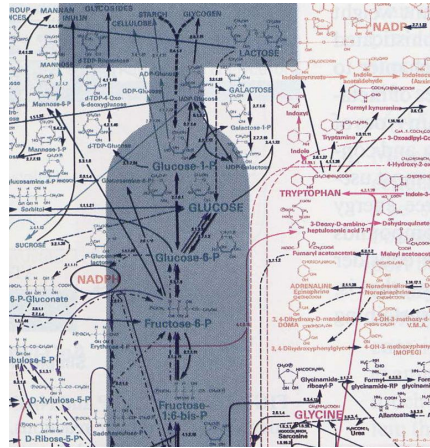
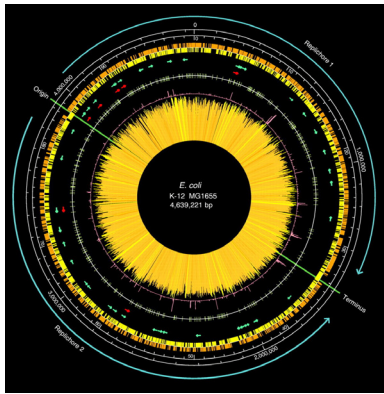


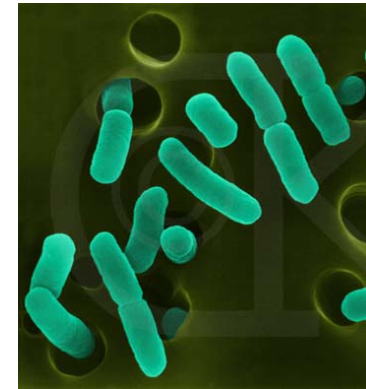
Using Optimization to Explore and Leverage Biochemical Networks



$$S = \begin{pmatrix} 0 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$S \cdot v = 0$$

$$v_{\min,i} \leq v_i \leq v_{\max,i}$$



Genome

Networks

Models

Phenotypes

Jennifer Reed
Assistant Professor
Dept of Chemical & Biological Engineering
University of Wisconsin- Madison



UW-Madison, Chemical & Biological Engineering

Modeling in Systems Biology

Gather information about important **components** and **component interactions** in biological networks

Component Information

Network Reconstruction

Glycolytic reactions



Organize and assemble component information at a **systems level** using a textual, graphical, or **mathematical representation**

Compare model **predictions** to **experimental data**, either retrospectively or prospectively

Experimental Data

Mathematical Model

$$S \begin{pmatrix} 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}$$

$$S \cdot v = 0$$

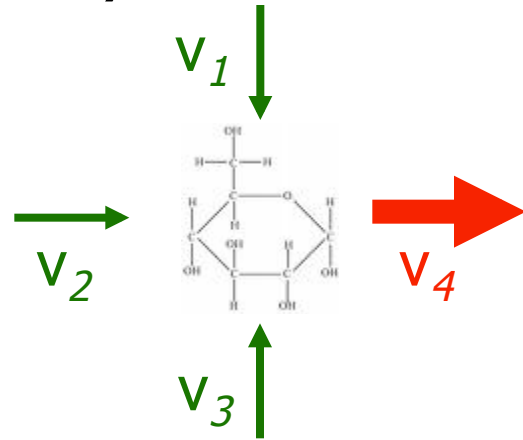
$$v_{min,i} \leq v_i \leq v_{max,i}$$

Convert the reconstruction into a model by introducing **variables** and **equations** based on **chemical** and **physical principles**



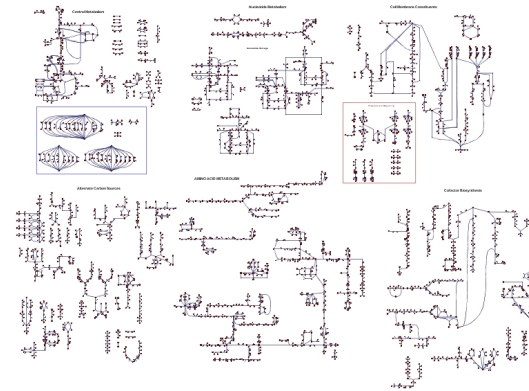
Constraints on Metabolic Networks

1. Steady-State Mass Balance Constraints



For each metabolite:

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



For all metabolites:

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

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

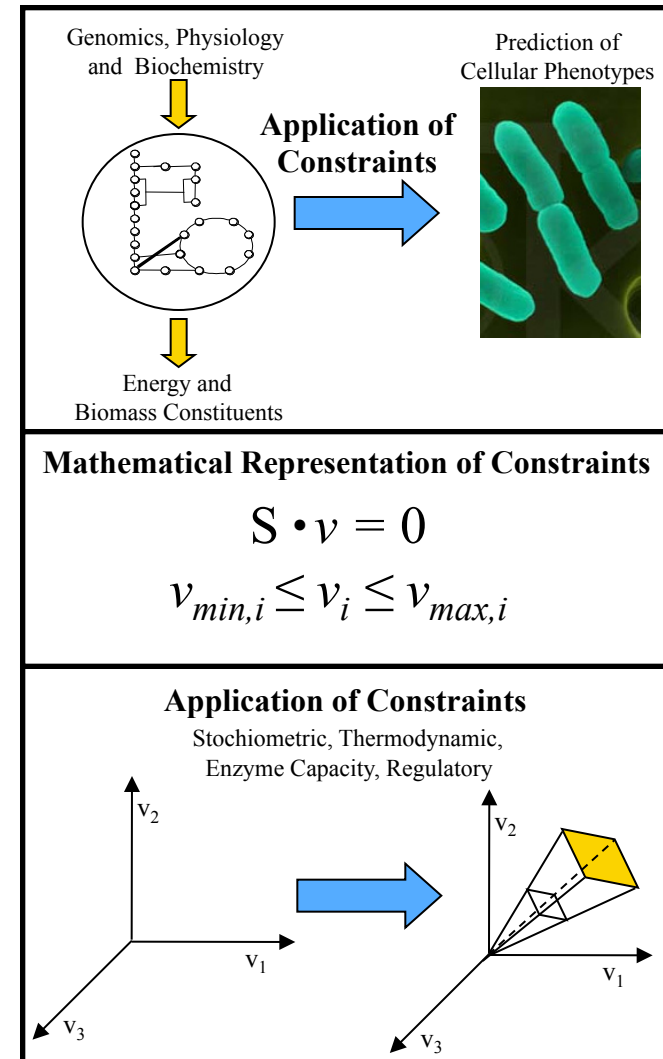


How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

–Sherlock Holmes, A Study in Scarlet



UW-Madison, Chemical & Biological Engineering



Exploring Biochemical Networks by Integrating Models & Data

5

1. Evaluating Network Structure:

- Metabolism and Regulation
- *Escherichia coli*
- Genomic and phenotypic data

2. Evaluating Network Usage:

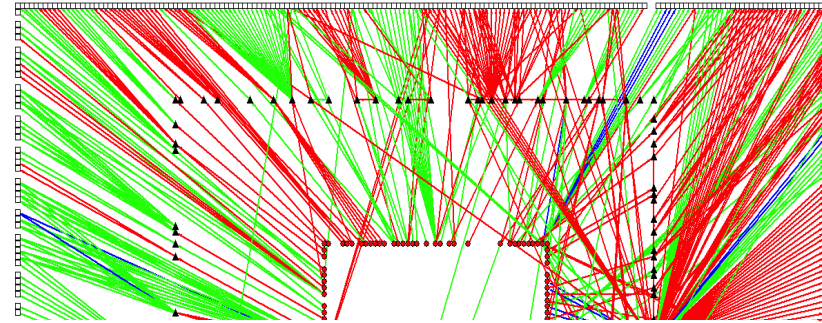
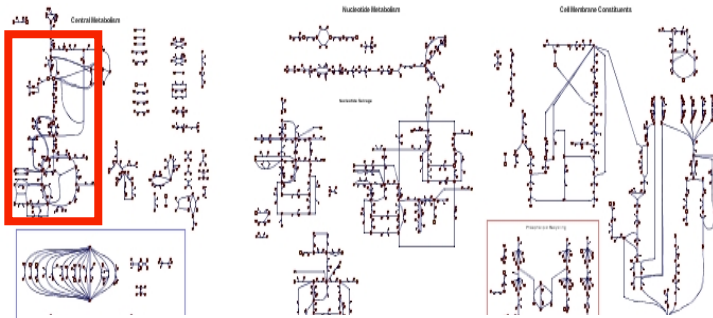
- Metabolism
- *Shewanella oneidensis* MR-1
- Genomic, transcriptomic, and phenotypic data



