SMILEY (no integer-cuts)

```
calcobj.. Obj=e=sum( k,a(k))+ sum( i,b(i));
massbalance(i).. sum( j,S(i,j)*v(j) )+ sum( k,U(i,k)*y(k) )  -z(i) =e=0;
ranges_up_y(k).. y(k)=l=( a(k)*y_max(k) );
ranges_low_y(k).. y(k)=g=( a(k)*y_min(k) );
ranges_up_z(i).. z(i)=l=( b(i)*z_max(i) );
ranges_low_z(i).. z(i)=g=( b(i)*z_min(i) );
growth.. v('Biomass')=g=0.05;
```

Calcobj: Calculates the number of a and b variables that are 1, and hence the associated fluxes that are non-zero

Mass balance: Now metabolite production and consumption can be balanced using reactions in genome-scale model (using y) or uptake/secretion into media (using z)

If a or b are zero then the associated fluxes must be zero using the ranges_up and ranges_low values.

Growth must be positive

OUTPUT: RequiredReactions.txt

```
"Model status:  ",1.00
"Solver status:  ",1.00
"Number of Added Fluxes",4.00

"Reactions from Universal Database"
"XYLI1",1.2337,-1.00,"xyl-D",1.00,"xylu-D"
"XYLK",1.2337,-1.00,"atp",-1.00,"xylu-D",1.00,"adp",1.00,"h",1.00,"xu5p-D"
"XYLabc",1.2337,-1.00,"atp",-1.00,"h2o",-1.00,"xyl-D_e",1.00,"adp",1.00,"h",1.00,"pi",1.00,"xyl-D"

"Transport Reactions (Pos=Secretion, Neg=Uptake)"
"xyl-D_e",-1.2337
```

Displays the Number of Fluxes you need to add

Tells you the name of the reaction, the flux value, and the reaction (stoichiometric coefficients followed by metabolite)

Tells you the name of the transport reaction and flux value (negative flux means the metabolite had to be added and positive flux means the metabolite needed to be consumed)

Example: Expand *E.coli* Core Model Using Reactions from Genome-Scale Model

- How many reactions do you need to add to the core model to get aerobic growth on fumarate (fum) as a carbon source?

- How many reactions do you need to add to the core model to get aerobic growth on arabinose (arab-L) as a carbon source?

Example: Expand *E.coli* Core Model Using Reactions from Genome-Scale Model

- How many reactions do you need to add to the core model to get aerobic growth on fumarate (fum)as a carbon source?
  - Since there already is an exchange flux in the core model, change carbon source use: LowerLimits('EX_fum_e')=-10;
  - No reactions are needed, meaning the core model can already do this
- How many reactions do you need to add to the core model to get aerobic growth on arabinose (arab-L) as a carbon source?
  - Since there already is no exchange flux in the core model, use: z_min('arab-L_e')=-10;
  - 5 reactions are needed, do these make sense?

Example: Expand *E.coli* Core Model Using Reactions from Genome-Scale Model

- How many reactions do you need to add to produce Hexadecanoate (hdca) a n-C16:0 fatty acid from glucose under aerobic conditions?

- Try and find another solution that does not secrete ppi (or does not use the PPA reaction).

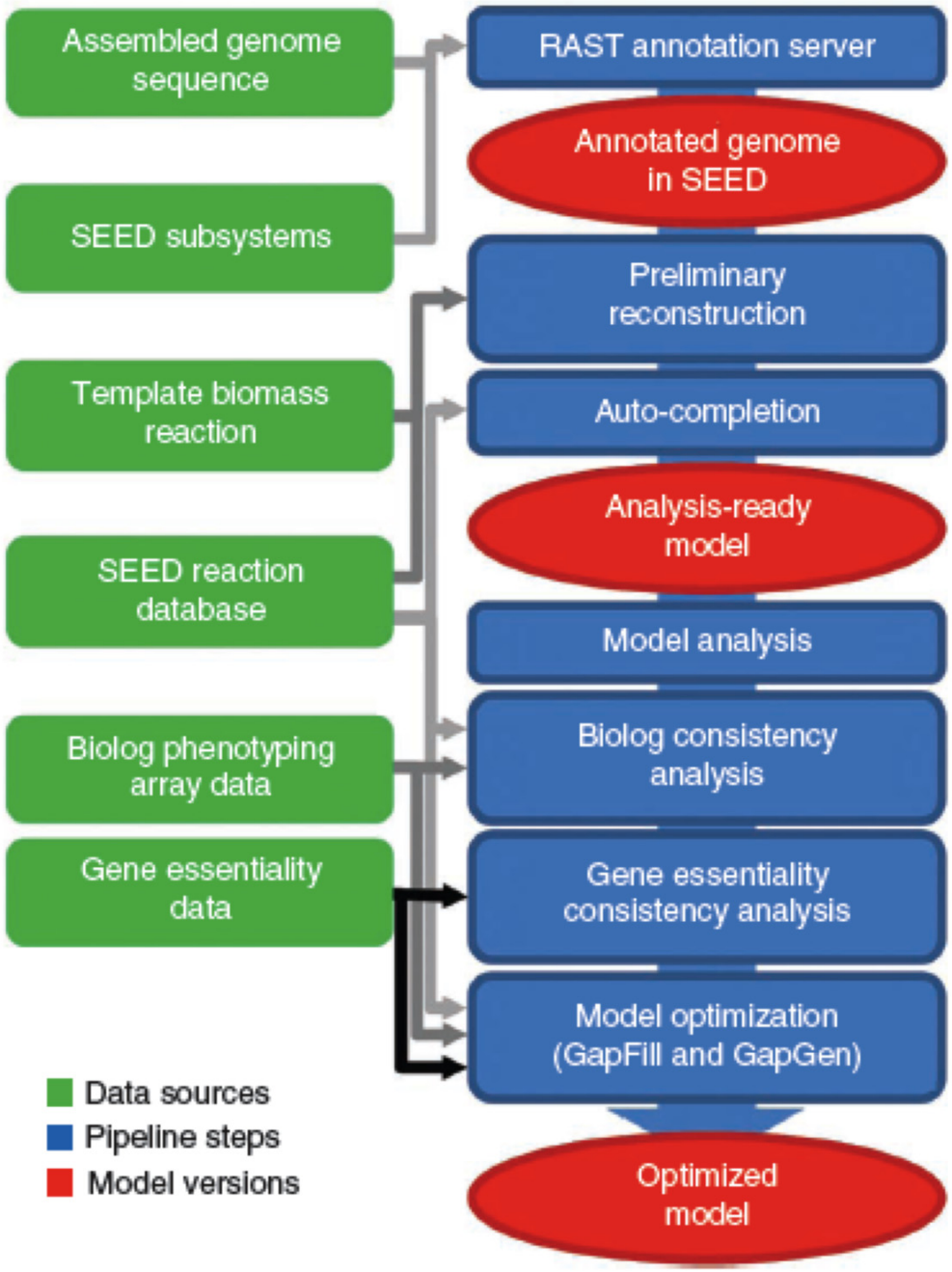
- How many reactions do you need to make biomass using the genome-scale biomass reaction (BiomassEcoli). Note make sure that nh4 and so4 are in the media by setting their corresponding z_min values to –Vmax;

Example: Expand *E.coli* Core Model Using Reactions from Genome-Scale Model

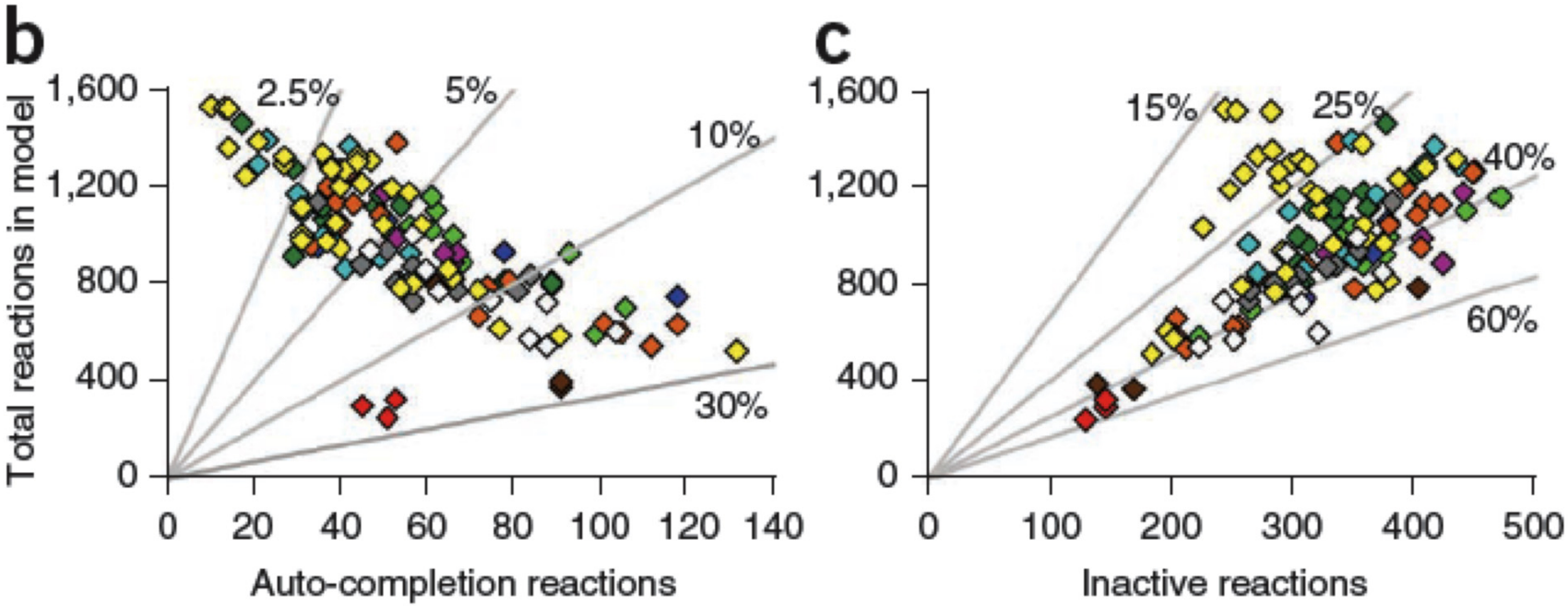
- How many reactions do you need to add to produce hdca from glucose under aerobic conditions?
  - Two Options: Change growth constraint to z('hdca')=g=1; or add line before the solve statement that z.lo('hdca')=1;
  - 21 reactions are needed.
- Try and find another solution that does not secrete ppi (or does not use the PPA reaction).
  - Two Options: add line before the solve statement that z.fx('hdca')=0; or y.fx('PPS')=0;
  - The two options are to either secrete ppi or convert ppi into (2)pi using the PPA reaction
- How many reactions do you need to make biomass using the genome-scale biomass reaction (BiomassEcoli). Note make sure that nh4 and so4 are in the media by setting their corresponding z_min values to –Vmax;
  - Two options: Change growth constraint to y('BiomassEcoli')=g=0.05; or add line before the solve statement that y.lo('BiomassEcoli')=0.05;
  - 227 Reactions are needed. These are all the biosynthetic pathways for amino acids, nucleotides, etc.

SEED Database

Overview of ModelSEED Process



How many reactions are needed to complete the models so they can predict growth? How many reactions can carry flux?



Henry et al. Nat Biotech. 28(9): 977-984 (2010).

Four Steps to Improve/Optimize Models

1. Evaluate consistency with Biolog data (growth phenotypes in different conditions) to identify missing transporters.
2. Evaluate consistency with gene essentiality data to find GPR conflicts.
3. Add reactions from universal database (transport, enzymatic, reversibility changes).
4. Create gaps to remove reactions.

Henry et al. Nat Biotech. 28(9): 977-984 (2010).

Adding Reactions To Model

$$\text{Minimize} \sum_{i=0}^{R} \left(1 + P_{T,i} + P_{K,i} + P_{SS,i} + P_{F,i} - f_{SS,i} - f_{p,i}\right) z_i$$

z_i – binary variable indicating whether reaction is added
$P_{T,i}$ – penalty for transport rxns (4=biomass or 2=non-biomass)
$P_{K,i}$ – penalty for non-kegg rxns (0=kegg or 2=non-kegg)
$P_{SS,i}$ – penalty for seed subsytems(0, 1 or 3)
$f_{SS,i}$ – bonus if other rxns in same subsystem are already in model
$f_{P,i}$ – bonus if other rxns in short linear pathway are already in model

Henry et al. Nat Biotech. 28(9): 977-984 (2010).

Access to ModelSEED

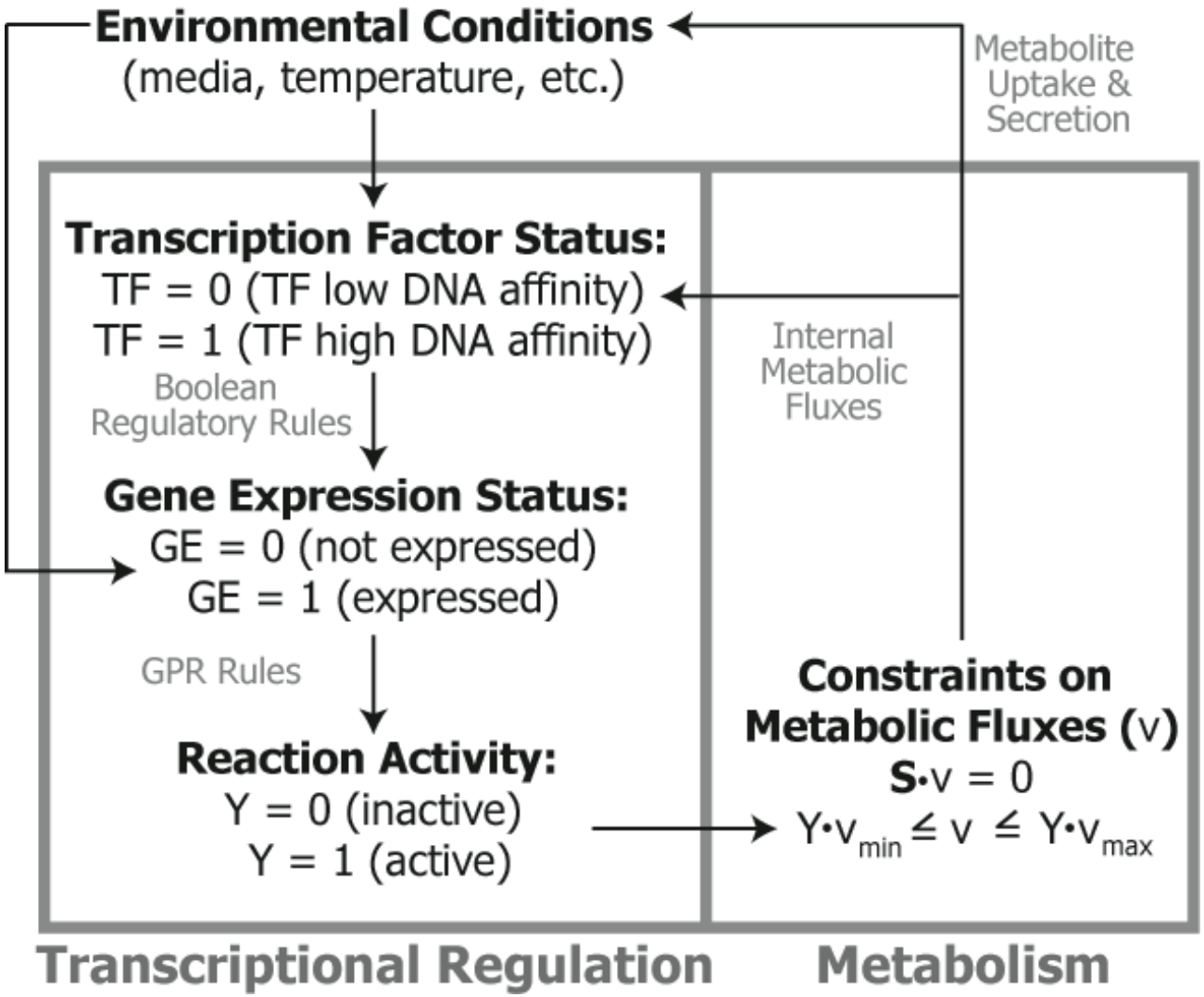
- http://seed-viewer.theseed.org/seedviewer.cgi?page=ModelView

- Recommend using firefox for ModelSEED

- If you want to have private models you will need to have your own account.

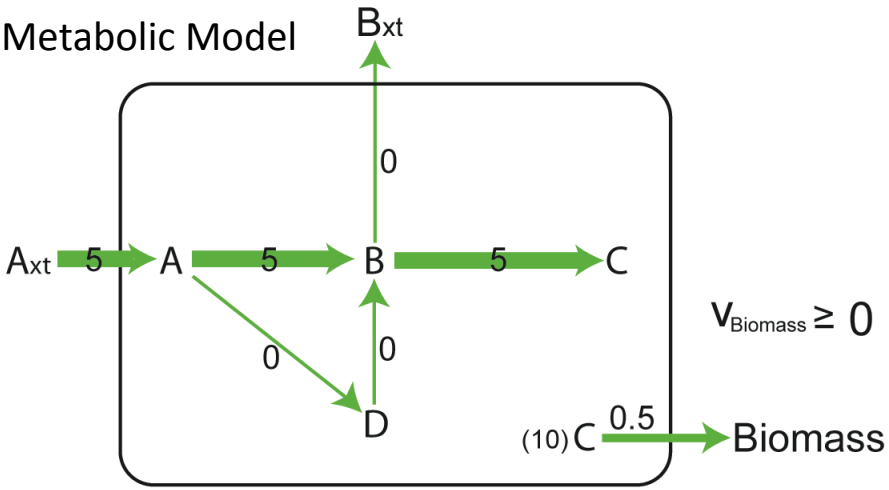
What About Transcriptional Regulatory Models

Integrated Models of Metabolism and Regulation



Approach for Relaxing Regulatory Constraints to Improve Accuracy

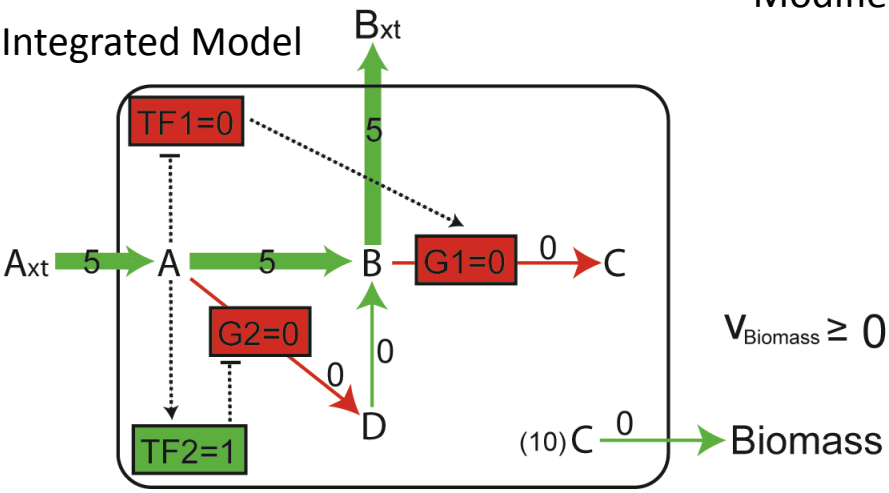
Metabolic Model



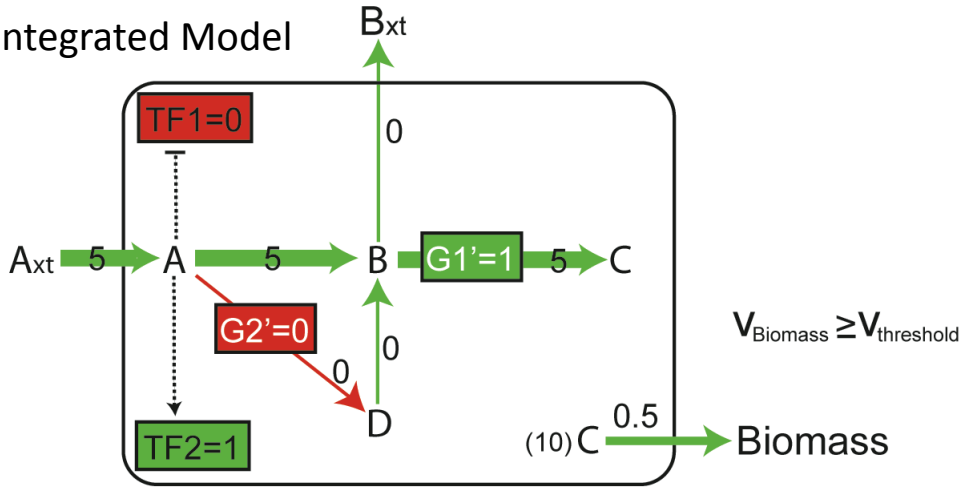
OPTIMIZATION PROBLEM

Find the MINIMUM number of genes (currently OFF) that must be turned ON.

Integrated Model



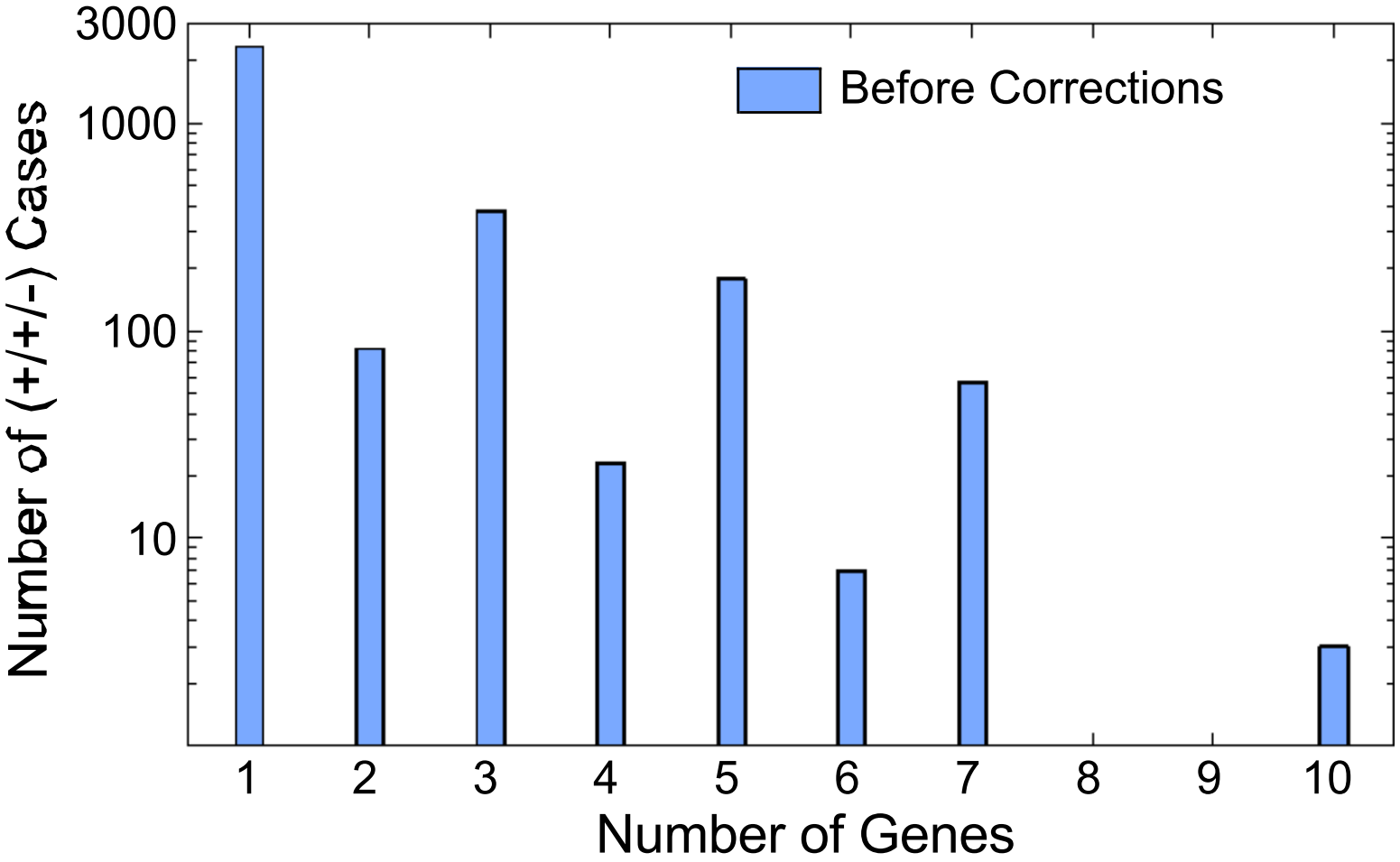
Modified Integrated Model



Barua, Kim and Reed. PLoS Comput Biol 6(10):e1000970 (2010)

How Many Changes Are Needed to Correct Each False Prediction?

Total of 3,079 Cases (+/+/-; exp/met/metreg)



11 Common *Gene*Force Corrections

Table 1. *E. coli* model refinements and the conditions under which they were identified by *Gene*Force.

Refinement Step	Gene	Original Rule	Refined Rule	Condition#	Comment
A	*metINQ*	(NOT MetJ)	GPR correction	Gly-Met (N) Met-Ala (N)	Unknown transporter for L-methionine (PMID: 4604763)
A	*glmU*	(NagC)	(ON)	N-acetyl-D-glucosamine (C,N) N-acetyl-D-mannosamine (C,N) N-acetyl-neuraminic acid (N)	Essential gene (PMID: 8407787)
A	*ilvY*	(NOT val-L(e)>0)	(ON)	b3773 (*ilvY*)	á-acetolactate or á-acetohydroxybutyrate inducer for *ilvY* (PMID: 10588699)
A	*ilvC*	(IlvY)	(IlvY AND NOT (val-L(e)>0)) OR (NOT IlvY)	b3773 (*ilvY*)	Constitutive expression of *ilvC* in *ilvY* strain (PMID: 6783625)
A	*sdaC**	(Crp AND (NOT Lrp OR (leu-L(e)>0)))	((Crp AND (NOT Lrp OR (leu-L(e)>0))) OR (ser-L(e)>0)	L-serine (N)	Transporters for ser-L; *sdaC* ser-L specific, *sstT* major, *tdcC* anaerobic (PMID: 8026499)
A	*cycA*	(NOT Lrp OR (leu-L(e)>0))	(NOT GcvB)	D-alanine (C,N)	No Lrp binding; CycA transporter for 6 amino acids (PMID: 19118351)
A	*gcvB*		(NOT GcvR AND GcvA)	D-alanine (C,N)	New regulatory small RNA (PMID: 10972807)
A	*dsdX*		GPR correction DsdC or (DsdC and Crp)	D-serine (C,N)	New ser-D transporter (This study, PMID: 16952954); regulation (PMID: 7592420)
A	*rpiR*	(NOT (rib-D(e)>0))	(NOT ((all-D(e)>0) OR (rib-D(e)>0)))	b2914 (*rpiA*)	*i*JR904 requires *rpiB* for *rpiA* strain (PMID: 10559180)
A	*acnA*	(SoxS)	(ON)	b0118 (*acnB*)	Two aconitases (PMID: 9202458)
A	*ilvA**	(NOT Lrp OR (leu-L(e)>0))	(ON)	b2797 (*sdaB*)	L-serine/L-threonine deaminases; SdaA (anaerobic), TdcB (anaerobic), IlvA (PMID: 13405870, 15155761)

E. coli's Regulation of D-Alanine Transporter

Experimental Result:
E.coli grows on D-alanine

Modeling Result:
Transporter is not expressed→ No growth

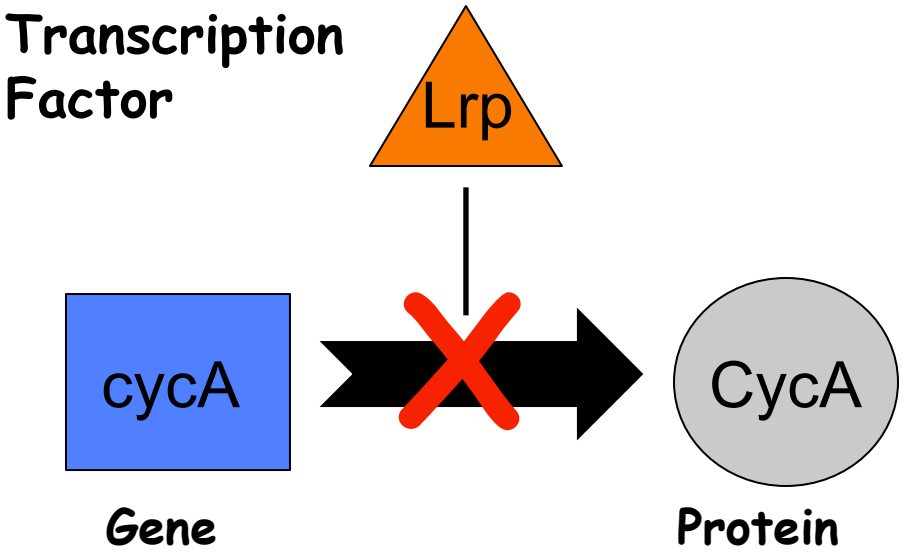
No Leucine

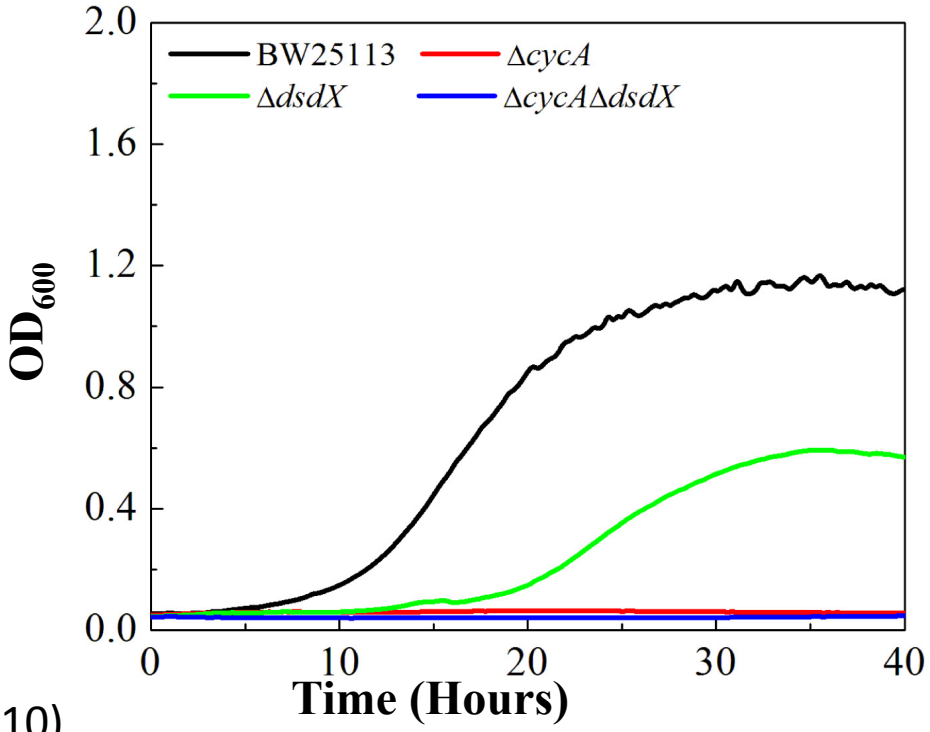
↓

Lrp Active

↓

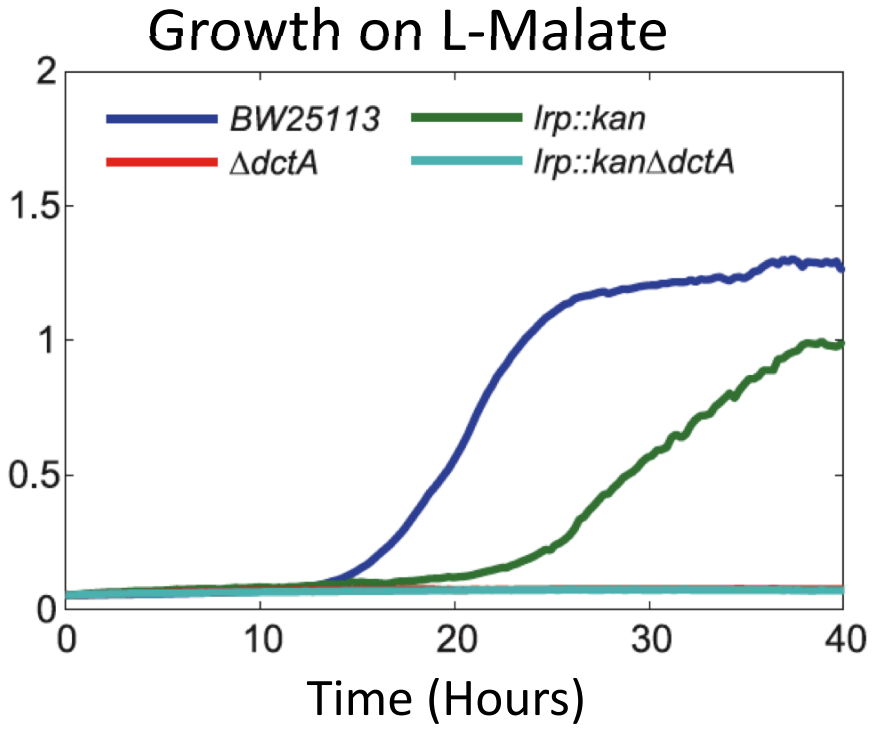
CycA Not Expressed

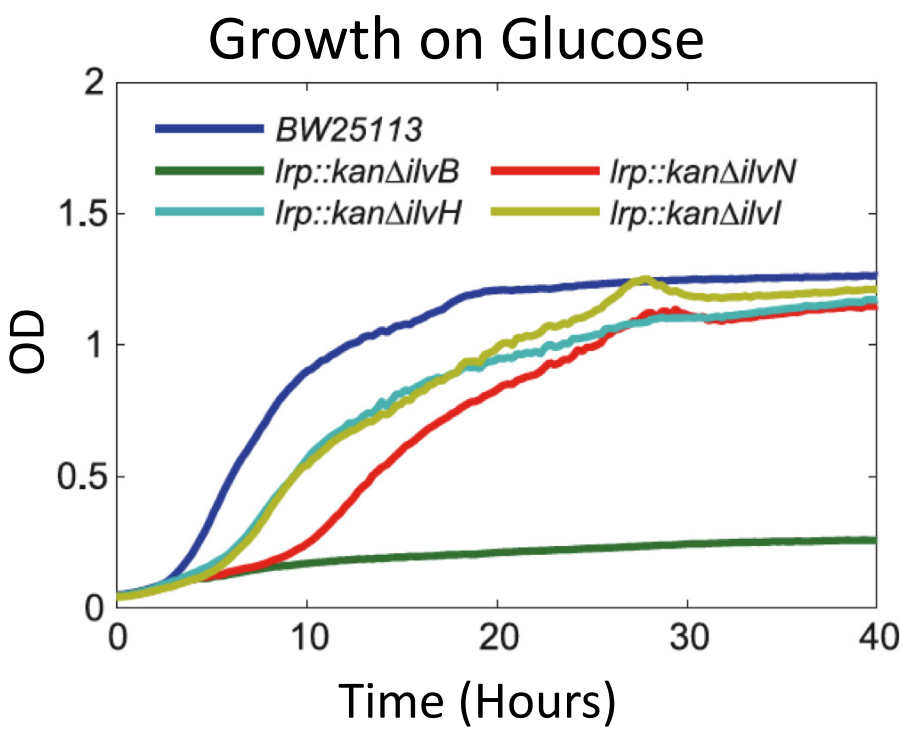




Barua, Kim and Reed. PLoS Comput Biol 6(10) (2010)

Other Experimentally Tested Corrections for Δ*lrp* Mutant

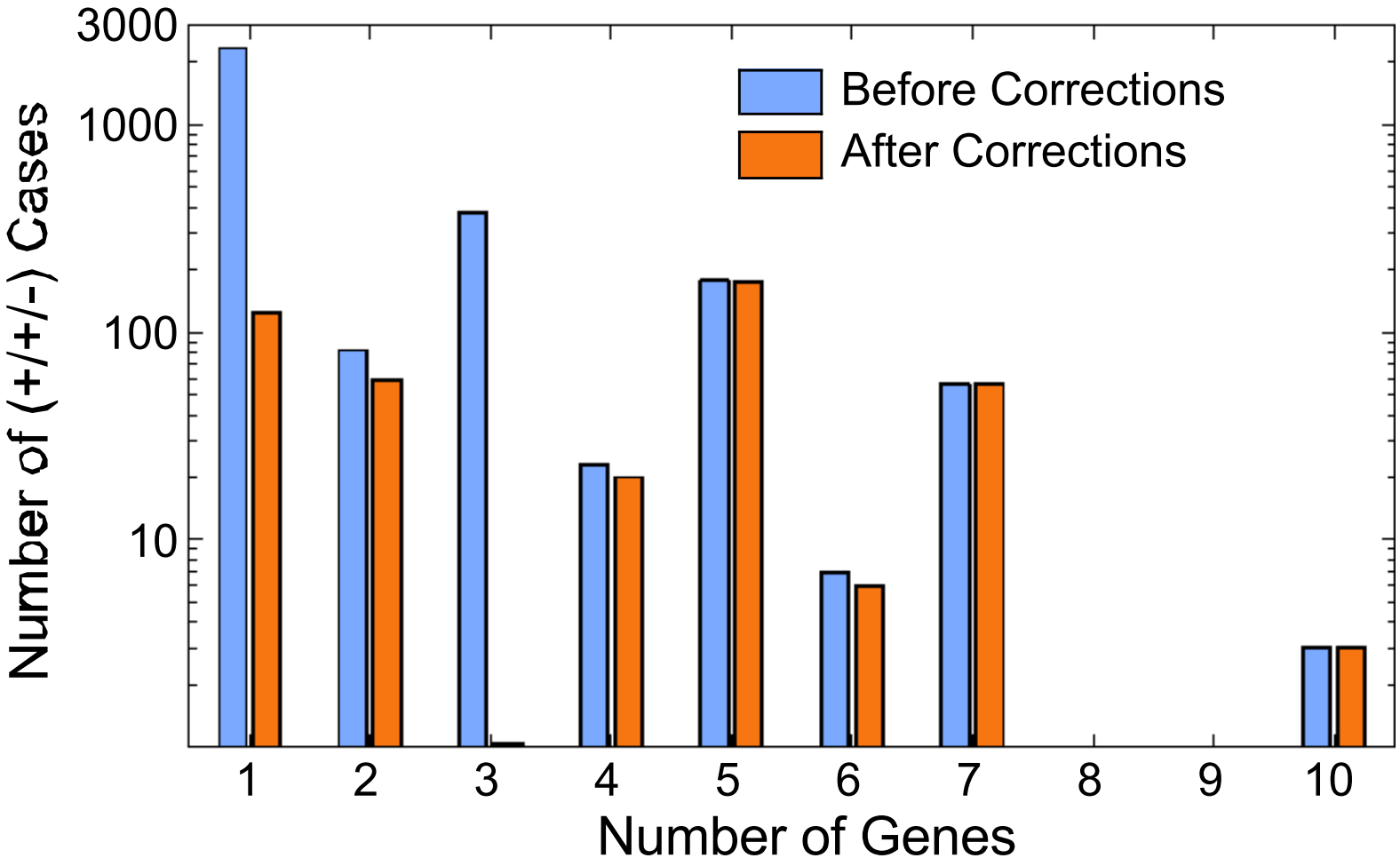




Gene	Original Rule	New Rule
dctA	(CRP NoMAN) AND NOT(ArcA) AND (DcuR)	ON
ilvB	NOT([leu]>0 OR [val]>0) AND Crp AND Lrp	ON
ilvN	NOT([leu]>0 OR [val]>0) AND Crp AND Lrp	ON

Effect of Model Corrections

Before 3,079 Cases; After 445 Cases



Rescue Non-Growth Phenotypes (cells can't grow due to regulation)

Table 3. Single genes or operons that are predicted to rescue non-growth phenotypes under aerobic conditions.