Escherichia coli Networks

METABOLISM

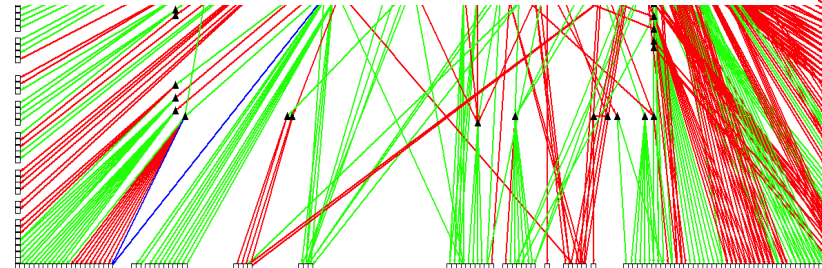
REGULATION



*How do we know if these networks are correct?
Are we missing nodes or connections?*



*>600 Metabolites
>900 Reactions
>900 Genes*



*~100 Transcription Factors
~500 Gene Targets
~50 Stimuli*

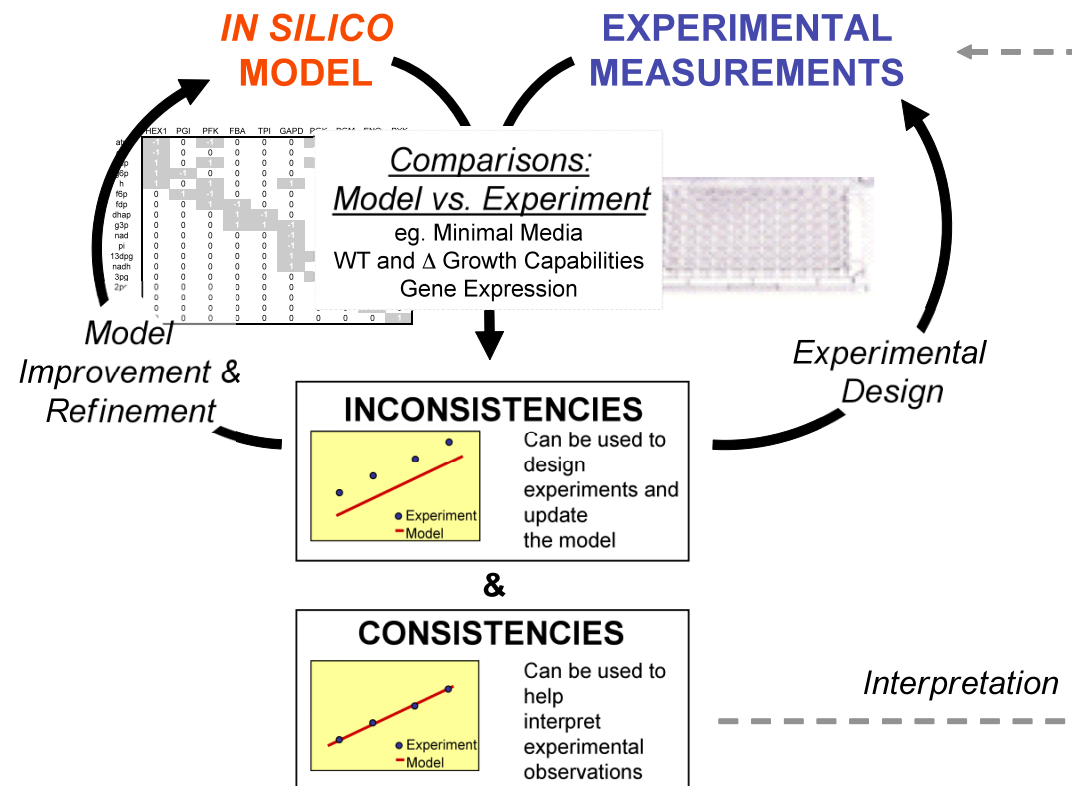
Reed, Vo, Schilling, and Palsson. *Genome Biology*. 4:R54.1-R54.12 (2003)

Covert, Knight, Reed, Herrgard, and Palsson. *Nature*. 429: 92-96 (2004).



UW-Madison, Chemical & Biological Engineering

Model Driven Discovery Via High Throughput Testing



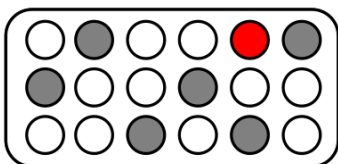
Growth Data Comparisons: Two Failure Modes

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

Phenotypes vs. Model Predictions



- Exp. Growth (+)
Pred. Growth (-)
- Exp. Growth (-)
Pred. Growth (-)
- Exp. Growth (+)
Pred. Growth (+)

Computational Algorithm

minimize $\sum a_i + \sum b_j$
such that:

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

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

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

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

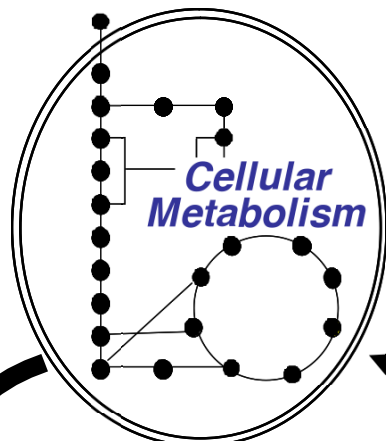
$$b_k \cdot z_{\min,k} \leq z_k \leq b_k \cdot z_{\max,k} \quad (5)$$

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

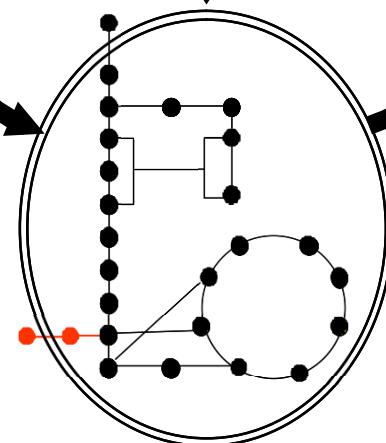
For multiple solutions, at iteration q include the additional $q-1$ constraints:

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

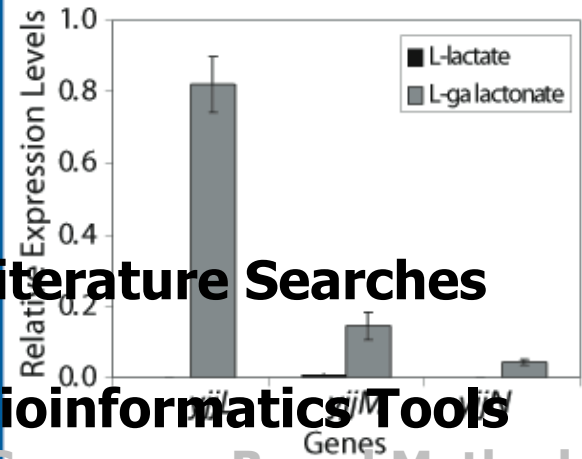
where $n=1 \dots q-1$



METABOLIC RECONSTRUCTION



Gene Expression

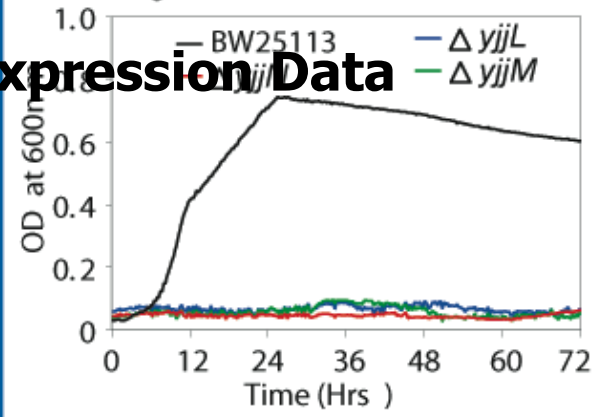


Literature Searches

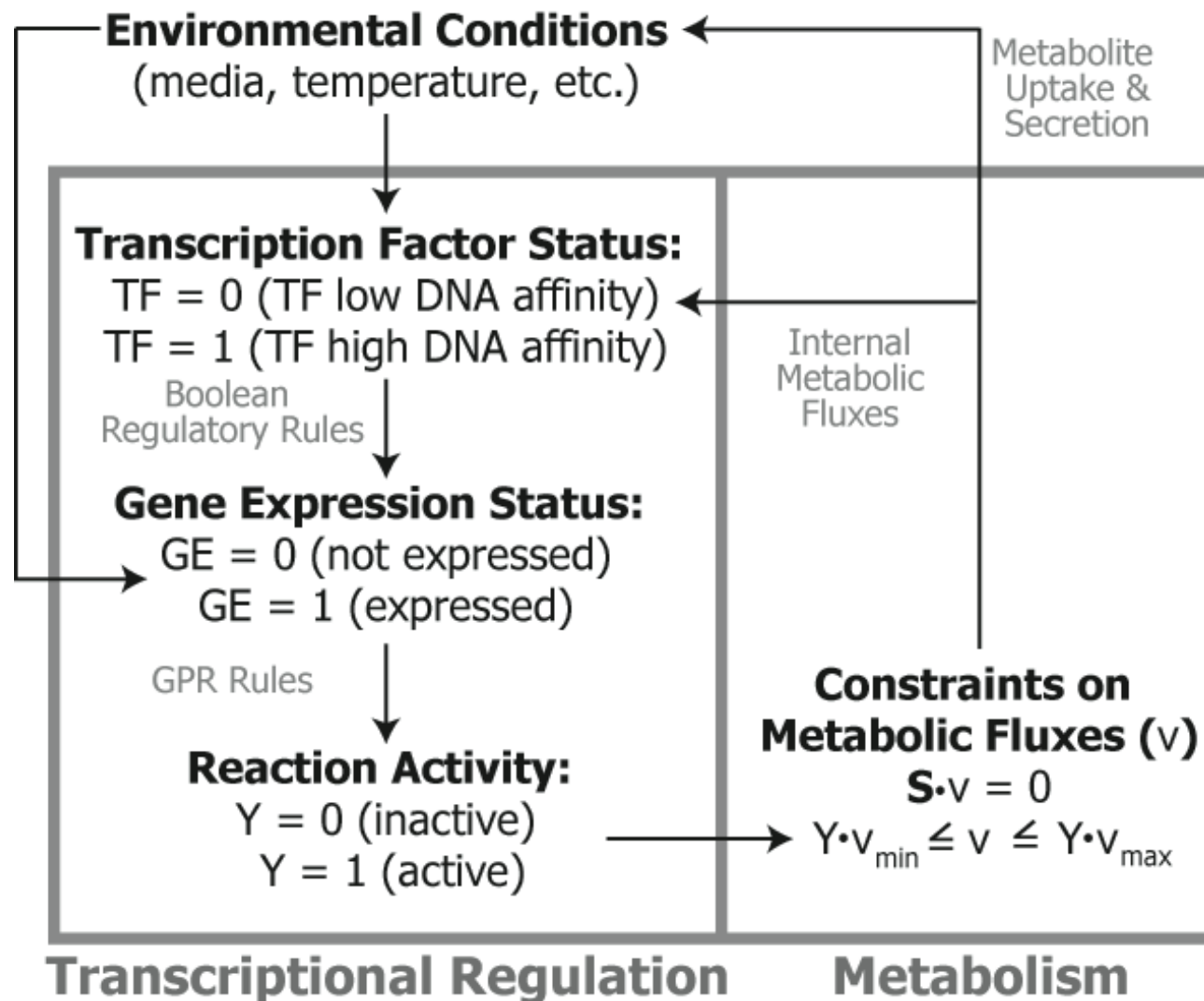
Bioinformatics Tools

- Sequence Based Methods
- Context Based Methods (eg. phylogenetic profiles)