Media	Gene	Condition
Citrate	*citT*	Carbon Source
Sucrose	*xylA*	Carbon Source
1,2 propanediol	*fucO*	Carbon Source
Butyrate	*atoDAEB*	Carbon Source
L-tartrate	*ttdAB*	Carbon Source
Allantoin	*allC*	Nitrogen Source
Nitrite	*nirBD*	Nitrogen Source

Overview of Constraint-Based Modeling Sessions

1. Reconstructing metabolic networks and flux balance analysis
2. Finding alternate solutions and predicting the effects of gene knockout
3. Improving models using optimization
4. Using models for metabolic engineering

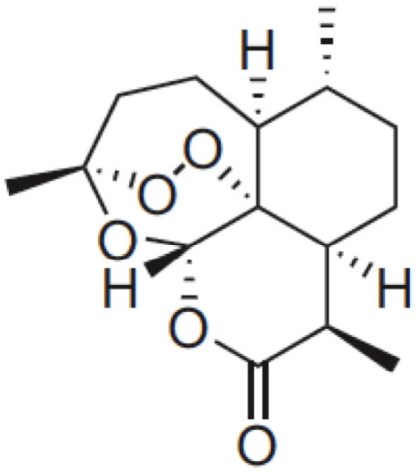
Metabolic Engineering

1. Knockout Prediction Tools (FBA, MOMA, ROOM)
2. OptKnock

Leveraging Biochemical Networks

Metabolic Engineering: Adjust metabolic behavior by engineering strains to produce useful chemicals

Drugs



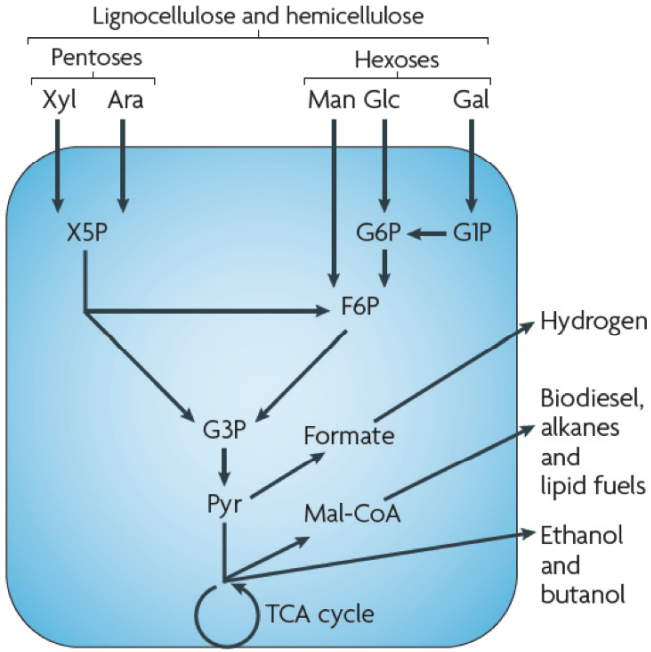
Artemisinin

Commodity Chemicals

DUPONT
The miracles of science™

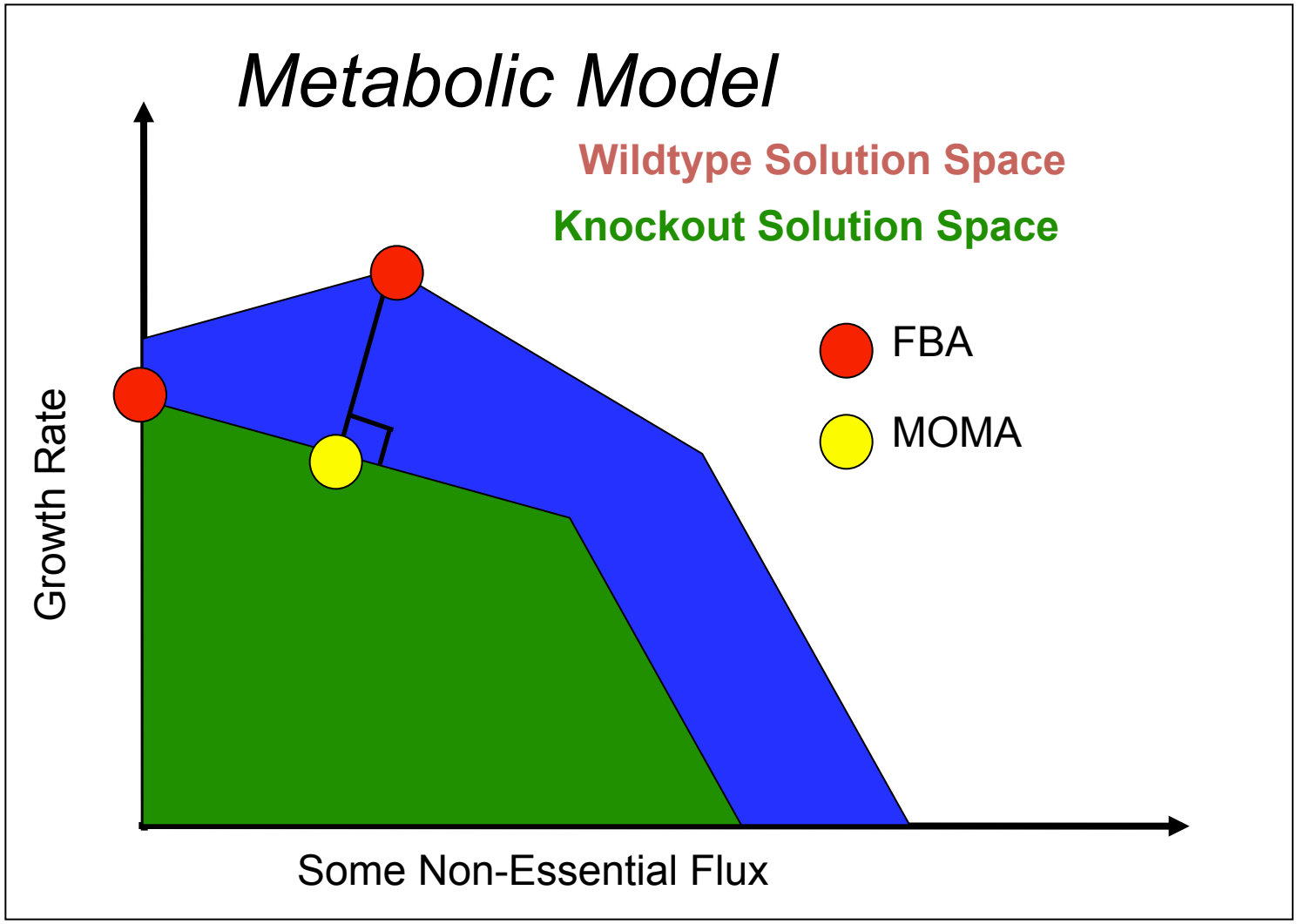
Sorona ®

Biofuels

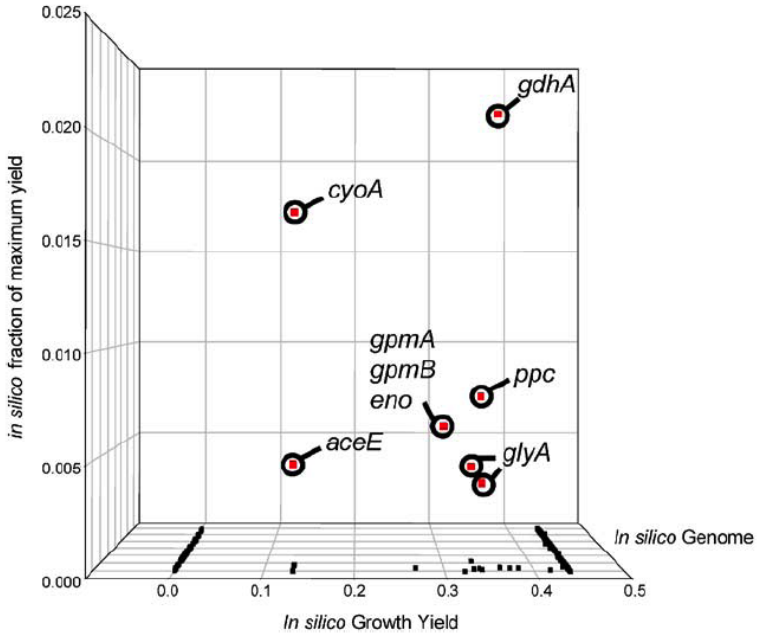


Alper & Stephanopoulos
Nat Rev Microbiol (2009)

MOMA: Minimize Distance Between Wildtype & Mutant Flux Distributions



MOMA for Increasing Lycopene Production



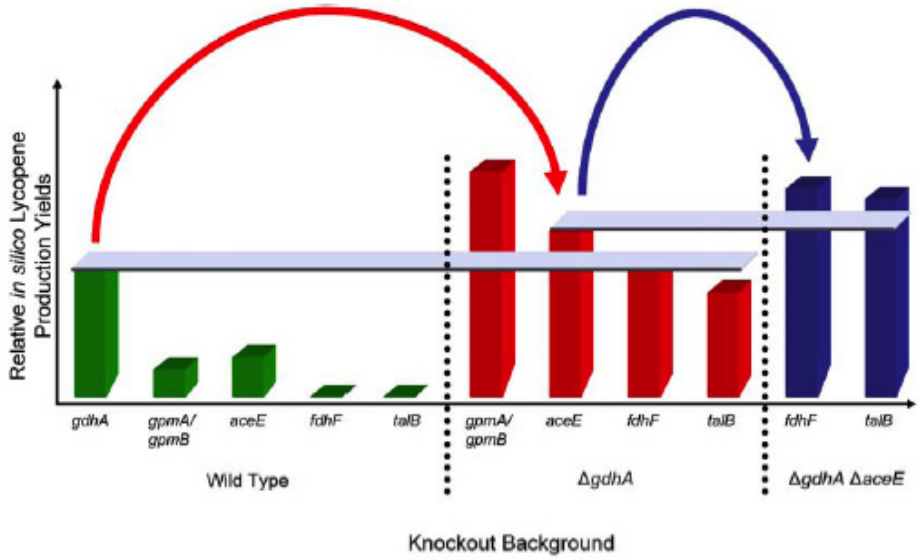


Table 1
Experimental results of single and multiple gene knockouts

Knockout construct	Growth rate	Percent increase in lycopene content (PPM)
None	0.67	0% (4700 PPM)
Single knockouts		
gdha	0.55	13% (±4)
gpma	0.44	−8% (±3)
gpmb	0.55	7% (±2)
acee	0.52	9% (±4)
fdhf	0.57	4% (±3)
Double knockouts		
gdhA, aceE	0.52	13% (±4)
gdhA, gpmA	0.37	12% (±3)
gdhA, gpmB	0.49	18% (±3)
gdhA, talB	0.46	3% (±4)
Triple knockouts		
gdhA, aceE, talB	0.44	19% (±4)
gdhA, aceE, fdhF	0.38	37% (±3) (6600 PPM)

Alper, Jin, Moxley, & Stephanopoulos. Metabolic Engineering. 7:155-64 (2005)

Improving Valine Production in *E. coli*

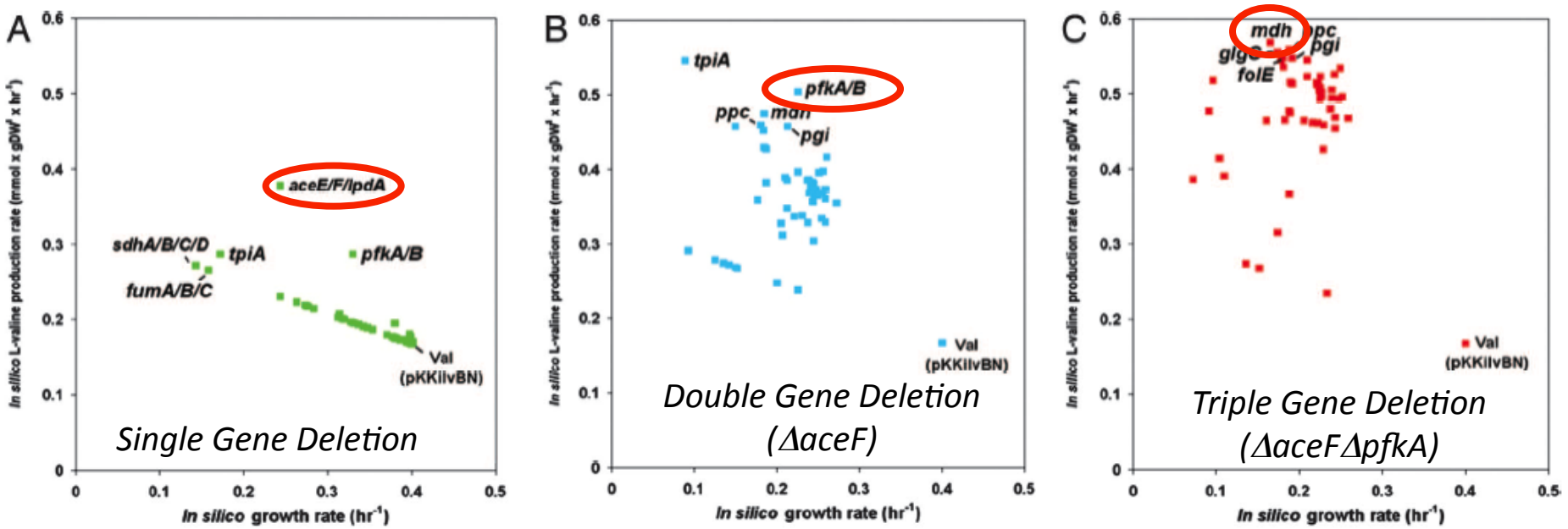


Fig. 3. Results of *in silico* gene knockout simulations by using the genome-scale metabolic model of *E. coli* MBEL979. The results of single (*A*), double (*B*), and triple (*C*) gene knockout simulations with respect to L-valine production and growth rates are shown. Only the five best candidates with respect to the L-valine production rate are shown for each stage of knockout simulation. Slashes indicate isoenzymes or subunits of the enzyme complex. The L-valine production and growth rates of the control Val strain harboring pKKilvBN are also indicated for comparison.

Park, J.H. Lee, K.H., Kim, T.Y., and Lee, S.Y. *PNAS*, 104(19):7797-7802 (2007).

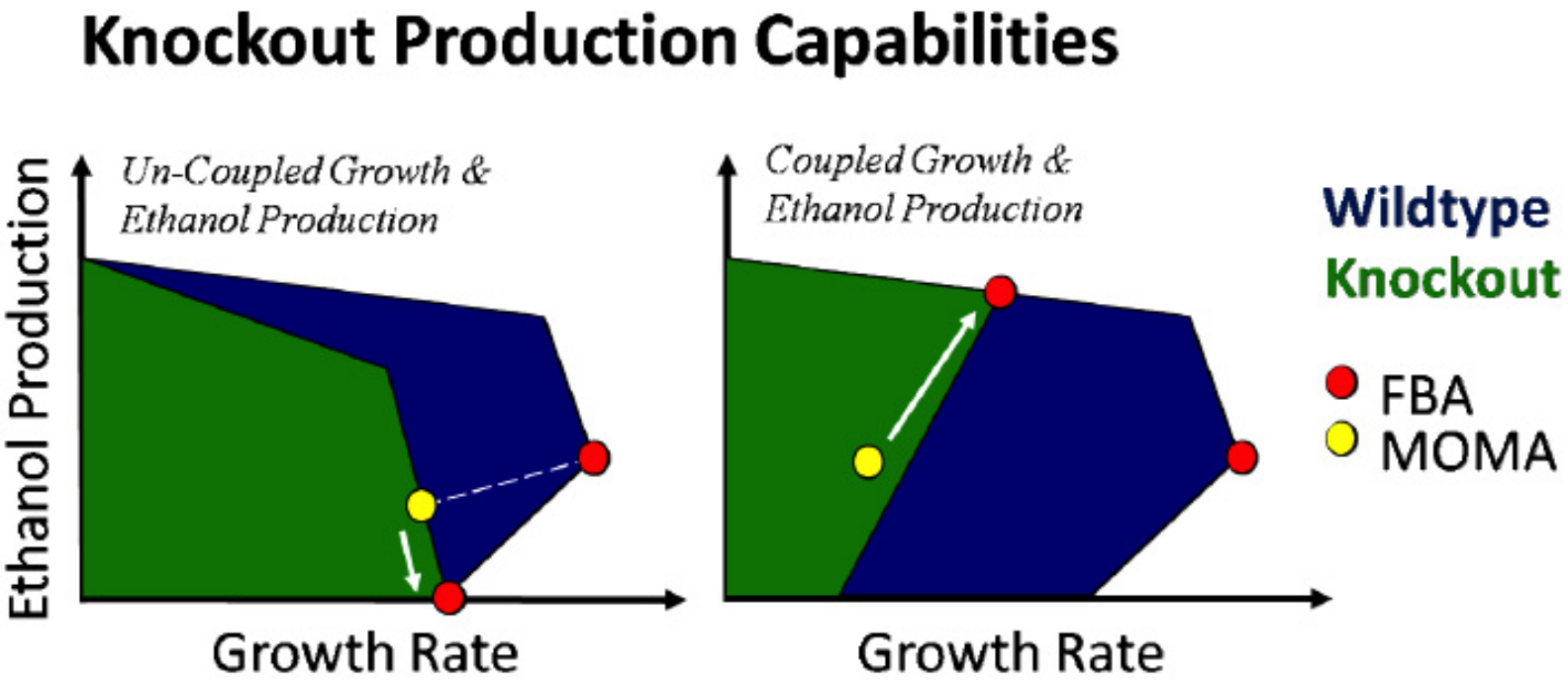
Model calculations led to an improved strain design for valine production (~2 fold increase in valine yields)

OptKnock:

- *Finds reactions, that if removed, couple of biomass production and metabolite production (ie. higher growth requires higher metabolite production levels)*

- *REFERENCES:*
  - *Burgard, Pharkya, Maranas. Biotechnology & Bioengineering. 84(6): 647-657 (2003)*
  - *Pharkya, Burgard, Maranas. Biotechnology & Bioengineering. 84(7): 887-899 (2003)*
  - *Pharkya, Burgard, Maranas. Genome Research. 14(11): 2367-76(2004)*
  - *Fong, et al. Biotechnology & Bioengineering. 91(5): 643-648*

OptKnock: Identifies Mutants with Coupled Biomass & Metabolite Production



Finds reactions, that if removed, couple biomass production to metabolite production (ie. higher growth =higher production)

So even if mutants initially have low production, by adaptively evolving strains using growth rate as selection pressure, the mutants should improve their productivity

Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

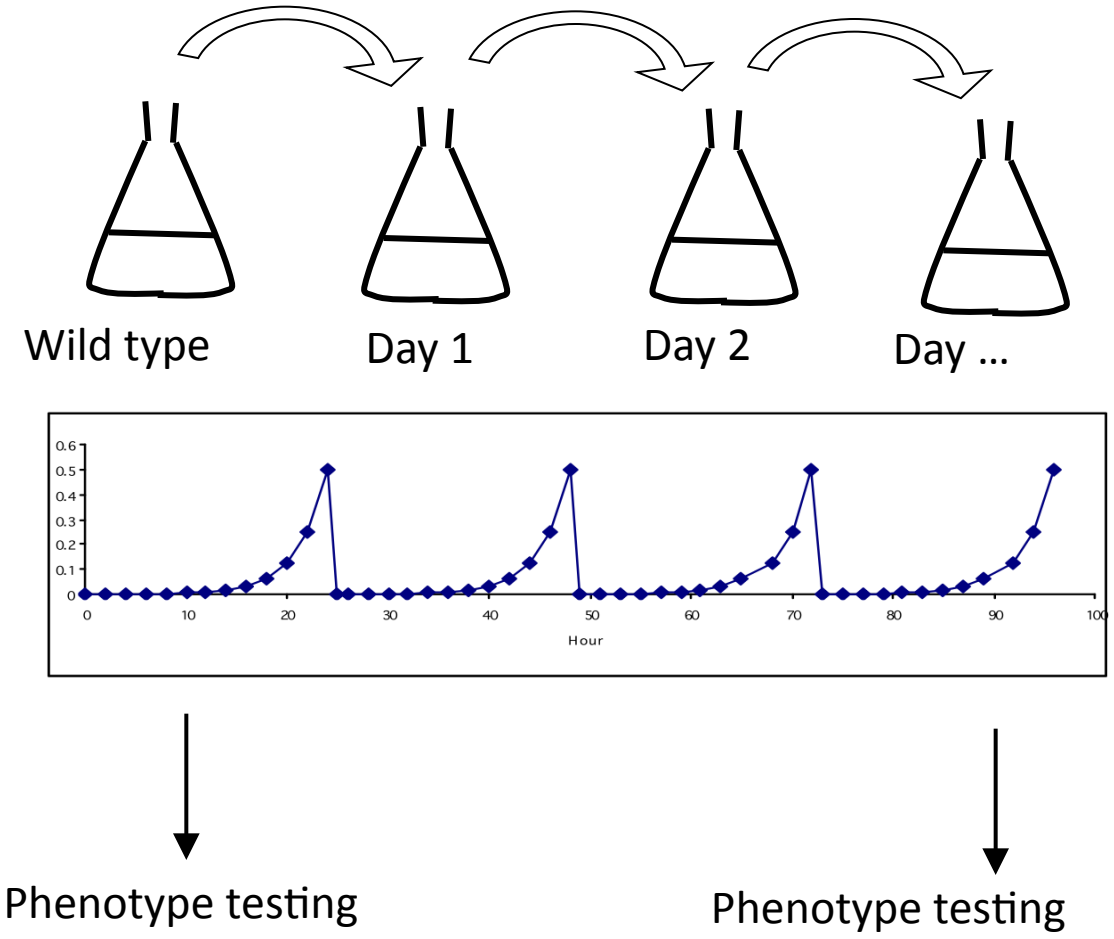
OptKnock Variations

- **OptStrain[1]:** Two step process, where (1) non-native pathways are identified that lead to product formation, and then (2) OptKnock is carried out to identified coupled phenotypes.
- **OptReg[2]:** Rather than consider reaction deletions, this also considers significant changes in fluxes.
- **OptGene[3]:** Uses genetic algorithms instead of optimization procedures to find the solutions.
- **OptORF[4]:** evaluates metabolic and regulatory gene deletions by gene and not reaction

1. Pharkya, Burgard, & Maranas, Genome Research, 14:2367-76 (2004)
2. Pharkya & Maranas, Metabolic Eng, 8:1-13 (2006)
3. Patil, Rocha, Forster & Nielsen, BMC Bioinformatics, 6:308 (2005)
4. Kim & Reed, BMC Systems Biology, 4:53 (2010)

Methods – adaptive evolution

- Cultures grown in 250ml minimal medium supplemented with 2g/L carbon source
- Serial passage during exponential growth
- Stable growth rate achieved at end of evolution
- Cells frozen throughout evolution for phenotype testing



OptKnock Problem Statement

$$\begin{aligned}
&\underset{y_j}{\text{maximize}} && v_{chemical} && \textbf{(OptKnock)}\\
&\text{subject to} && \underset{v_j}{\text{maximize}}\ v_{biomass} && \textbf{(Primal)}\\
& && \text{subject to} && \sum_{j=1}^{M} S_{ij} v_j = 0,\\
& && && v_{pts} + v_{glk} = v_{glc_uptake}\\
& && && v_{atp} \geq v_{atp_main}\\
& && && v_{biomass} \geq v_{biomass}^{target}\\
& && && v_j^{\min} \cdot y_j \leq v_j \leq v_j^{\max} \cdot y_j, \quad \forall j \in \mathcal{M}\\
& && && y_j = \{0,1\}, \quad \forall j \in \mathcal{M}\\
& && && \sum_{j \in M} (1 - y_j) \leq K
\end{aligned}$$

Cells have to grow

If a reaction (j) is removed, set y_j=0 so that v_j has to equal 0.

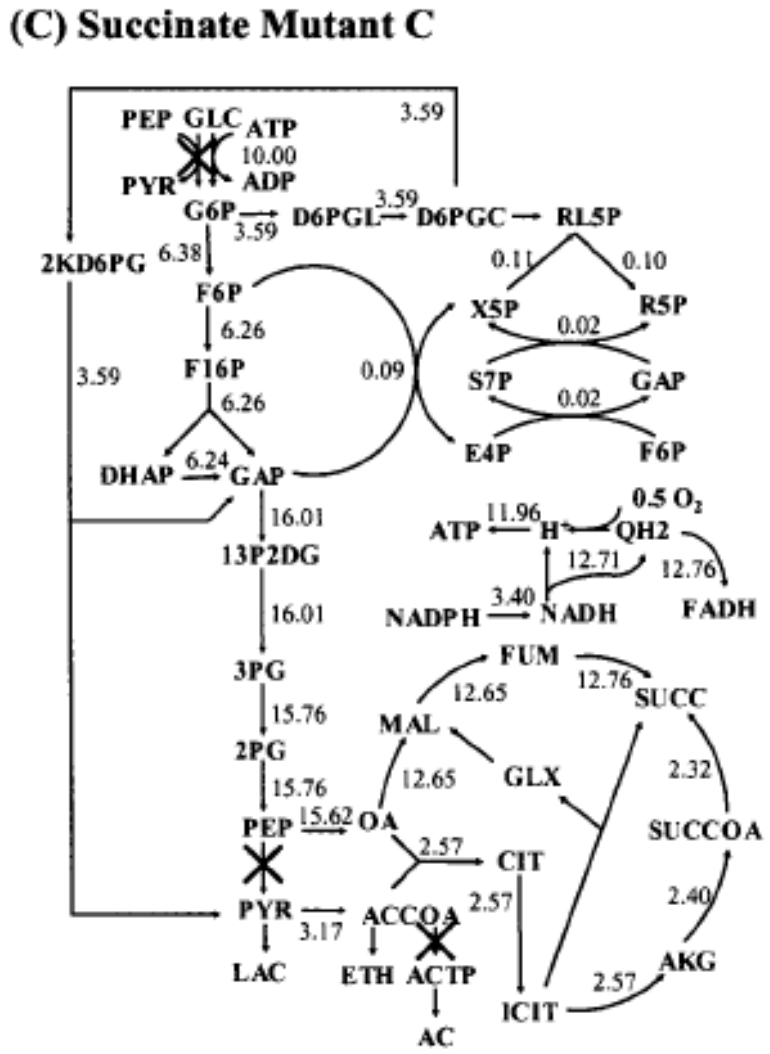
To solve this problem, you transform it by using the dual constraints for the primal problem, in addition to the primal constraints

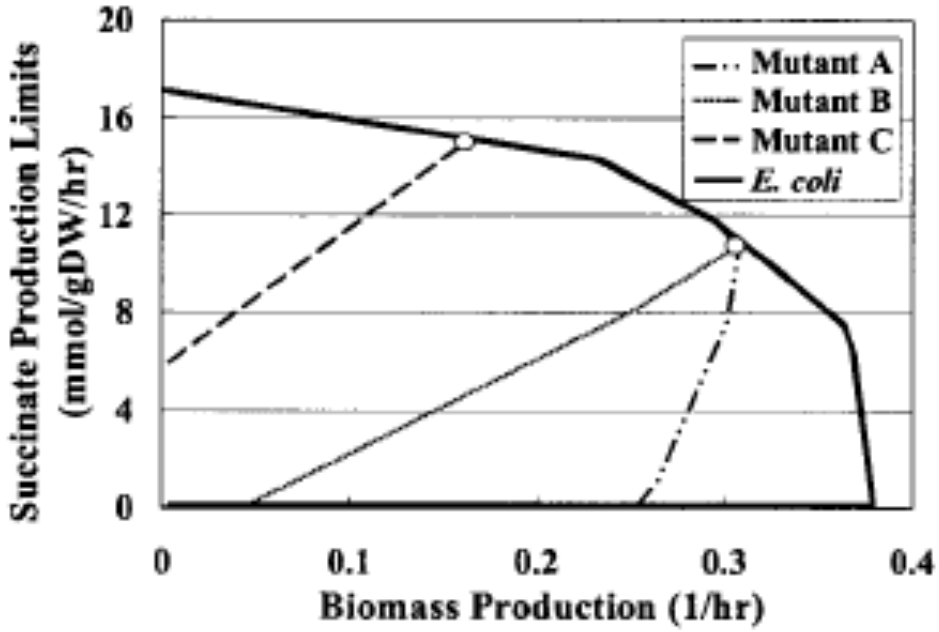
Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

Succinate				OptKnock		MOMA
ID		Knockouts	Enzyme	Biomass (1/hr)	Succinate (mmol/hr)	Succinate (mmol/hr)
Wild		''Complete network''		0.38	0.12	0
A	1	COA + PYR → ACCOA + FOR	Pyruvate formate lyase	0.31	10.70	1.65
	2	NADH + PYR ↔ LAC + NAD	Lactate dehydrogenase			
B	1	COA + PYR → ACCOA + FOR	Pyruvate formate lyase	0.31	10.70	4.79
	2	NADH + PYR ↔ LAC + NAD	Lactate dehydrogenase			
	3	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
C	1	ADP + PEP → ATP + PYR	Pyruvate kinase	0.16	15.15	6.21
	2	ACTP + ADP ↔ AC + ATP or	Acetate kinase			
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	3	GLC + PEP → G6P + PYR	Phosphotransferase system			

Lactate				OptKnock		MOMA
ID		Knockouts	Enzyme	Biomass (1/hr)	Lactate (mmol/hr)	Lactate (mmol/hr)
Wild		''Complete network''		0.38	0	0
A	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase	0.28	10.46	5.58
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	2	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
B	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase	0.13	18.00	0.19
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	2	ATP + F6P → ADP + F16P or	Phosphofructokinase			
		F16P ↔ GAP + DHAP	Fructose-1,6-biphosphatate aldolase			
C	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase	0.12	18.13	10.53
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	2	ATP + F6P → ADP + F16P or	Phosphofructokinase			
		F16P ↔ GAP + DHAP	Fructose-1,6-biphosphatate aldolase			
	3	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
	4	GLC + ATP → G6P + PEP	Glucokinase			