Expression Data

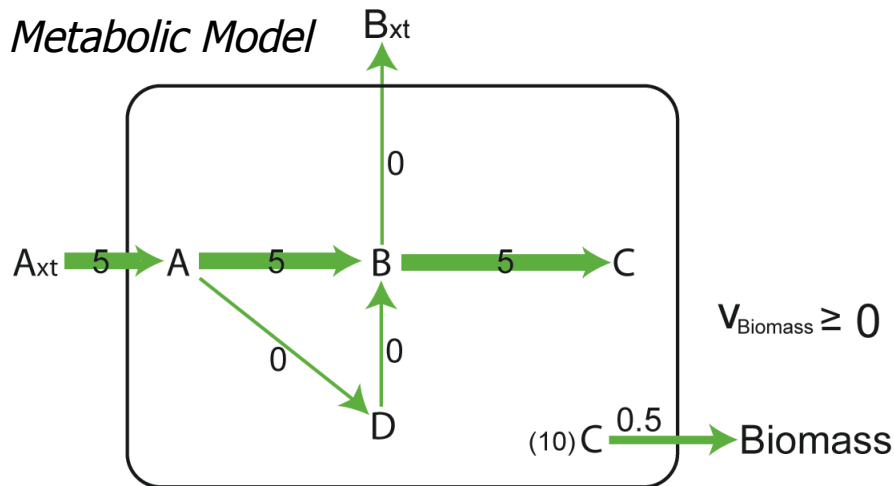


Integrated Models of Metabolism and Regulation ⁹

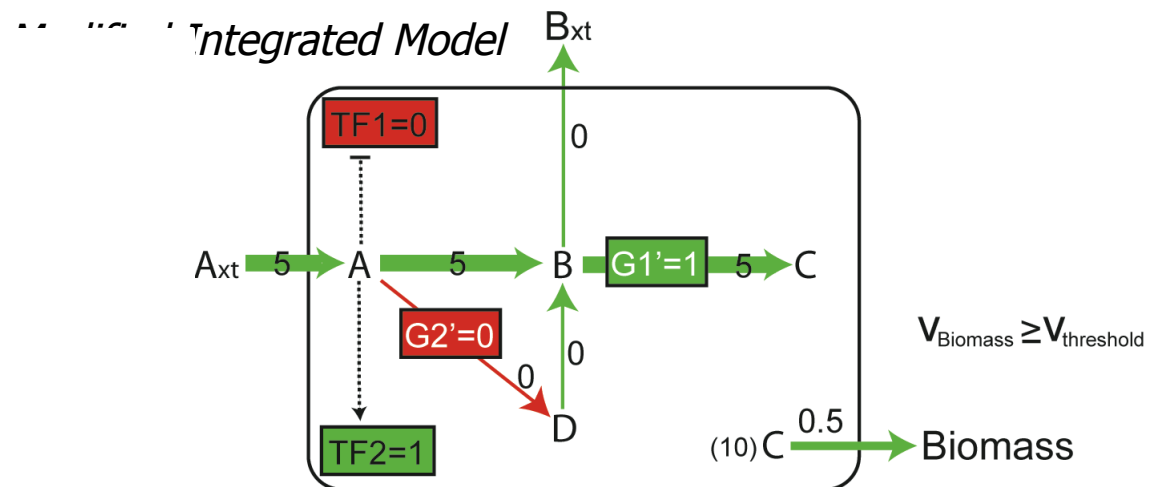


Approach for Relaxing Regulatory Constraints to Improve Accuracy

10



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

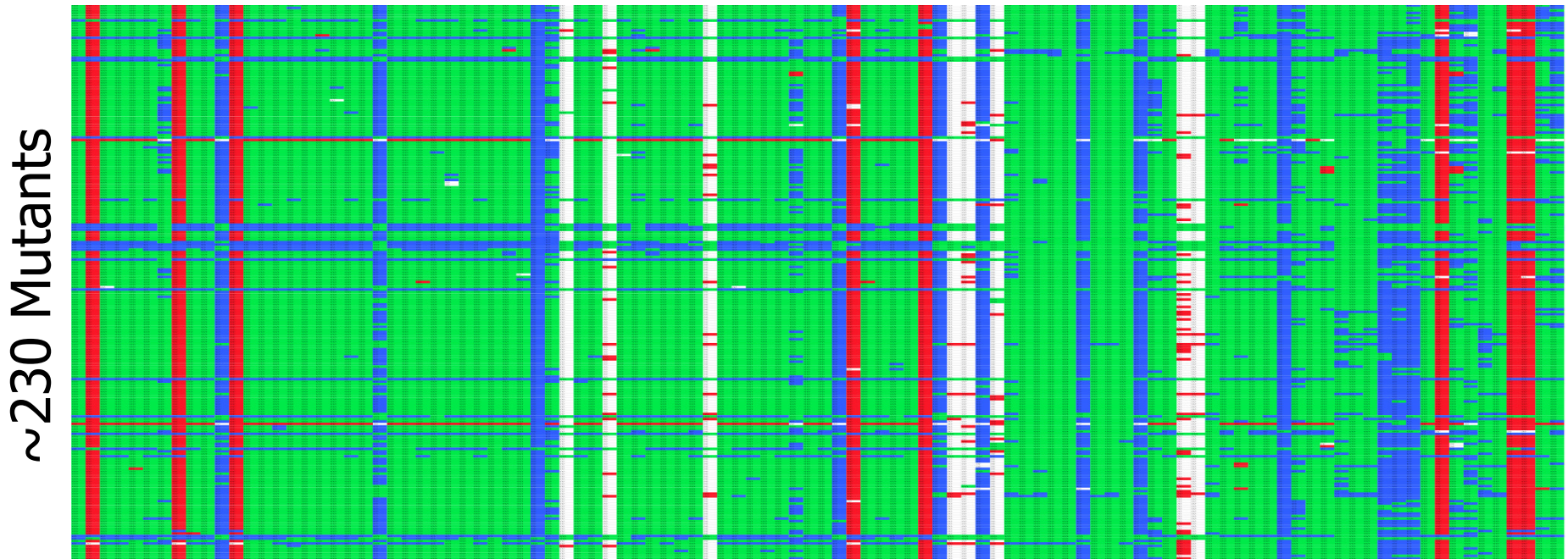


Barua, Kim and Reed. *PLoS Comput Biol* 6(10):e1000970 (2010)



Analysis of $\sim 32,000$ *E. coli* Mutant Growth Phenotypes

~ 130 Conditions



Reg. Model Correct

Both Models Incorrect

Reg. Model Incorrect

Both Models Correct

Data from ASAP Database

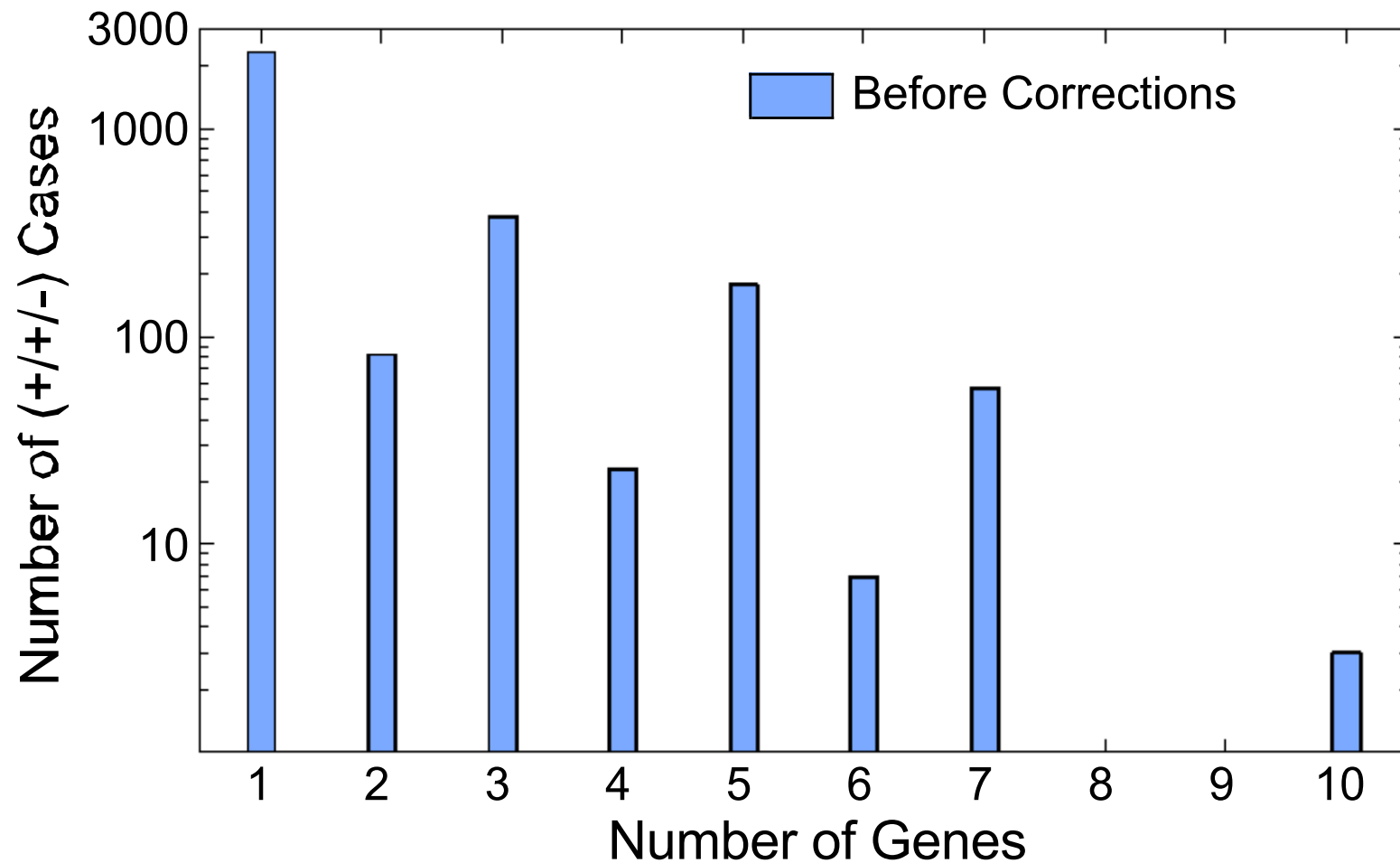


UW-Madison, Chemical & Biological Engineering

| | |
|--|----------------|
| Metabolic network ^a | iJR904 |
| Regulatory network ^b | iMC104 |
| Total comparisons ^c | 32,050 |
| Rule correction cases (+/+/-) | 3,079 |
| Rescue cases (-/+/-) | 2,041 |
| Integrated model accuracy ^d | 23,670 (73.9%) |

How Many Changes Are Needed to Correct Each False Prediction?

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



E. coli's Regulation of D-Alanine Transporter

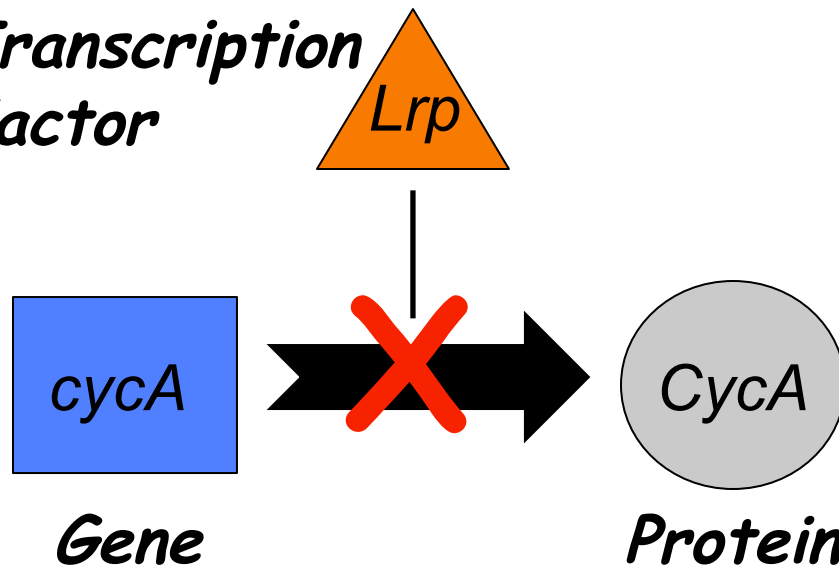
Experimental Result:

E. coli grows on D-alanine

Modeling Result:

Transporter is not expressed → No growth

**Transcription
Factor**



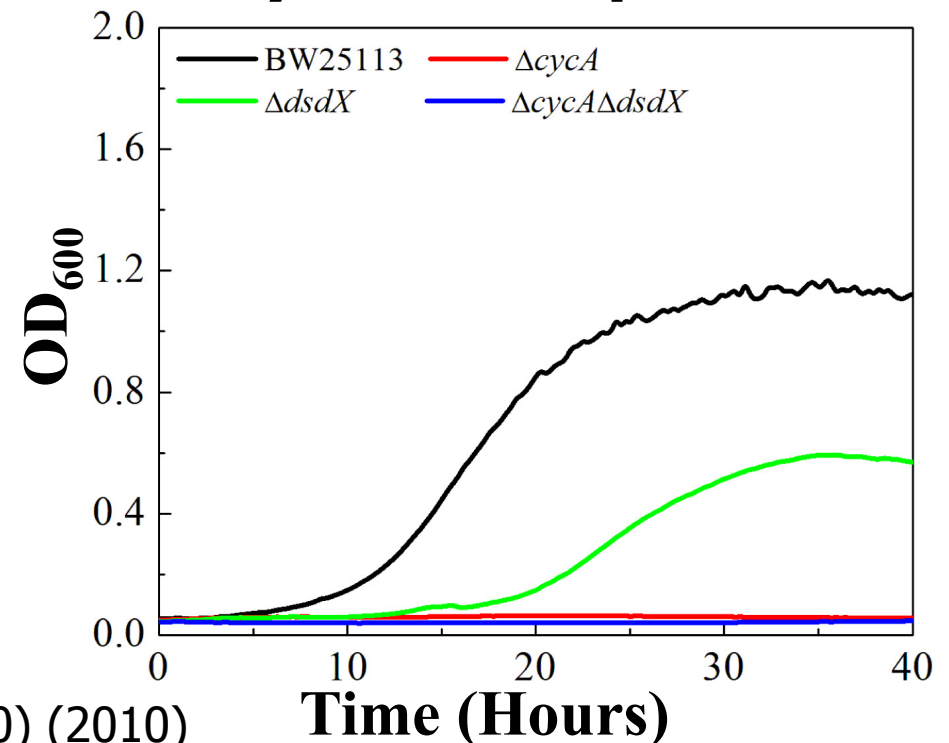
No Leucine



Lrp Active



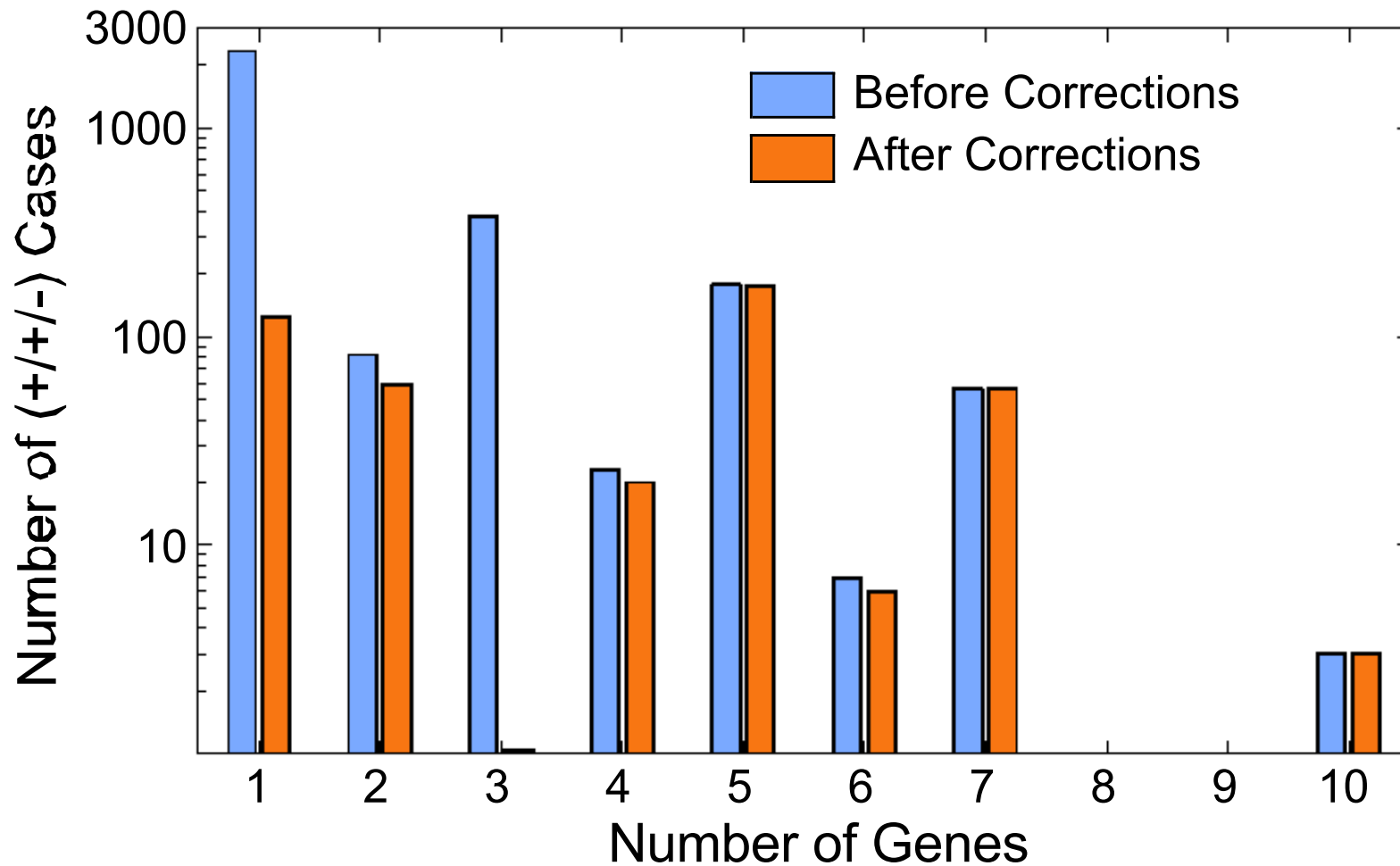
CycA Not Expressed



Effect of Only 11 Model Corrections

14

Before 3,079 Cases; After 445 Cases



Exploring Biochemical Networks by Integrating Datasets

15

1. Evaluating Network Structure:

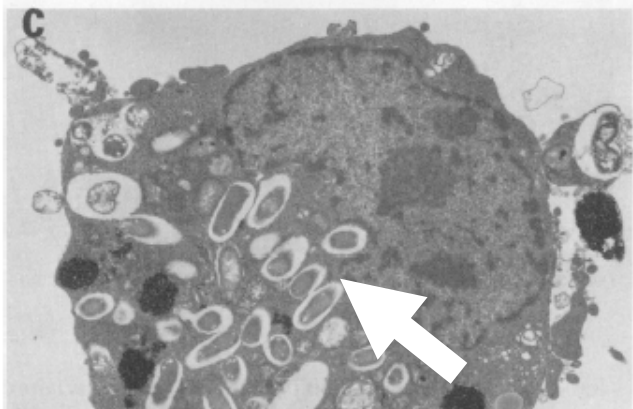
- Metabolism and Regulation
- *Escherichia coli*
- Genomic and phenotypic data

2. Evaluating Network Usage:

- Metabolism
- *Salmonella typhimurium* LT2 and *Shewanella oneidensis* MR-1
- Genomic, proteomic, transcriptomic, and phenotypic data



S. typhimurium Infection Requires Survival & ¹⁶ Growth in Host-Cell Environment



Carbohydrates

Glucose
Fructose
Fucose
Galactitol
Galactonate
Galactose
Glucarate
Galactarate
Glucosate

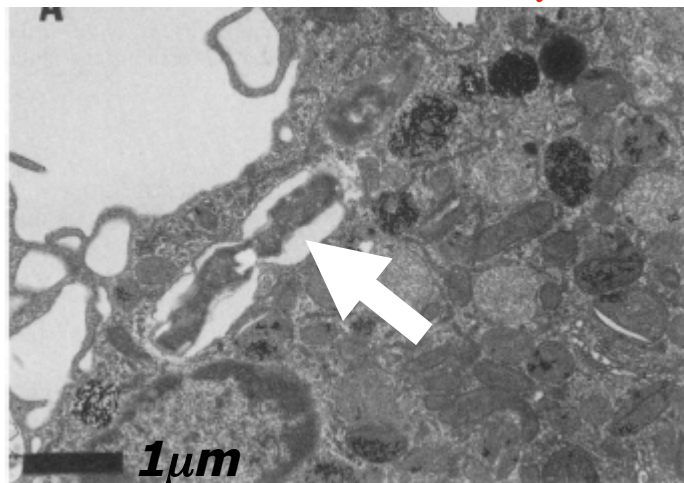
Amino Acids

Alanine
Arginine
Asparagine
Cysteine
Glutamic acid
Histidine
Isoleucine
Leucine
Lysine

Minerals & Inorganic Molecules

Sodium
Chloride
Sulfate
Potassium
Phosphate
Calcium
Magnesium
DMSO

What is the intracellular environment in the host cell providing to the bacteria?