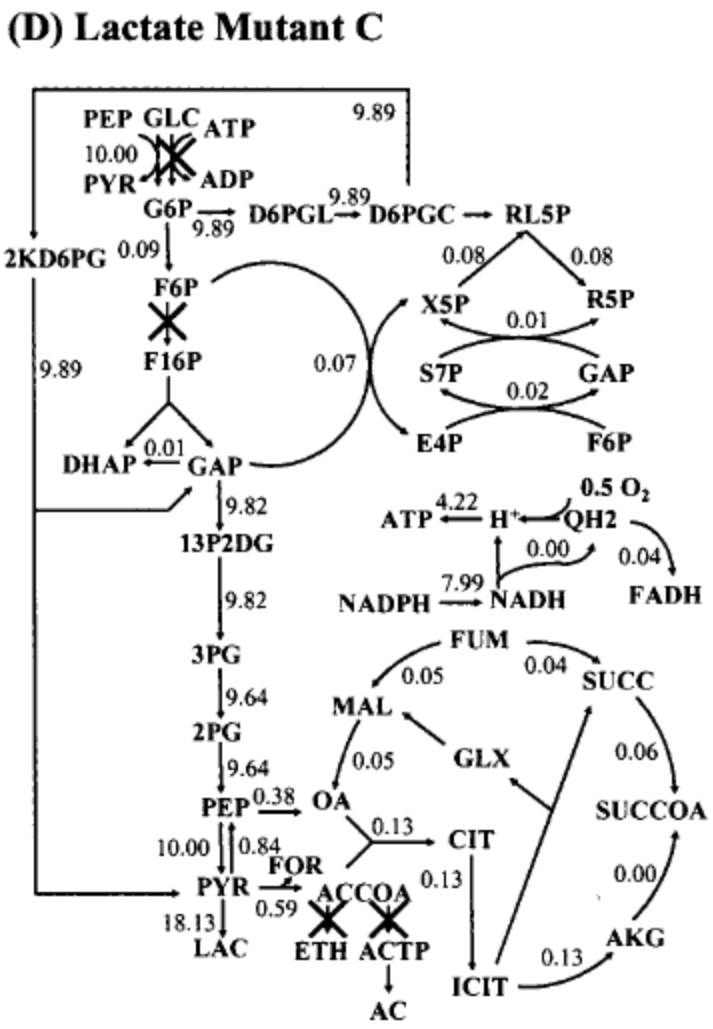
Succinate Production Strains

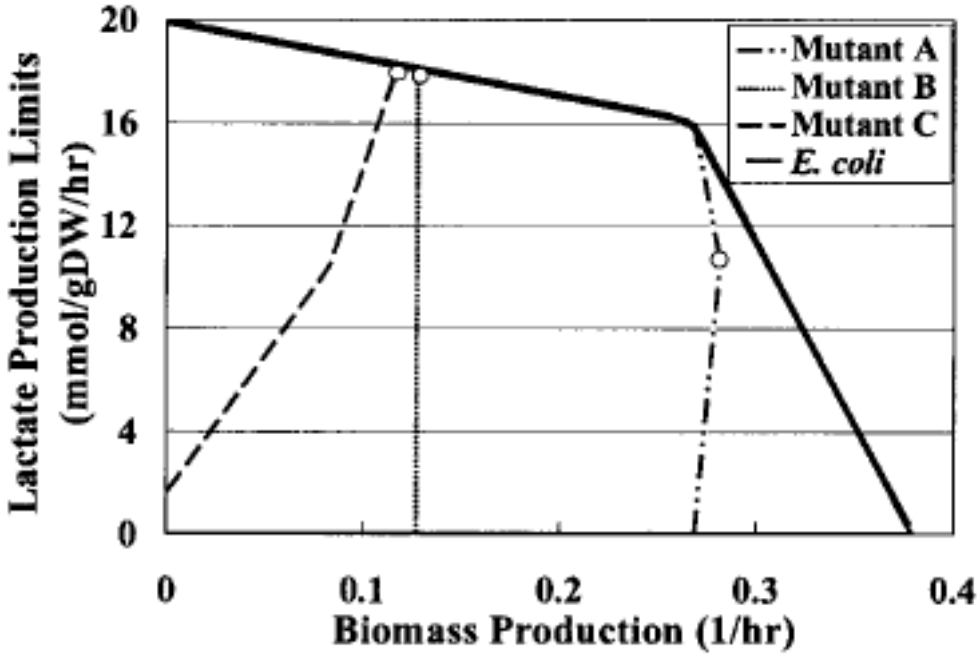




Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

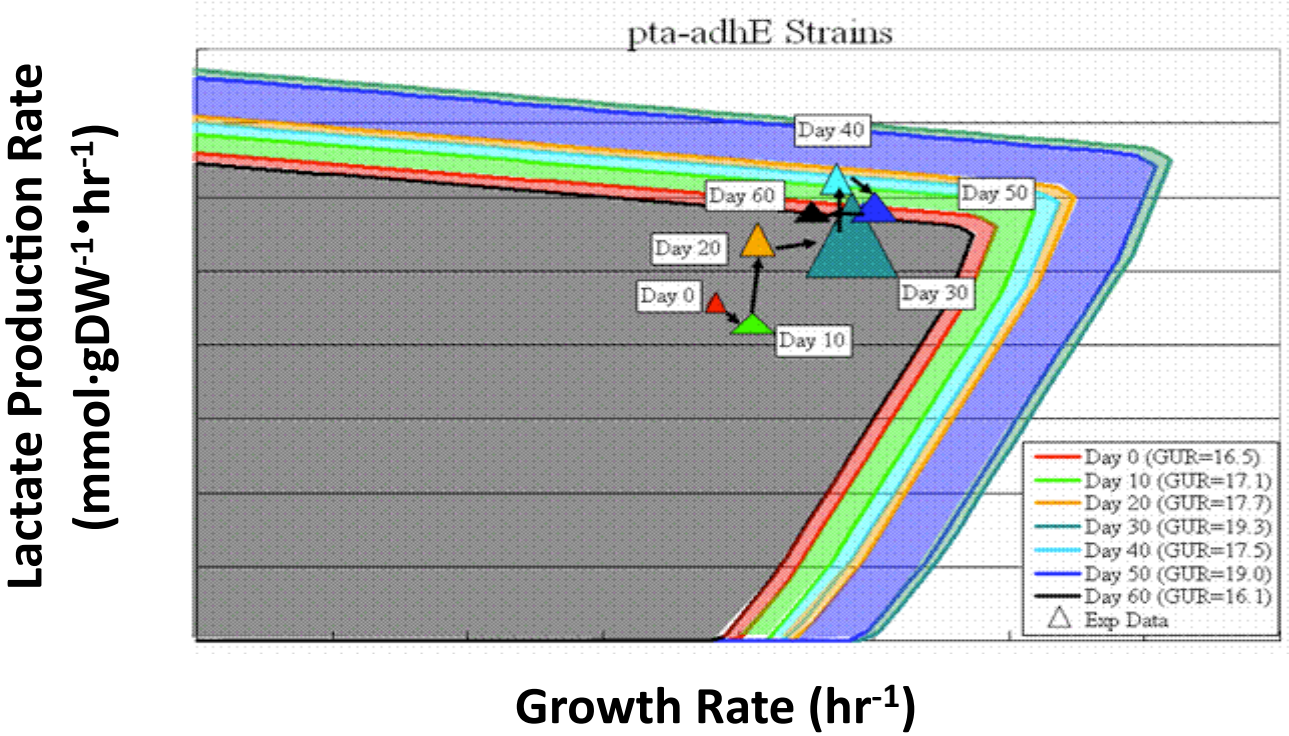
Lactate Production Strains



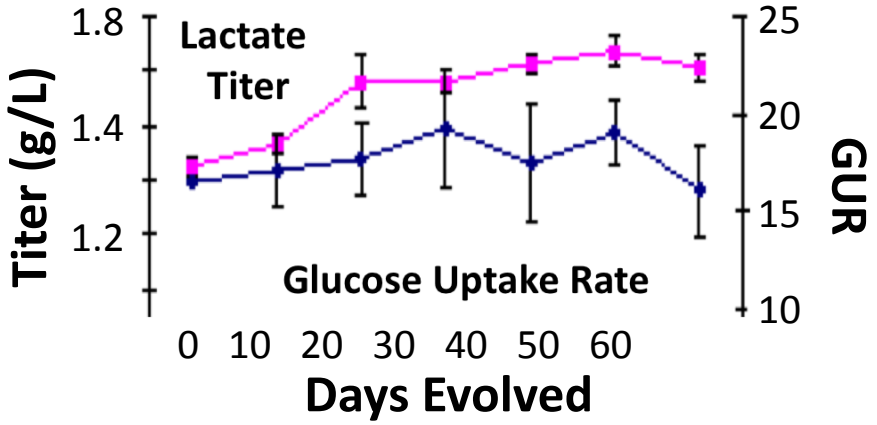


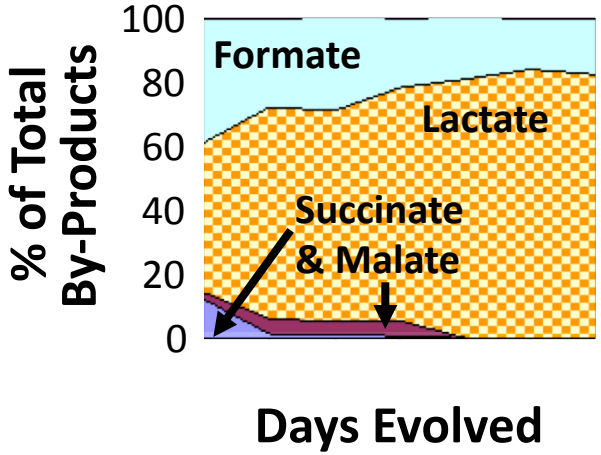
Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

Experimental Testing of a Lactate Strain



- Lactate secretion rate increased with increasing growth rate
- Lactate yield increased ~35% over 60 day evolution
- 2° by-product secretion decreased





S.S. Fong et al. B&B 2005

Metabolic Engineering

Ethanol Production

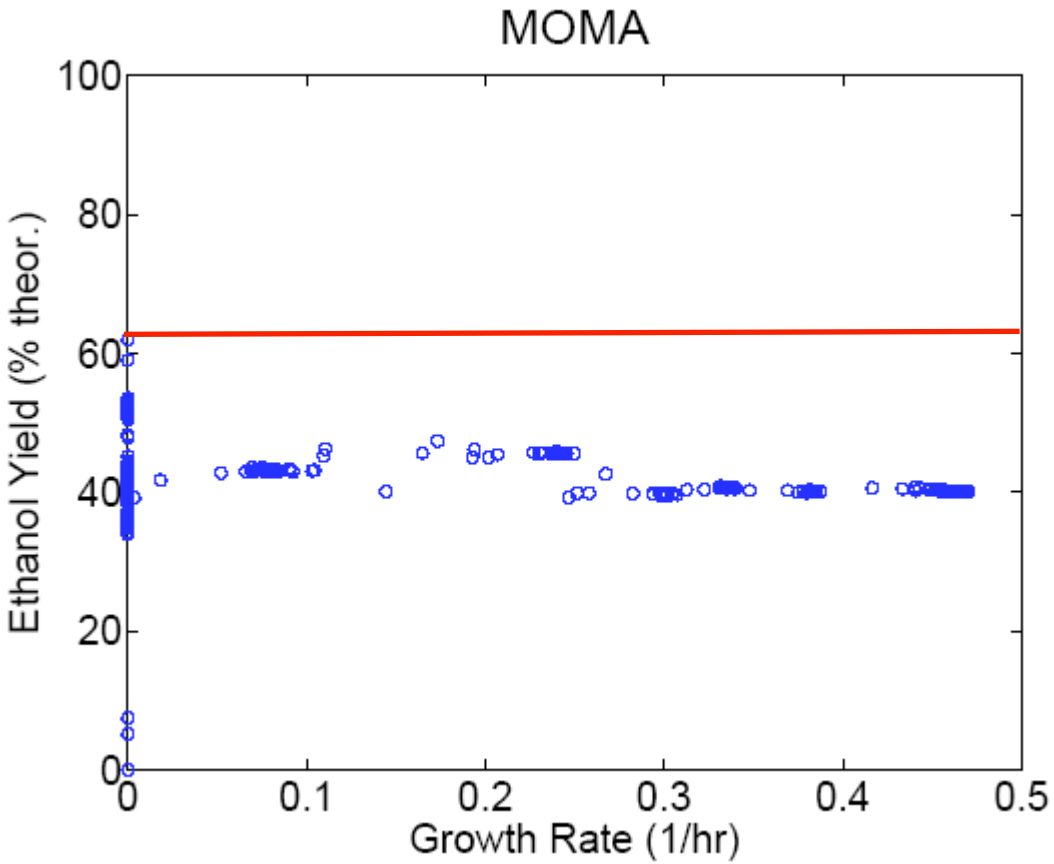
Exhaustive Search:
All Possible Double Gene Deletions

904 Metabolic Genes

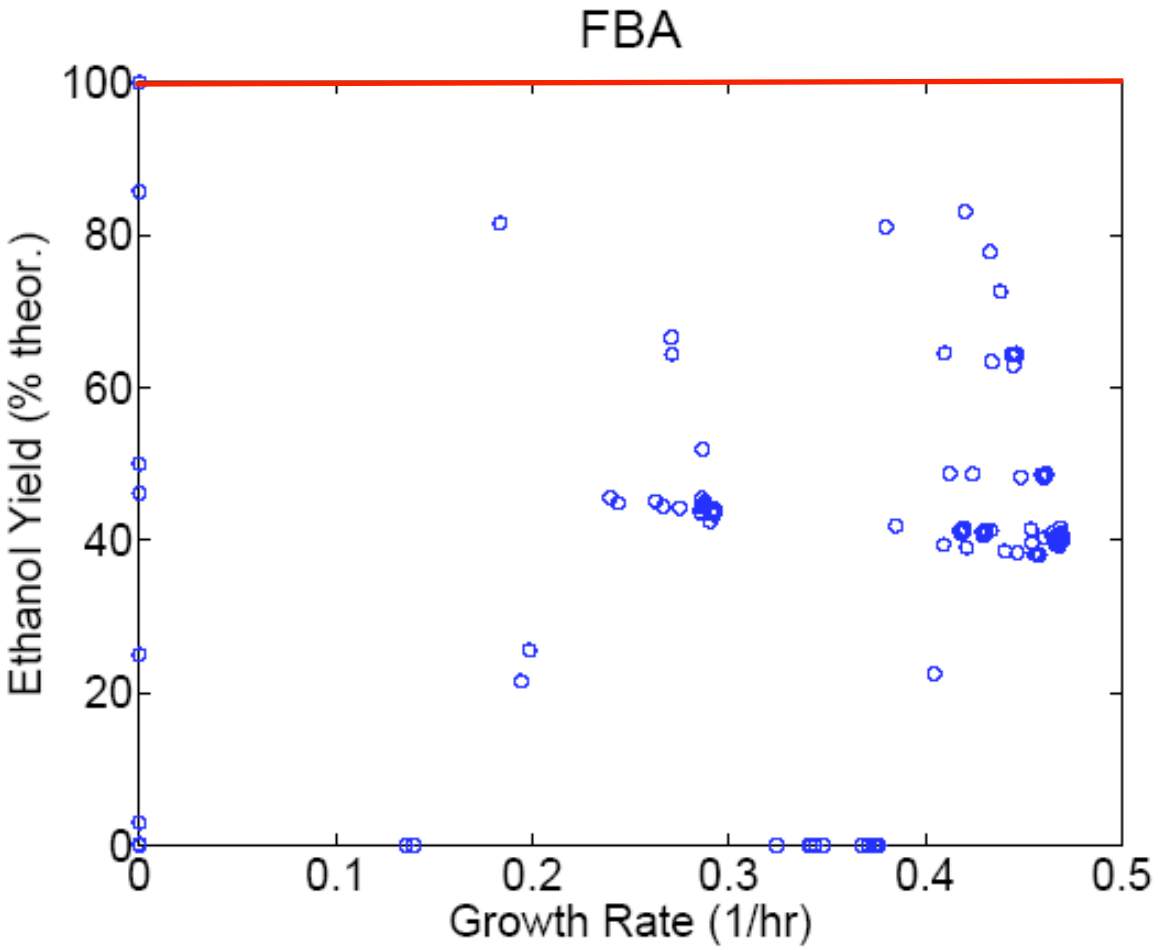
→408,000 Possible Double Knockouts

904 Metabolic Genes
-174 Essential Genes
-287 Unusable Genes
-81 Equivalent Genes
362 Genes of Interest

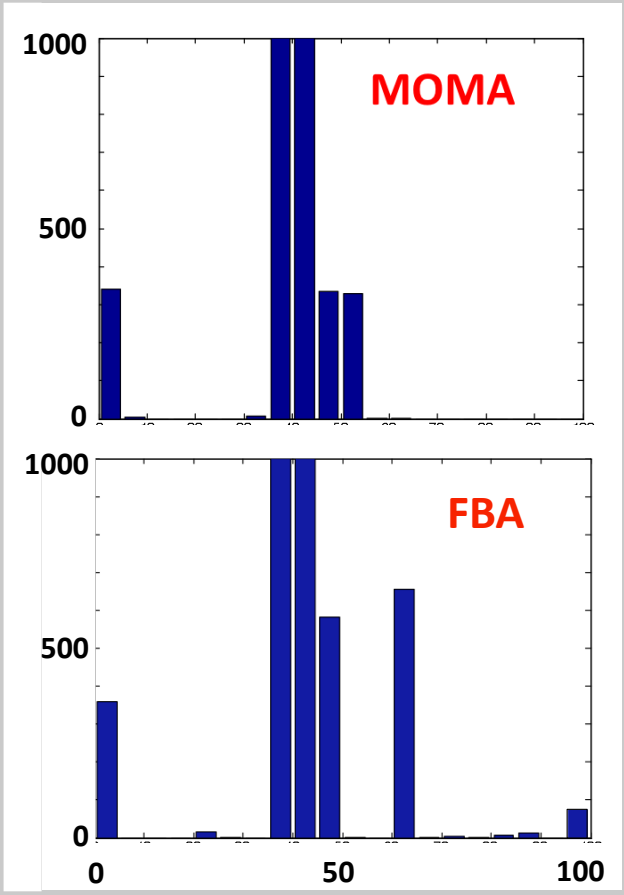
→ 57,000 Double Knockouts



Exhaustive Search:
All Possible Double Gene Deletions

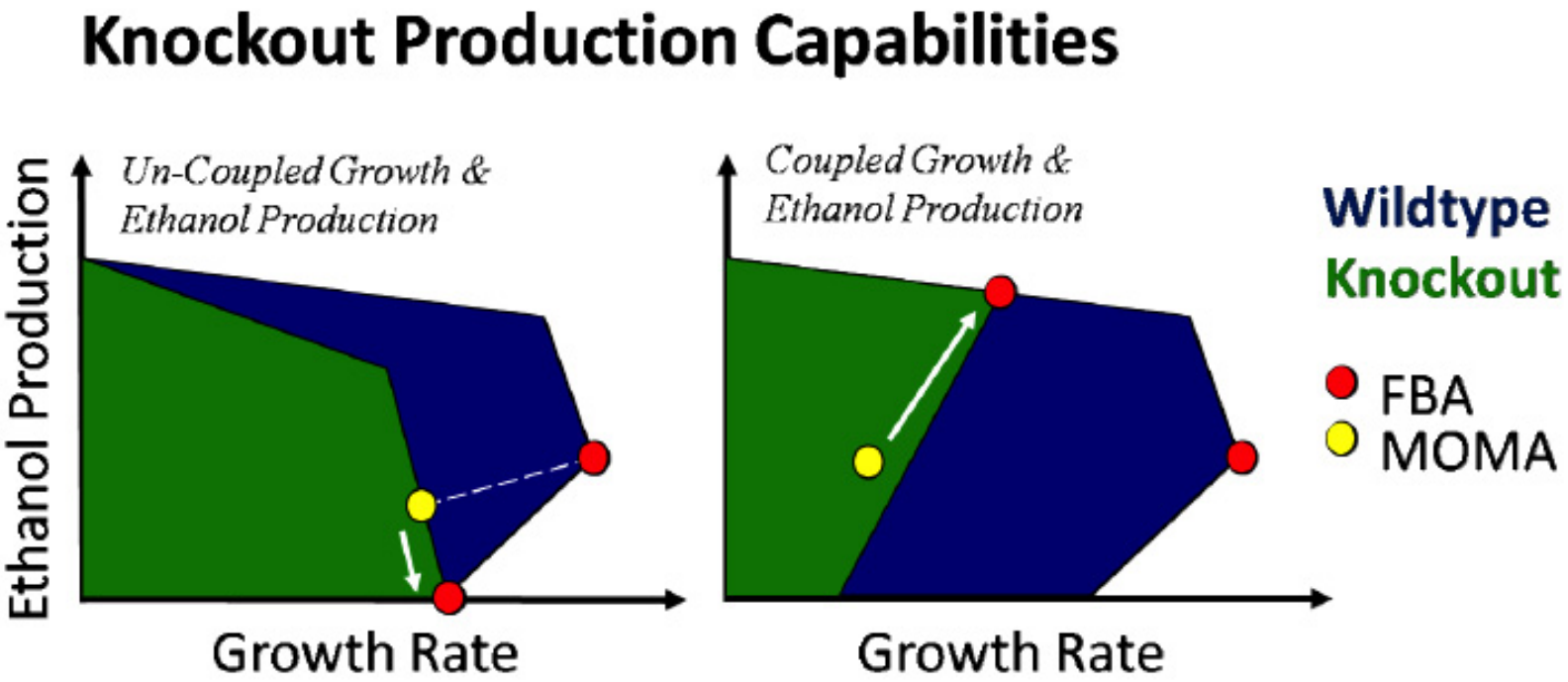


Mutant Histograms



Ethanol Yield

OptKnock: Identifies Mutants with Coupled Biomass & Metabolite Production

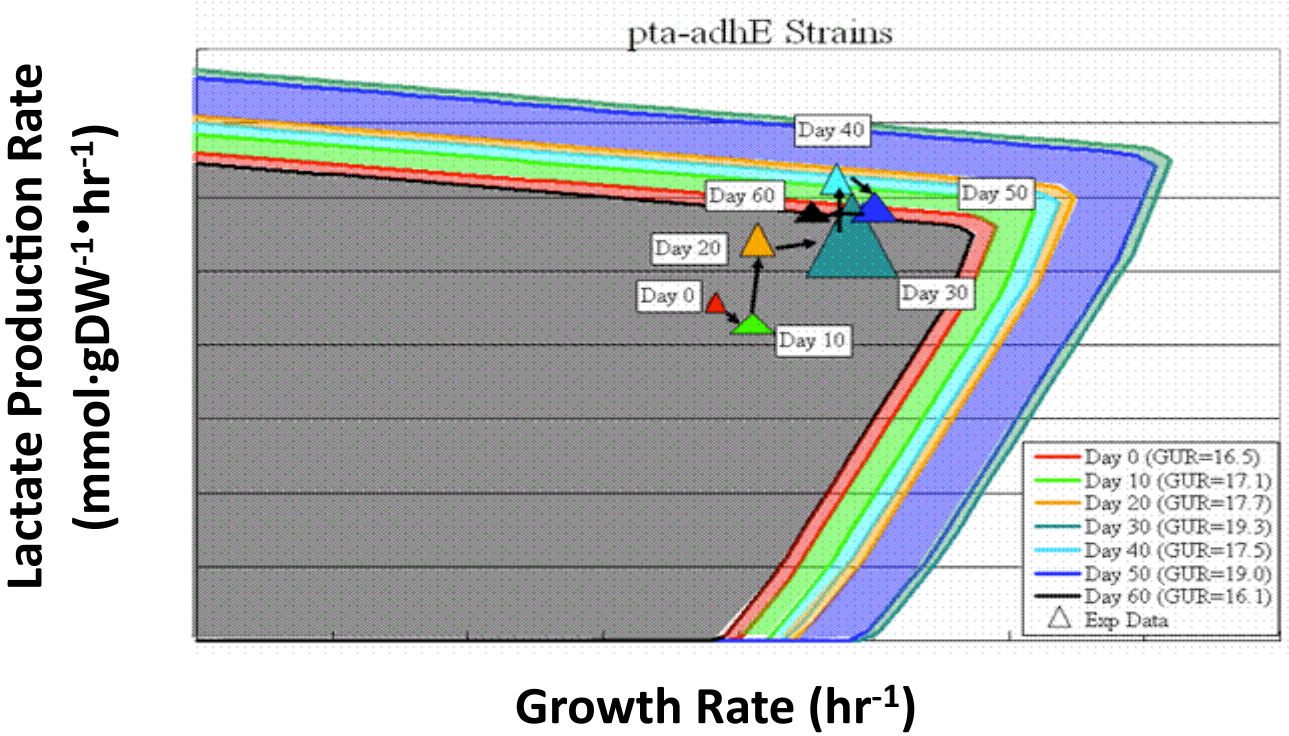


Finds reactions, that if removed, couple biomass production to metabolite production (ie. higher growth =higher production)