Mannose
Melibiose
N-acetylglucosamine
N-acetylneuraminic acid
Rhamnose
Ribose
Sorbitol
Cellobiose
1,2-propanediol

Nucleosides

Inosine
Hypoxanthine
Deoxycytidine
Thymidine
Uracil
Uridine
Deoxyadenosine
Adenosine
Guanosine
Cytosine

Sulfite

Amines

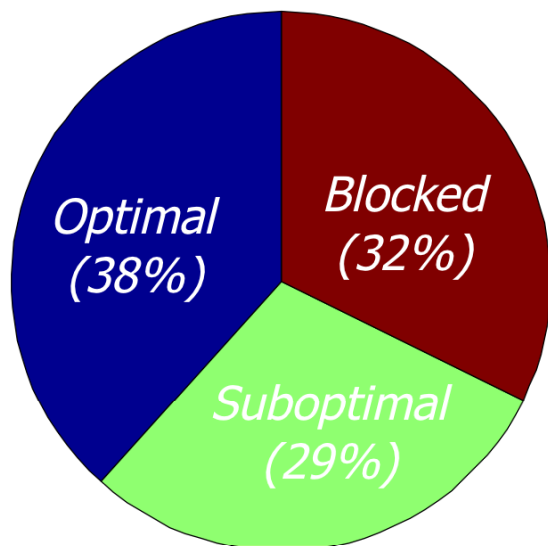
Allantoin
Carnitine
Polyamine
Ethanolamine

Vitamins

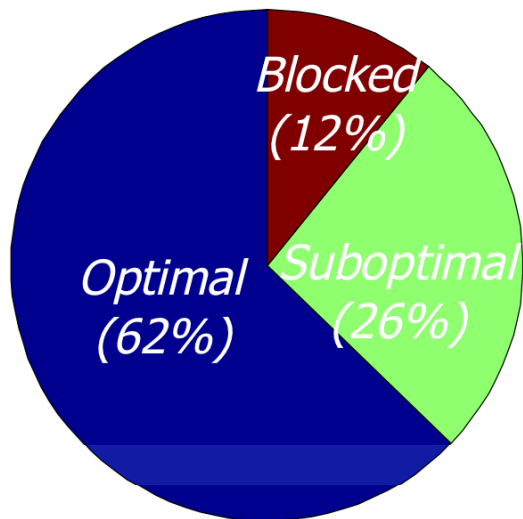
Thiamine (B1)
Pantothenate (B5)



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

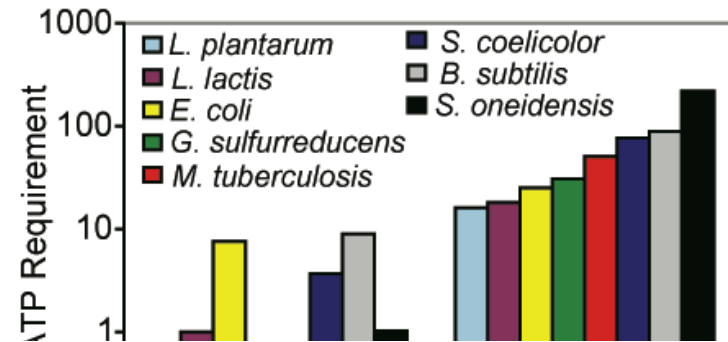
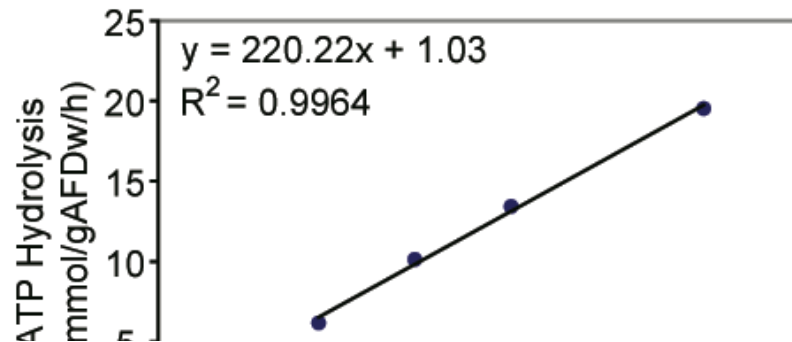
Purine Biosynthesis (4)

Amino Acid Biosynthesis (9):

- Threonine, Cysteine, Arginine, Asparagine



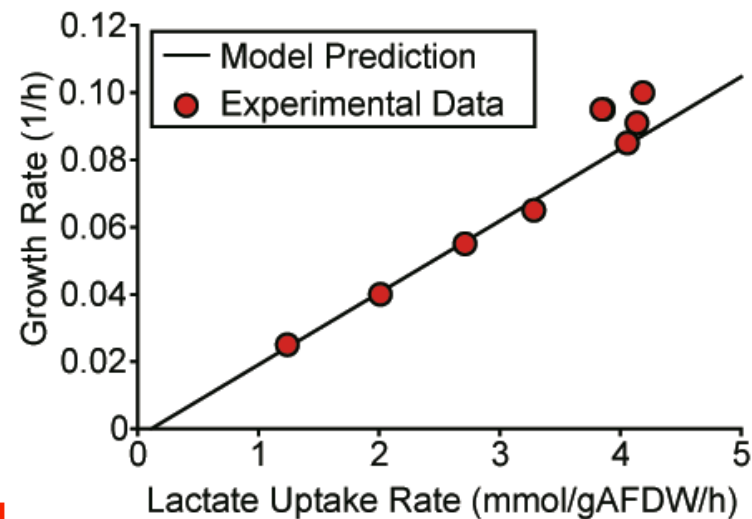
Shewanella Growth and Non-Growth Associated ATP Requirements



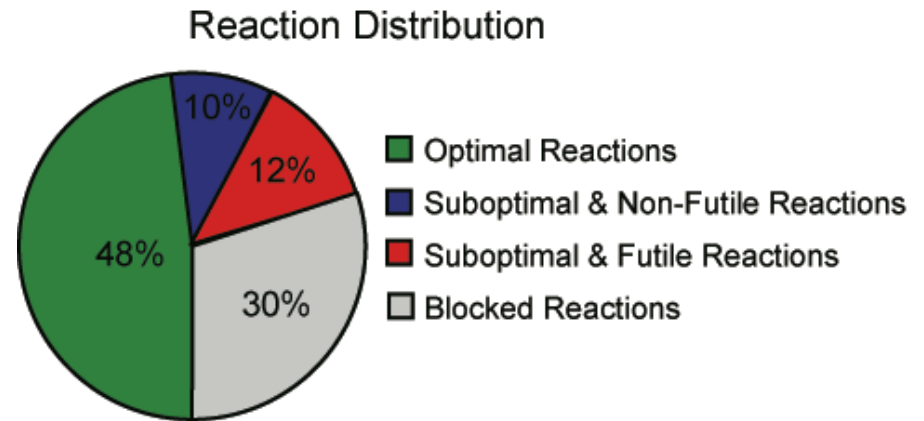
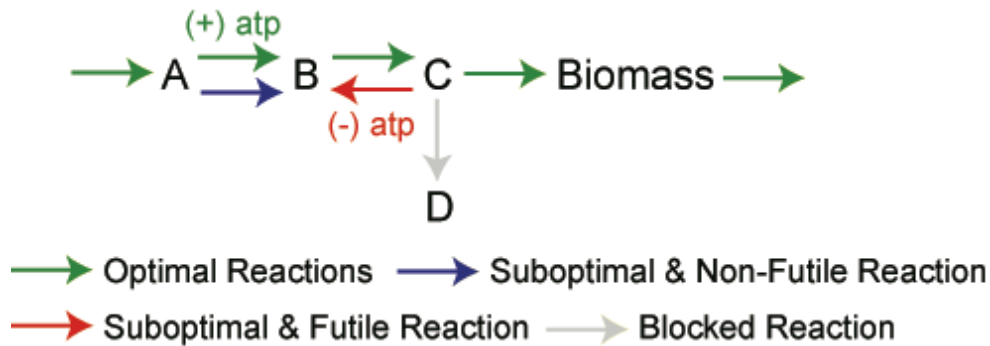
The high GAR is unrealistic. Why is MR-1 metabolically inefficient under these conditions?

$$\text{GAR} = 220.2 \text{ (mmol/gAFDW)}$$

$$\text{NGAR} = 1.03 \text{ (mmol/gAFDW/hr)}$$

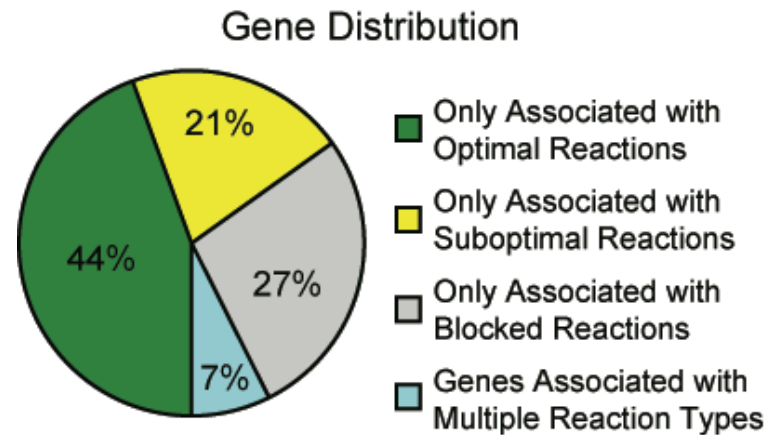


Characterization of Reactions (Aerobic Growth on Lactate)



Hypothesis:
*Expect High Expression
 of Optimal Genes*

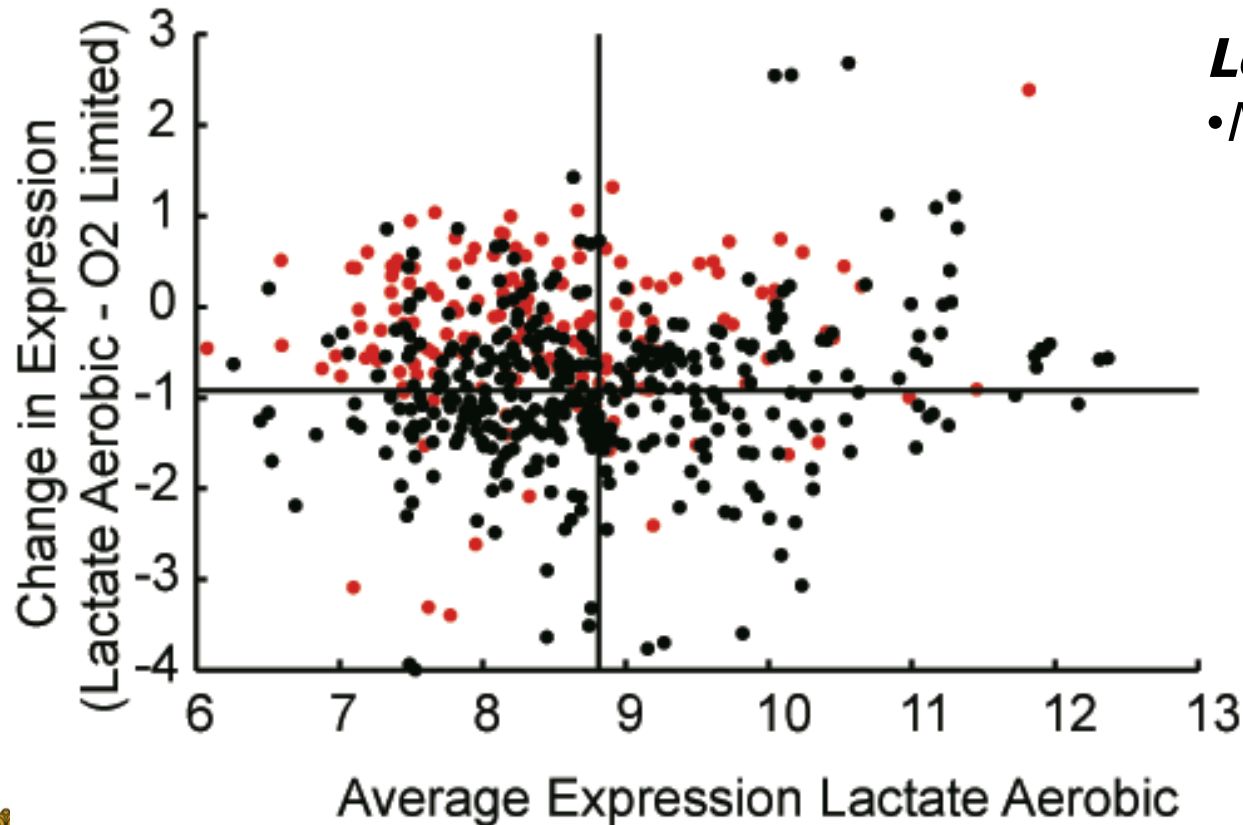
Expect Low Expression
 of Suboptimal Genes



Expression of Optimal and Suboptimal Genes

Potential Futile Cycles:

- *Pyruvate Kinase* + *Phosphoenolpyruvate synthase*
- *Phosphoenolpyruvate carboxylase* + *Malic enzyme*
- *Fatty acid synthesis* + *degradation*



Less Efficient Enzymes:

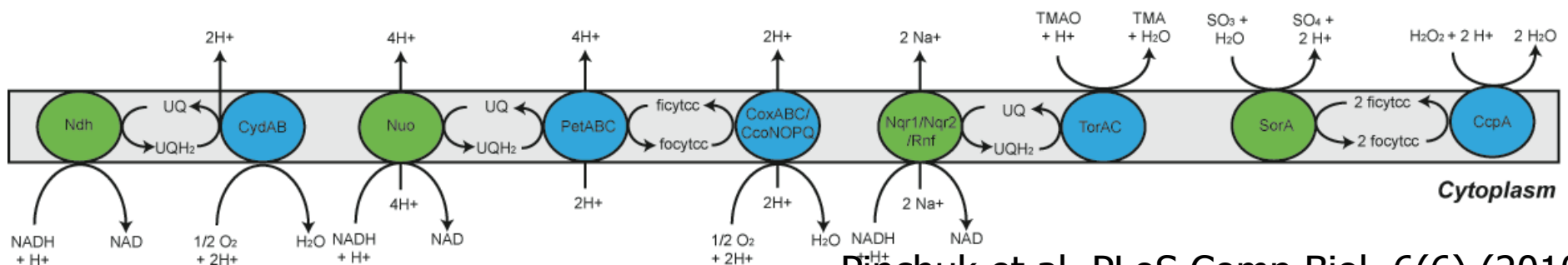
- *Ndh* and *Nqr* > *Nuo*

*Data from FedEx 2
Experiment from
M3D Database at BU*



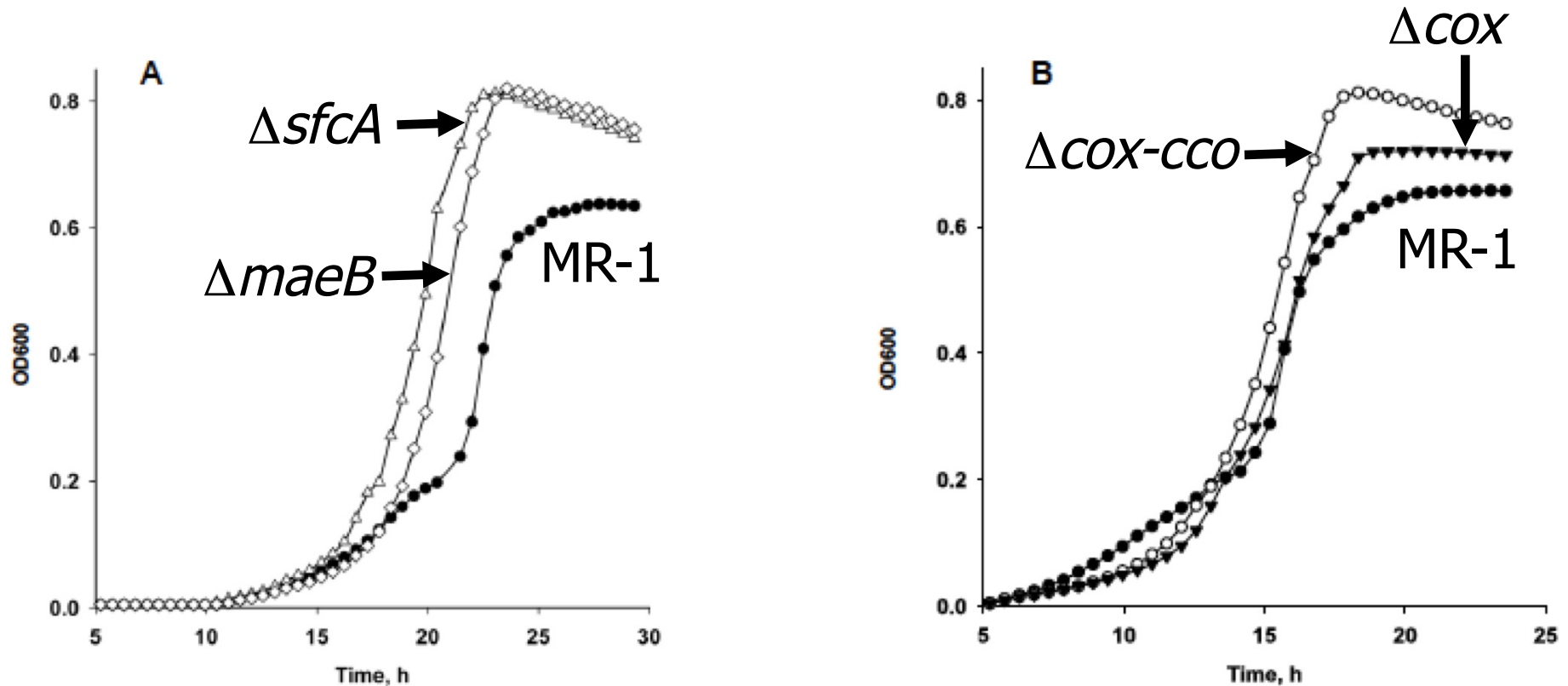
Possible Reasons for Less Efficient Growth (High apparent GAR)