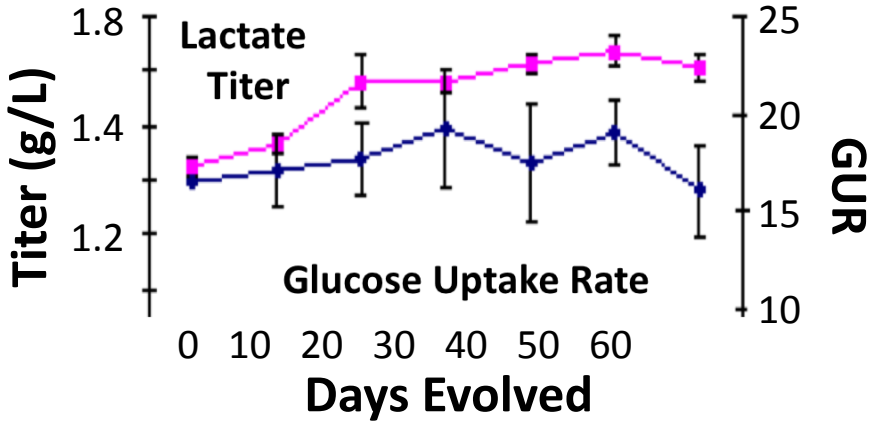
So even if mutants initially have low production, by adaptively evolving strains using growth rate as selection pressure, the mutants should improve their productivity

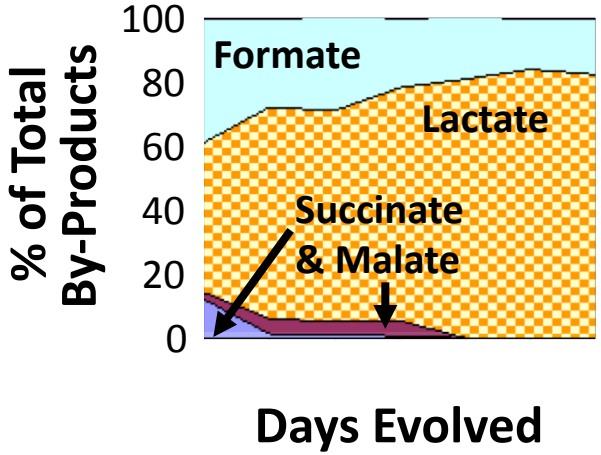
Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

Experimental Testing of a Lactate Strain



- Lactate secretion rate increased with increasing growth rate
- Lactate yield increased ~35% over 60 day evolution
- 2° by-product secretion decreased





S.S. Fong et al. B&B 2005

OptKnock Problem Statement

$$\begin{aligned}
&\underset{y_j}{\text{maximize}} && v_{chemical} && \textbf{(OptKnock)}\\
&\text{subject to} && \underset{v_j}{\text{maximize}}\ v_{biomass} && \textbf{(Primal)}\\
& && \text{subject to} && \sum_{j=1}^{M} S_{ij} v_j = 0,\\
& && && v_{pts} + v_{glk} = v_{glc_uptake}\\
& && && v_{atp} \geq v_{atp_main}\\
& && && v_{biomass} \geq v_{biomass}^{target}\\
& && && v_j^{\min} \cdot y_j \leq v_j \leq v_j^{\max} \cdot y_j, \quad \forall j \in \mathcal{M}\\
& && && y_j = \{0,1\}, \quad \forall j \in \mathcal{M}\\
& && && \sum_{j \in M} (1 - y_j) \leq K
\end{aligned}$$

Cells have to grow

If a reaction (j) is removed, set y_j=0 so that v_j has to equal 0.

To solve this problem, you transform it by using the dual constraints for the primal problem, in addition to the primal constraints

Burgard & Maranas. Biotechnol & Bioeng. 84(6):647-657 (2003)

Some Other Examples Where Bi-Level Optimization is Used

1. Met. Eng. Strain Design: Delete reactions/genes, add reactions/genes, change fluxes, alter regulation.
2. Calculate Obj. Functions: Given known fluxes find c so that max cv gives you known fluxes.
3. Synthetic Lethals: Find pairs of deletions where the resulting max. growth rate is 0.
4. Model Identification: Find which reactions need to be removed to match experimental observations.

Calculating the Flux Envelop

- This is a combination of flux variability analysis (for just the production flux) and robustness analysis (varying growth rate)!

```
upperLimits(j)=vmax;
*CARBON SOURCE: select upper and lower limits for exchange flux
LowerLimits('EX_glc_e')=-18.5;
UpperLimits('EX_glc_e')=0;
*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')=0;
LowerLimits('EX_pi_e')=-Vmax;

LowerLimits('ATPM')=7.6;
S(i,'Biomass')=1.3*S(i,'Biomass');

*Define the number of steps that you want to take eg. /step1*step25/ will have 25 steps
Sets
steps /step1*step15/
xaxis(j) /Biomass/
yaxis(j) /EX_lac_D_e/
maxmin /maxprod,minprod/;

Parameter
c(j)   used to define the objective function for optimization
n_steps  number of steps that will be taken and is defined by the elements in steps
range_max maximum flux value through the flux to be varied
range_min minimum flux value through the flux to be varied
flux_value(steps)  stores the values for the varied flux
store_obj(steps,maxmin) stores the value of the objective function for each iteration;
```

How many steps along the x-axis

Flux along the x-axis

Flux along the y-axis

Calculating the Flux Envelop

- This is a combination of flux variability analysis (for just the production flux) and robustness analysis (varying growth rate)!

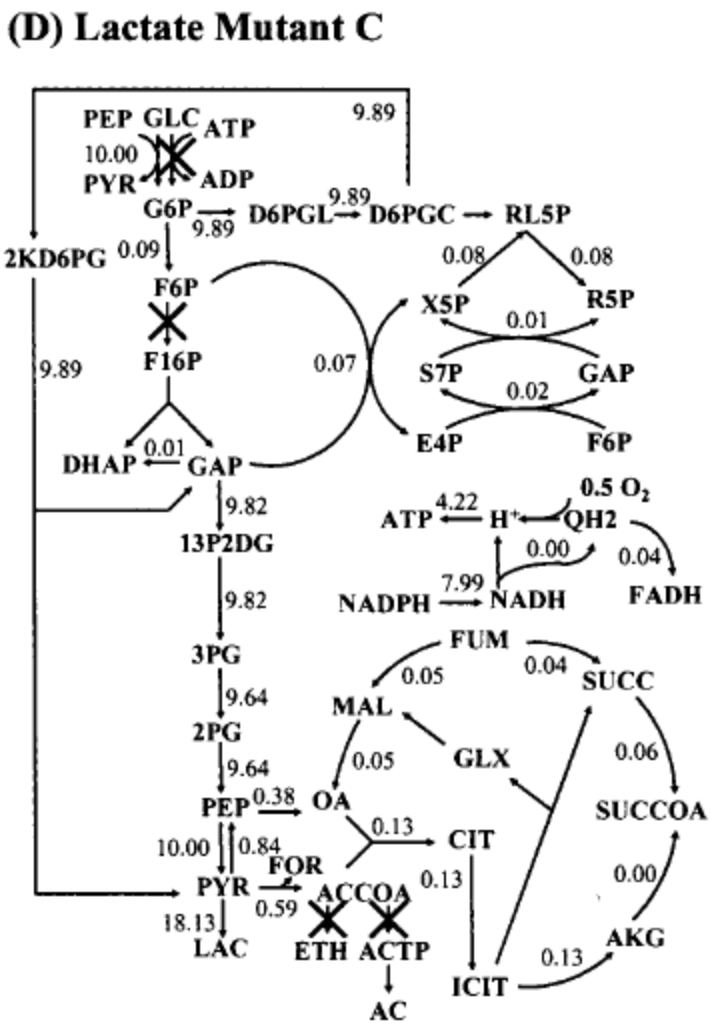
```
*calculate the allowable range for the chosen flux
c(xaxis)=1;
solve FBA using lp maximizing Obj;
range_max=Obj.l;
solve FBA using lp minimizing Obj;
range_min=Obj.l;

*reset the objective function to maximize objective of interest
c(j)=0;
c(yaxis)=1;
loop(steps,
        flux_value(steps)=range_min+(ord(steps)-1)*(range_max-range_min)/(n_steps-1);
        v.fx(xaxis)=flux_value(steps);
        solve FBA using lp maximizing Obj;
        store_obj(steps,'maxprod')=Obj.l;
        solve FBA using lp minimizing Obj;
        store_obj(steps,'minprod')=Obj.l;
);
```

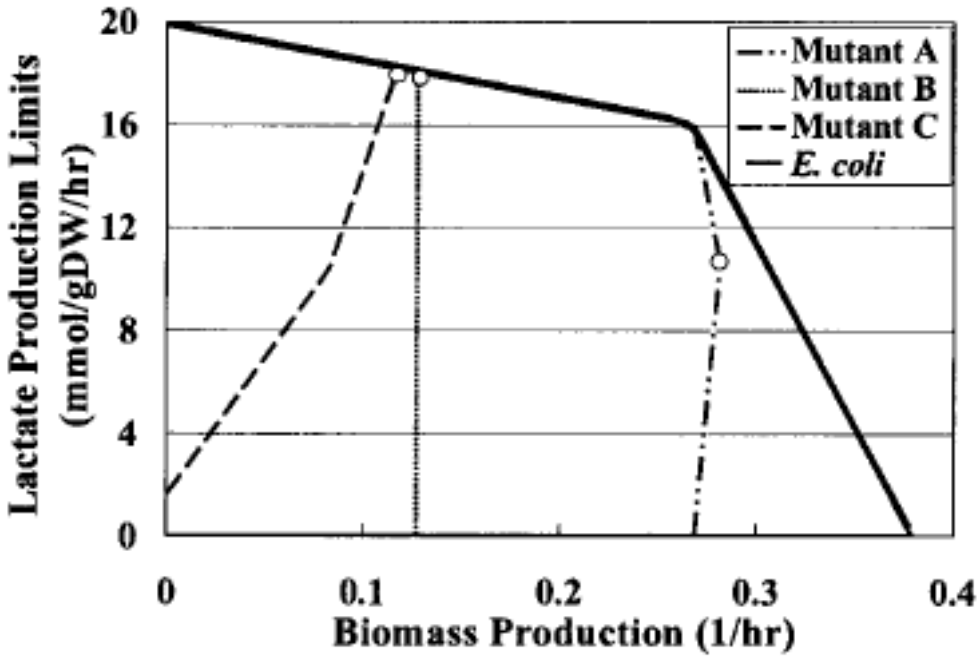
Find the range for the x-axis (min and max flux value)

Change the objective from x-axis flux to y-axis flux.

Fix flux on x-axis and calculate the min and max flux on the y-axis

Lactate Production Strains





Glucose Uptake = 10

Oxygen Uptake = 0

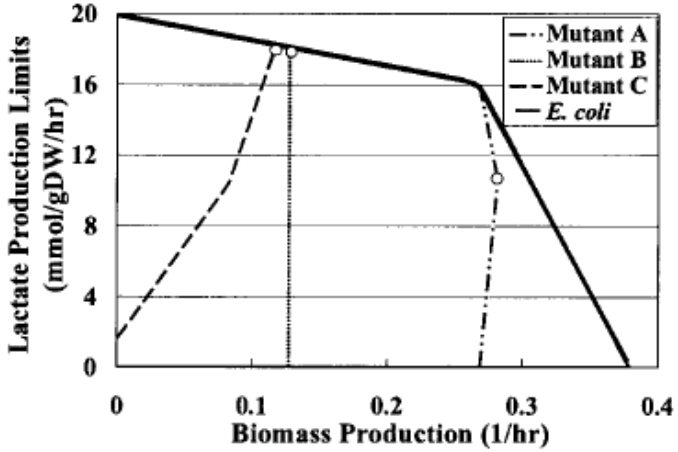
Calculations

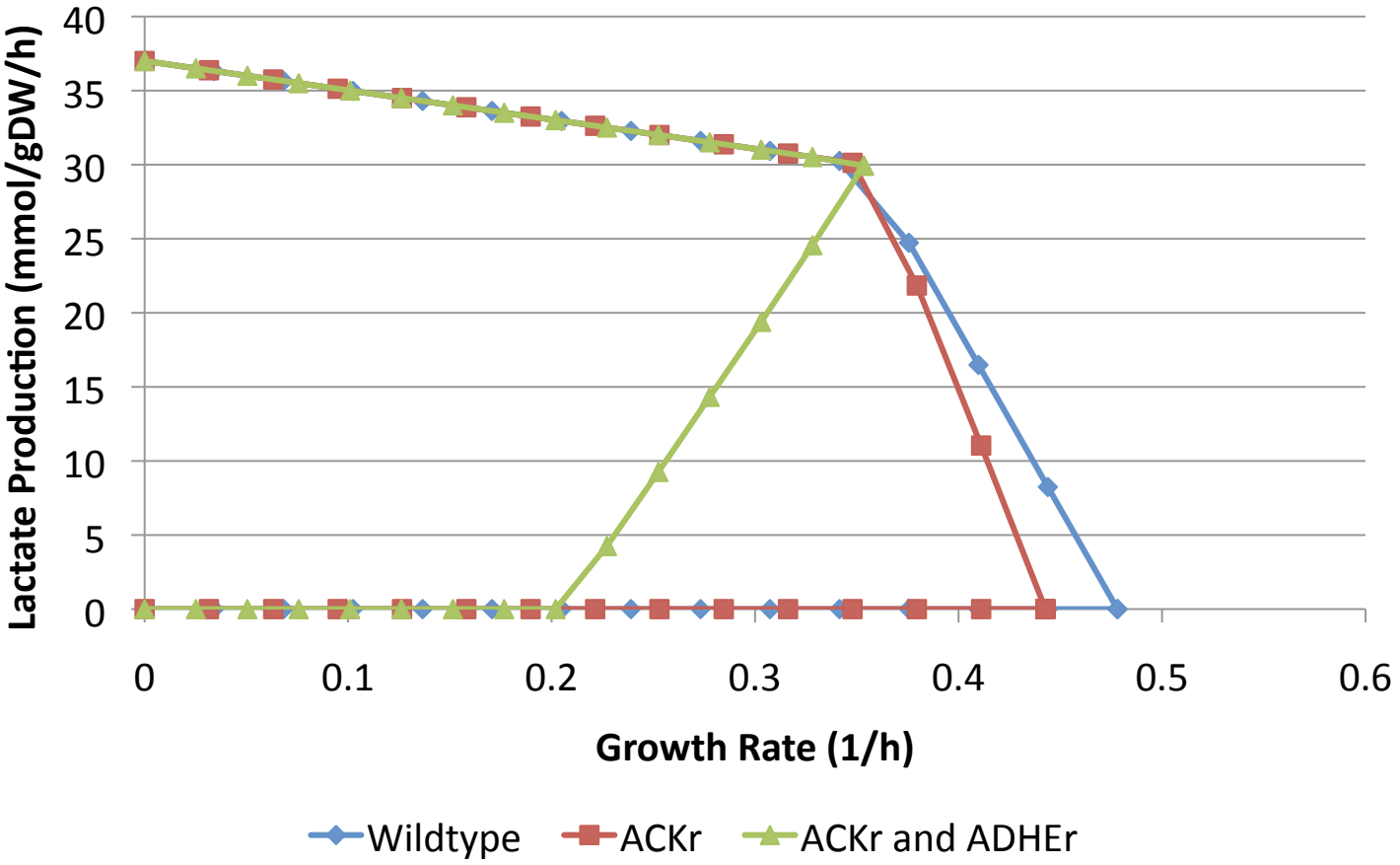
- Calculate and graph the flux envelops for lactate production under glucose ANAEROBIC conditions for:
  - Wildtype solution
  - Acetate Kinase mutant (ACKr reaction)
  - Acetate Kinase & Aldehdehyde Dehydrogenase double mutant (ACKr and ADHEr reactions)

  **We will use physiological measurements for glucose anaerobic uptake and ATP maintenance

  **Remember to delete a reaction we can either change upper and lower limits or use v.fx('NAME')=0;

- Why might your graphs look different from those in the publication?





- Why might your graphs look different from those in the publication?
  - Different metabolic networks are used (their network overpredicts anaerobic growth rates)!
  - Different glucose uptake rates are used (they used a value of 10 rather than 18.5)!

OptKnock Code

Set minimum growth rate

```
* Target growth rate
LowerLimits('Biomass')=0.01;

UpperLimits('EX_glc_e')=LowerLimits('EX_glc_e');
```

No. of Solutions

```
Sets
store how many optknock solutions you want to find /solution1*solution3/
exclude(j) subset of reactions that you do not want OptKnock to consider (usually transporters and
/Biomass,EX_ac_e,EX_akg_e,EX_co2_e,EX_etoh_e,EX_for_e,EX_fum_e,EX_glc_e,EX_h2o_e,EX_h_e,EX_lac_D_e,
ACt2r,AKGt2r,CO2t,D_LACt2,ETOHt2r,FORt,FUMt2_2,GLCpts,H2Ot,O2t,PIt,PYRt2r,SUCCt2_2,SUCCt2b/;
```

Reactions you don't want to consider deleting

```
Parameter
outer_obj(j) used to define the objective function for Optknock /EX_ac_e 1/
inner_obj(j) used to define the objective function for inner problem /Biomass 1/
Umax used to limit dual variables /1000/
maxKnockouts maximum number of knockouts /1/
storeA(j,store) stores the a values from previous iterations
storeRxns(store) stores the number of reactions that were removed from previous iterations
storeV(j,store) stores the fluxes at the maximum growth rate from previous iterations
alpha /1e-4/;
```

Objective Functions

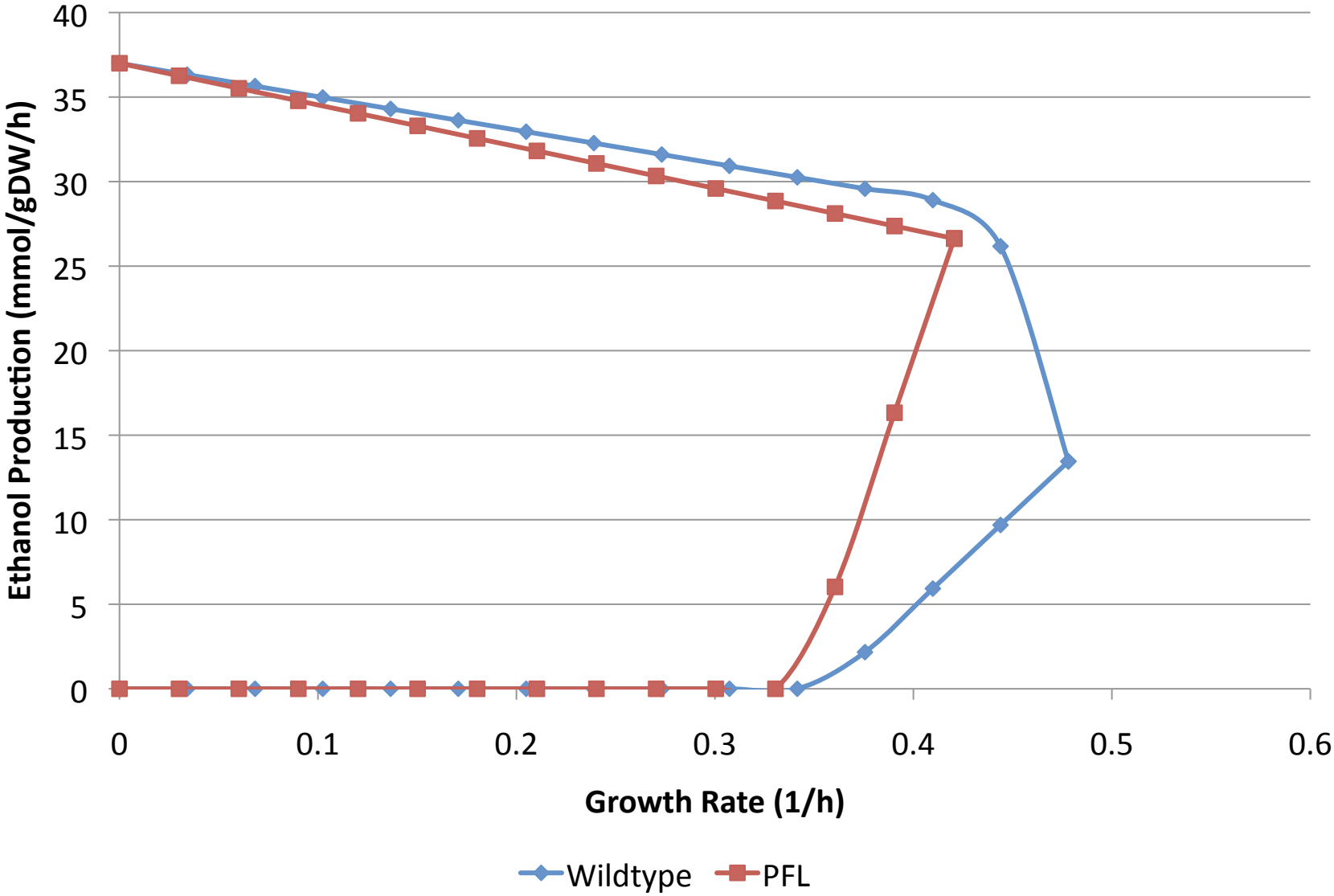
Max No. Knockouts

Questions:

1. Calculate what the maximum theoretical yield is for ethanol from glucose using FBA.
2. Use FBA, to determine what the ethanol production for the wildtype strain would be under anaerobic conditions (glucose uptake =18.5).
   - Include LowerLimits('ATPM')=7.6;
   - Include S(i,'Biomass')=1.3*S(i,'Biomass');
3. Identify six single gene deletion strains that would couple biomass to ethanol production.
4. Which strain would result in the highest ethanol production?
5. Calculate the production envelop for ethanol for both the wildtype and best OptKnock strain

Questions:

1. Calculate what the maximum theoretical yield is for ethanol from glucose using FBA.
   - ANS: 2 mol ethanol per mol glucose
2. Use FBA, to determine what the ethanol production rate for the wildtype strain would be under anaerobic conditions (glucose uptake =18.5).
   - ANS: at maximum growth rate (μ=0.48 hr^{-1}), ethanol production is 13.4 mmol/gDW/hr
3. Identify six single gene deletion strains that would couple biomass to ethanol production.
   - ANS: PFL,PTAr, ACKr, ATPS4r, THD2,TKT2
   - Note: ΔPYK is only slightly higher than wildtype in terms of ethanol production
4. Which strain would result in the highest ethanol production?
   - ANS: PFL, then ACKr and PTA

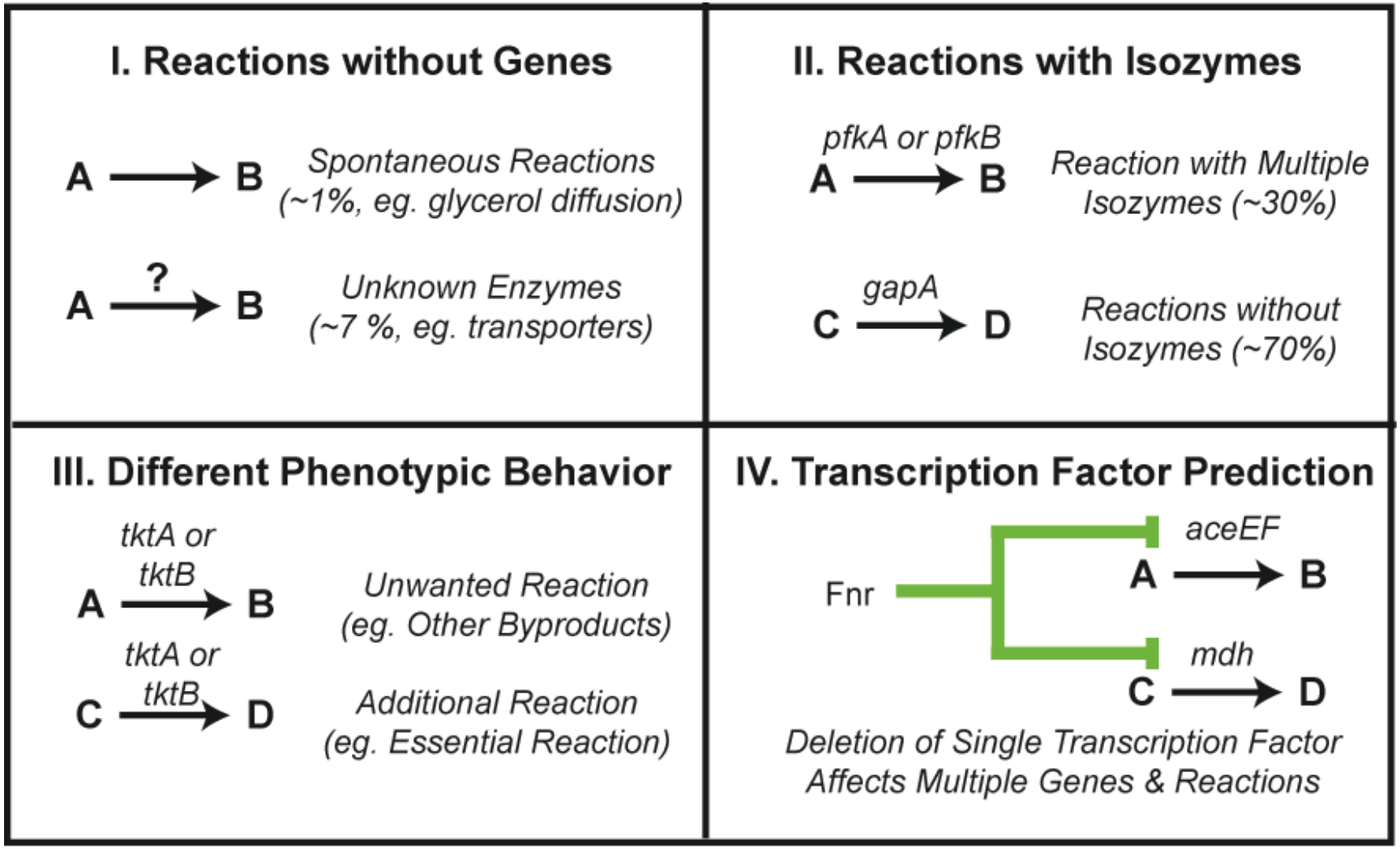

Ethanol Production (mmol/gDW/h)
40
35
30
25
20
15
10
5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
Growth Rate (1/h)
Wildtype
PFL

OptKnock Variations

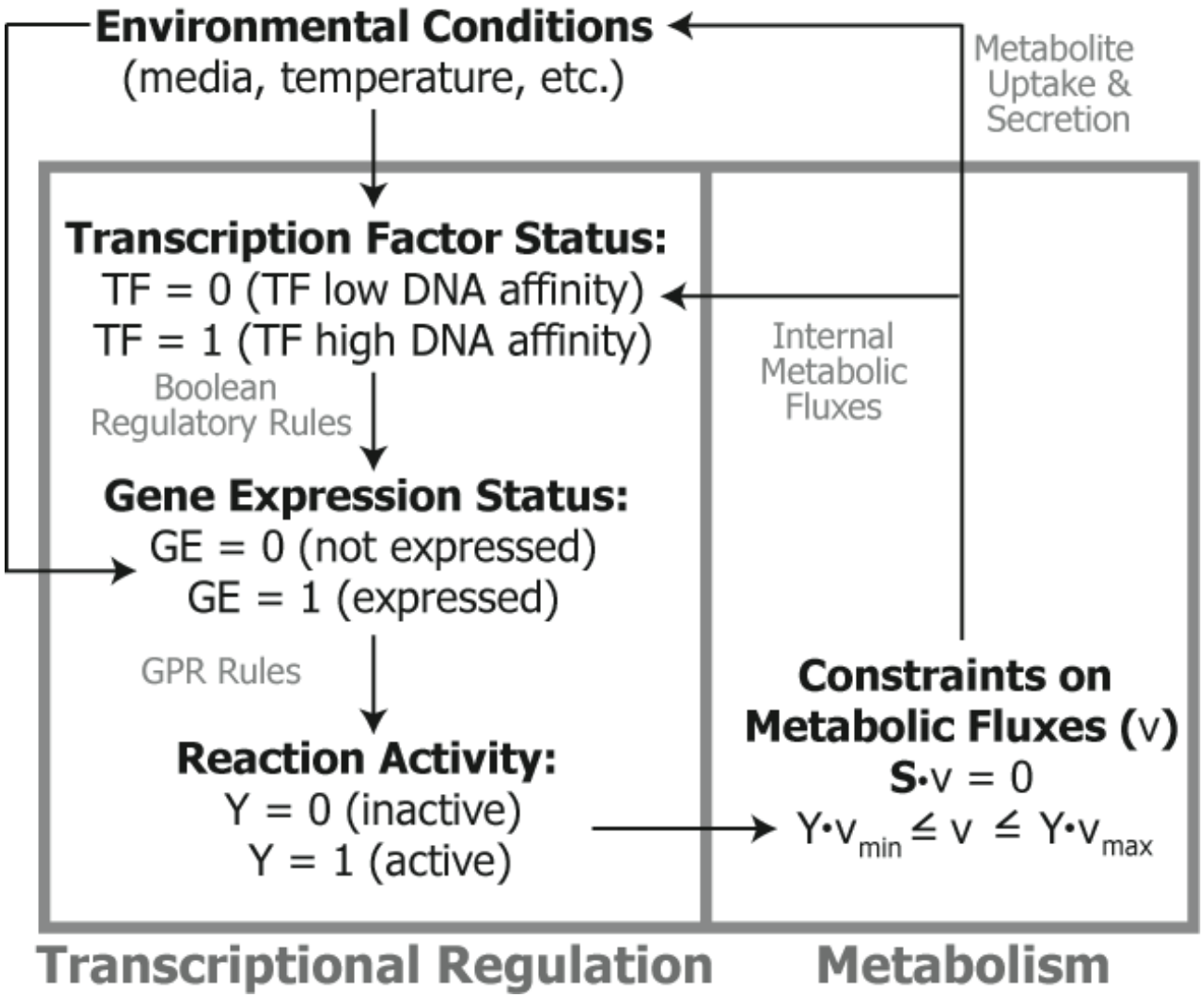
- **OptStrain[1]:** Two step process, where (1) non-native pathways are identified that lead to product formation, and then (2) OptKnock is carried out to identified coupled phenotypes.
- **OptReg[2]:** Rather than consider reaction deletions, this also considers significant changes in fluxes.
- **OptGene[3]:** Uses genetic algorithms instead of optimization procedures to find the solutions.
- **OptORF[4]:** evaluates metabolic and regulatory gene deletions by gene and not reaction

1. Pharkya, Burgard, & Maranas, Genome Research, 14:2367-76 (2004)
2. Pharkya & Maranas, Metabolic Eng, 8:1-13 (2006)
3. Patil, Rocha, Forster & Nielsen, BMC Bioinformatics, 6:308 (2005)
4. Kim & Reed, BMC Systems Biology, 4:53 (2010)

Benefit of Considering Genes and Regulation



Integrated Models of Metabolism and Regulation



OptORF: Strain design with gene deletions and transcriptional regulatory effects

maximize	*biochemical production*	
subject to	**maximize**	*cellular growth*
	subject to	steady-state mass balance
		enzyme capacity
		thermodynamics
		reaction deletions
	GPR associations	
	transcriptional regulations	
	gene deletions and overexpressions	
	limited number of gene deletions	
	limited number of gene overexpressions	

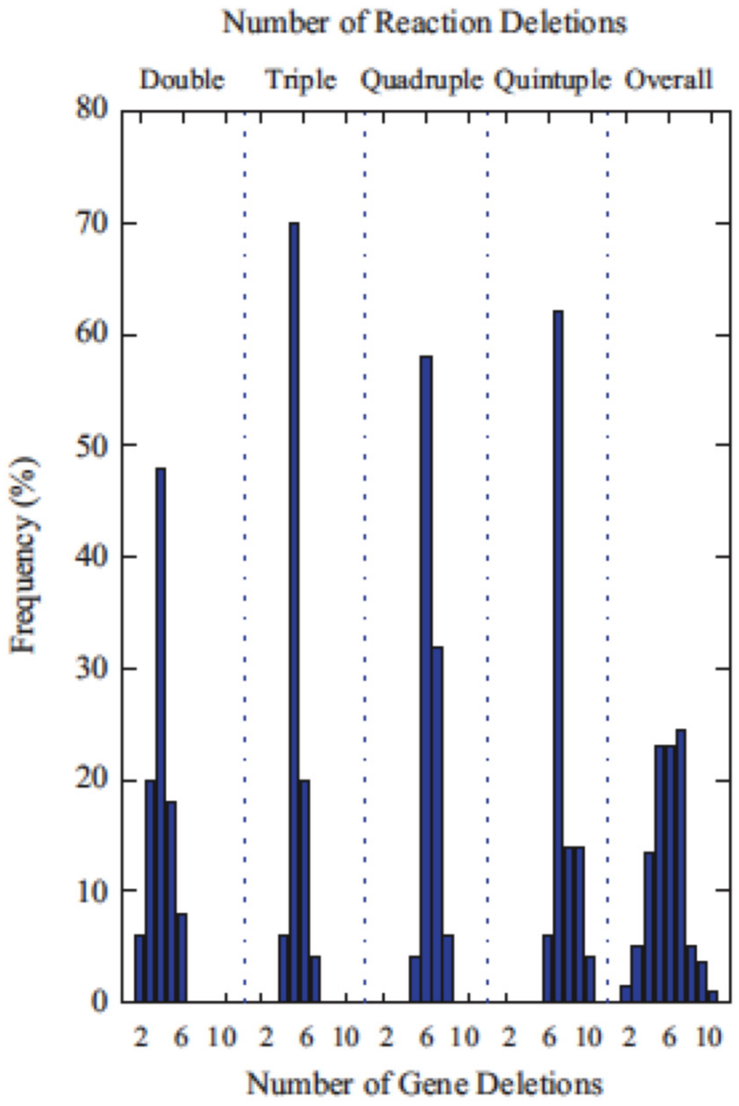
Kim and Reed. BMC Systems Biol 4:53 (2010)

Deleting by Gene versus Reaction

1. 200 Total OptKnock Strategies
   - 50 Double Reaction Deletions
   - 50 Triple Reaction Deletions
   - 50 Quadruple Reaction Deletions
   - 50 Quintuple Reaction Deletions

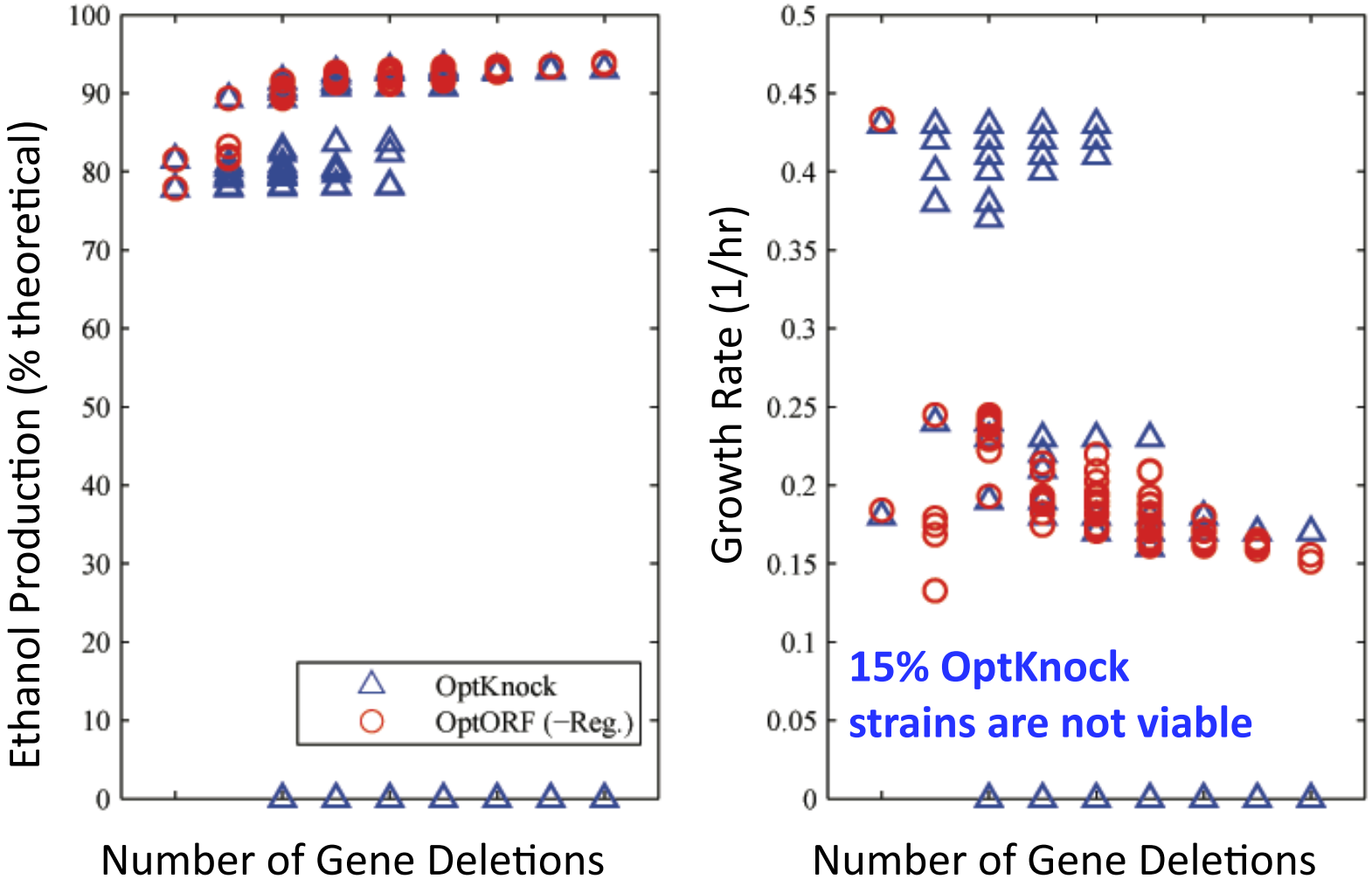
2. Mapped reaction deletions to gene deletions
   - OptKnock Strategies had between 2 and 10 genes

3. Found OptORF strategies with the same number of gene deletions



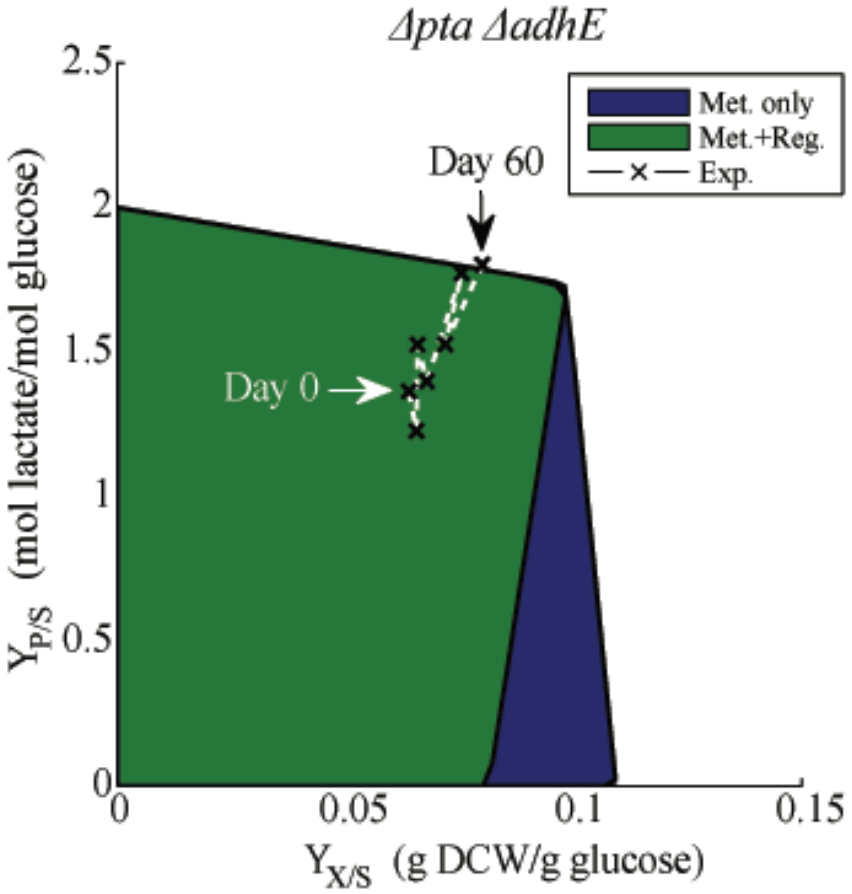
Kim and Reed. BMC Systems Biol 4:53 (2010)

Deleting by Gene versus Reaction

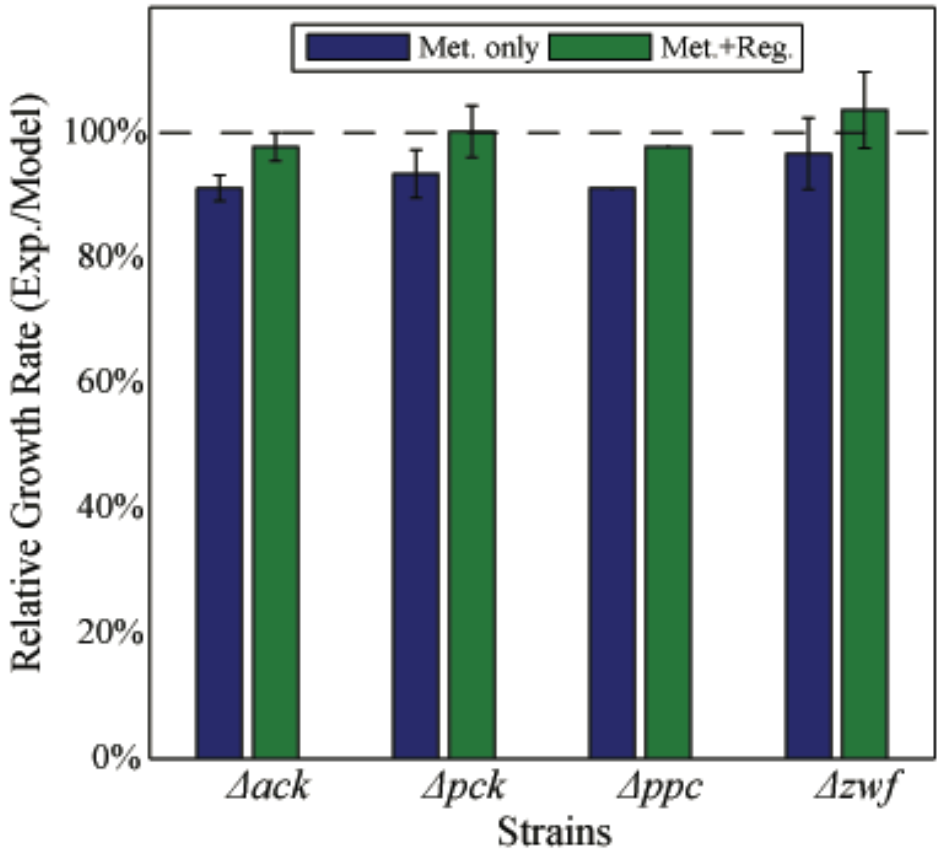


Kim and Reed. BMC Systems Biol 4:53 (2010)

Adaptive Evolutionary Outcomes are Consistent with Regulatory Predictions



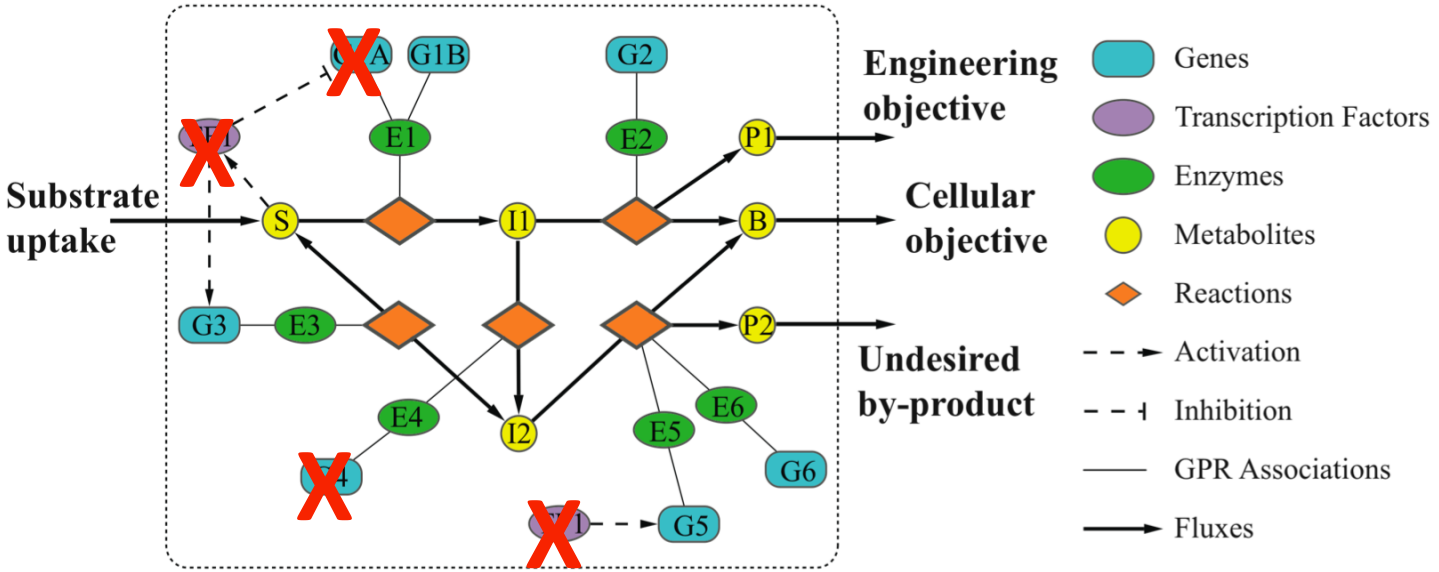
Data from S.S. Fong et al. Biotech & Bioeng (2005)



Data from S.S. Fong et al. Nature Genetics. (2004)

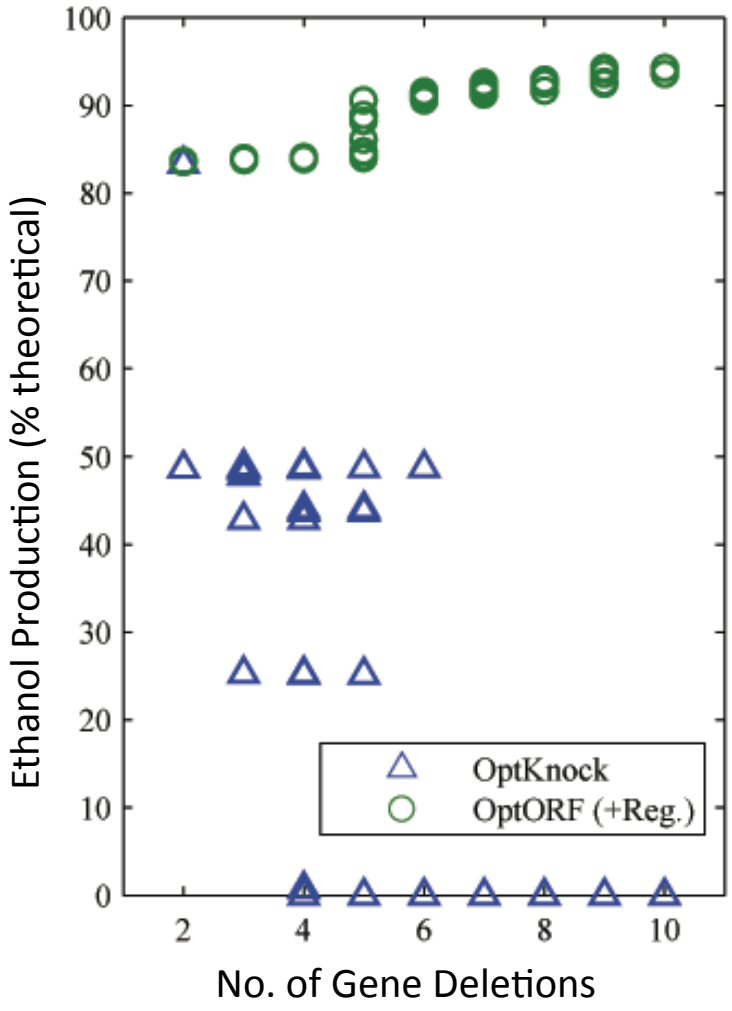
Kim and Reed. BMC Systems Biol 4:53 (2010)

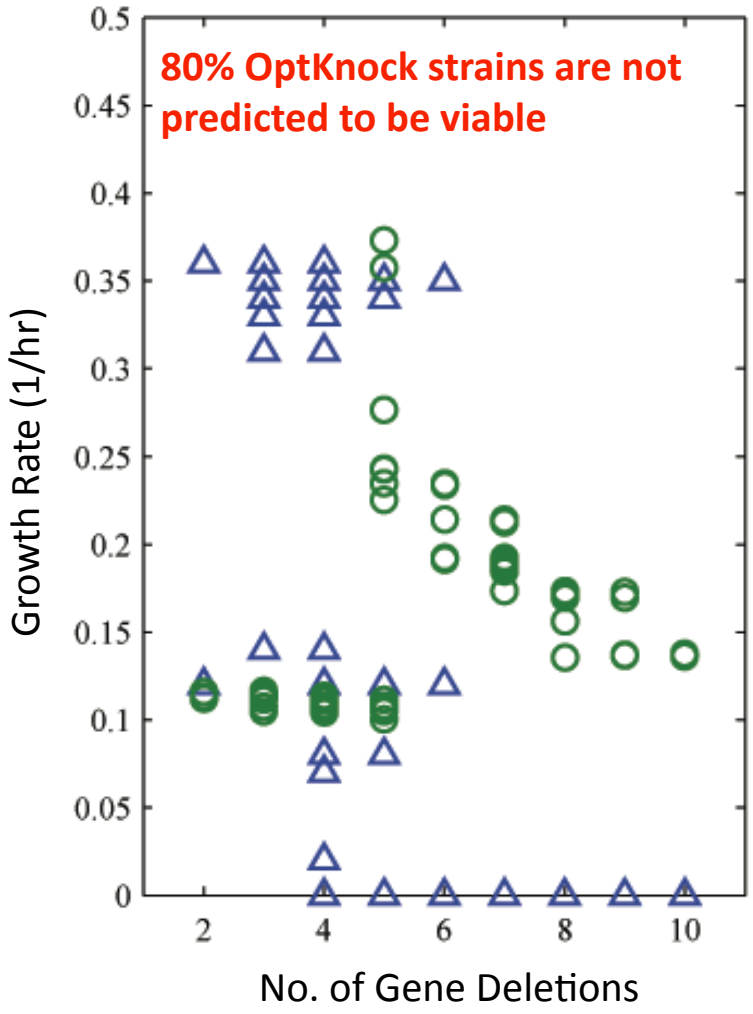
Example Network: Designing Around Transcriptional Regulation



Kim and Reed. BMC Systems Biol 4:53 (2010)

Transcriptional Regulation Restricts Growth and Ethanol Production





Kim and Reed. BMC Systems Biol 4:53 (2010)

Strains for Ethanol and Isobutanol (via BCAA pathways) Production

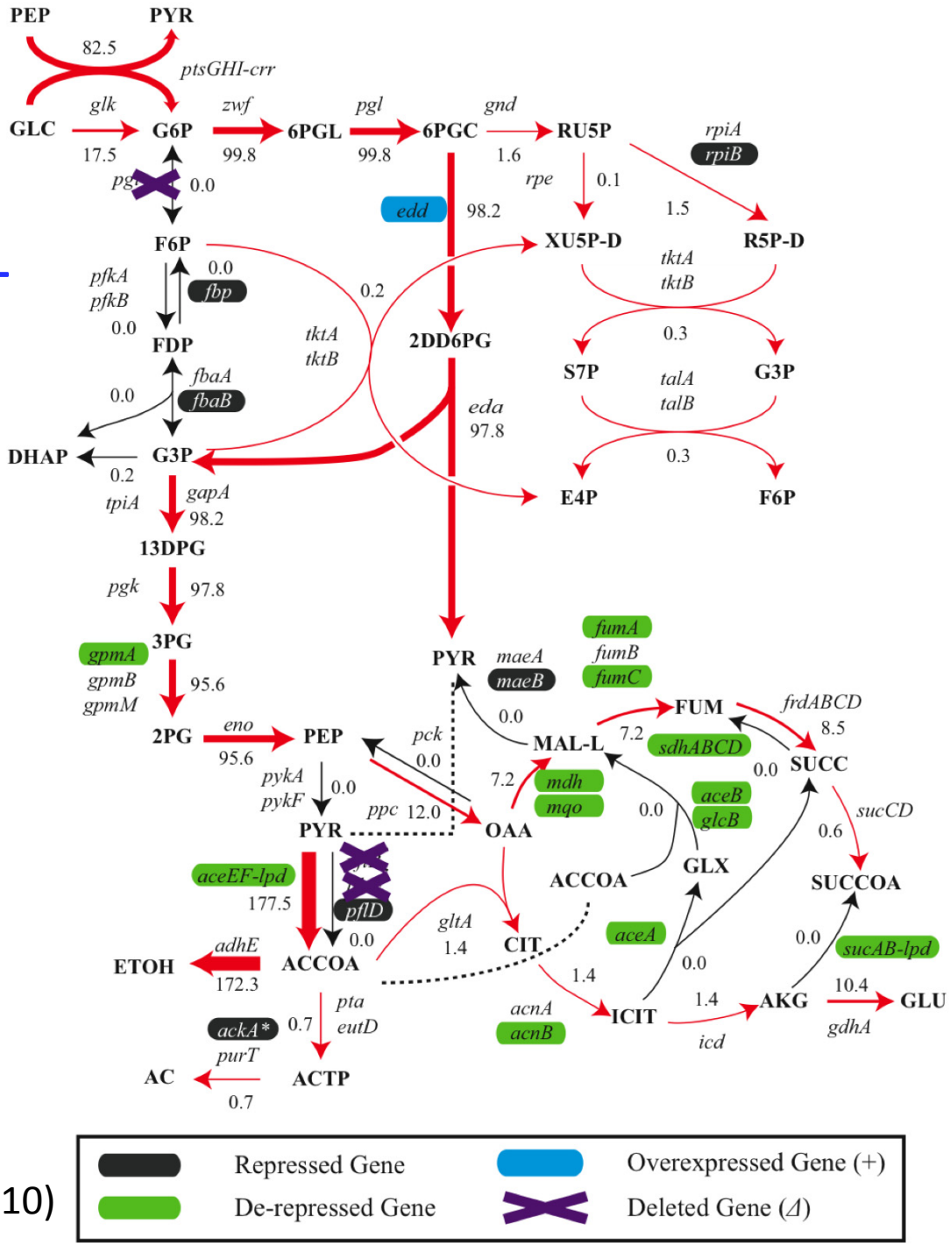
ETHANOL: Gene Deletions	Gene Over-Expression	Growth Rate	Ethanol Production (% max yield)
Δfnr ΔpflB ΔtdcE Δpgi	*edd*	0.225	86.2%
Δfnr ΔpflB ΔtdcE Δtpi	*edd*	0.235	90.5%
Δfnr ΔpflB ΔtdcE Δtpi ΔgdhA	*edd*	0.214	91.4%
ΔarcA Δpta ΔeutD Δtpi ΔptsH	*edd*	0.192	91.6%

ISOBUTANOL: Gene Deletions	Gene Over-Expression	Growth Rate	Isobutanol Production (% max yield)
ΔadhE ΔgdhA		0.223	89.5%
ΔgntR ΔadhE Δpgi		0.128	93.8%
ΔadhE Δtpi	*edd+fbp*	0.128	94.3%
ΔadhE ΔpntA Δnuo	*edd+fbp*	0.110	95.1%
ΔadhE ΔpntA ΔgdhA	*edd+fbp*	0.102	95.5%

Strategy with Gene Deletions & Gene Over-expression

<u>Mutations</u>
Δfnr ΔpflAB
ΔtdcE Δpgi
+edd

Predicted Ethanol Yield:
86%
Predicted Growth Rate:
0.225 hr^{-1}



Kim and Reed. BMC Systems Biol 4:53 (2010)