- Futile Cycling
 - Three times lactate uptake rate: GAR ~ 80
- Protein Turnover
 - Each peptide bond hydrolyzed 7 times: GAR ~ 80
- Inefficient Use of Electron Transport Chain
 - 0.5 to 2.5 ATP per electron pair
 - Simulations done with 1.7 ATP per electron pair
 - Using 0.5 ATP per electron pair: GAR ~ 80



S. oneidensis MR-1 Mutant Phenotypes

22



Deletion of malic enzymes improves growth
→ Futile cycle involving malic enzyme

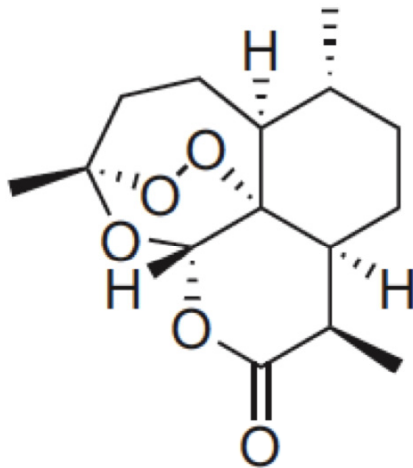
Deletion of Cox-Cco improves growth
→ MR-1 uses Cyd ($2H^+/2e^-$)



Leveraging Biochemical Networks

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

Drugs



Artemisinin

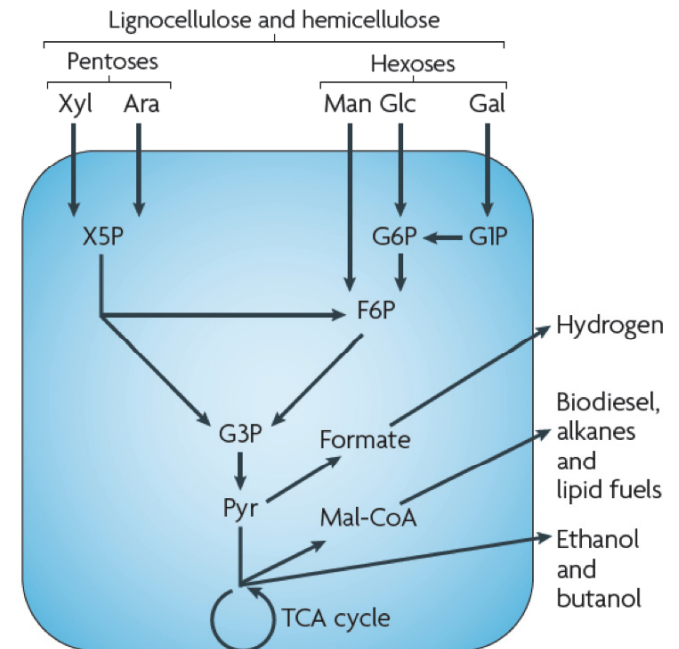
Commodity Chemicals



The miracles of science™

Sorona®

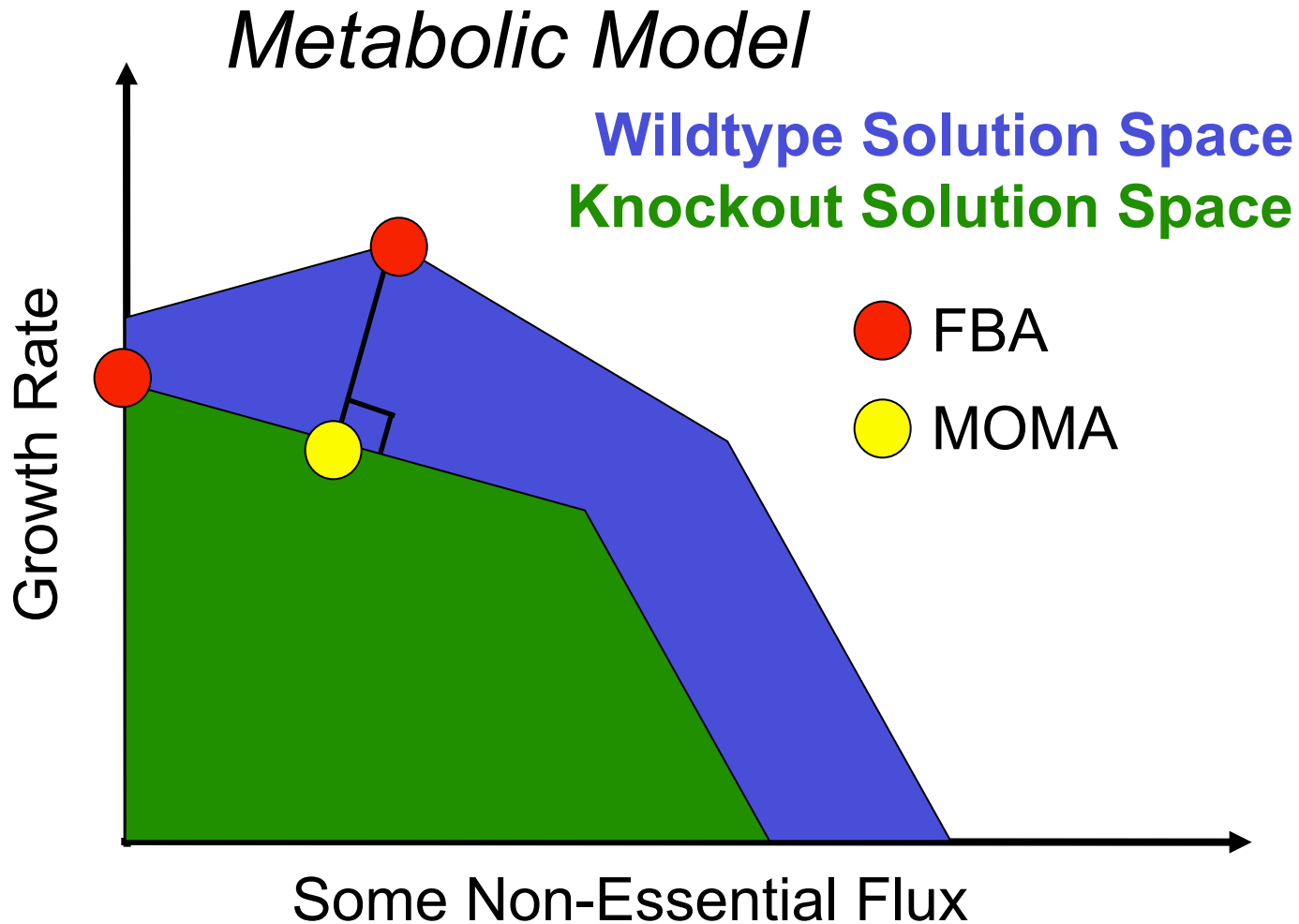
Biofuels



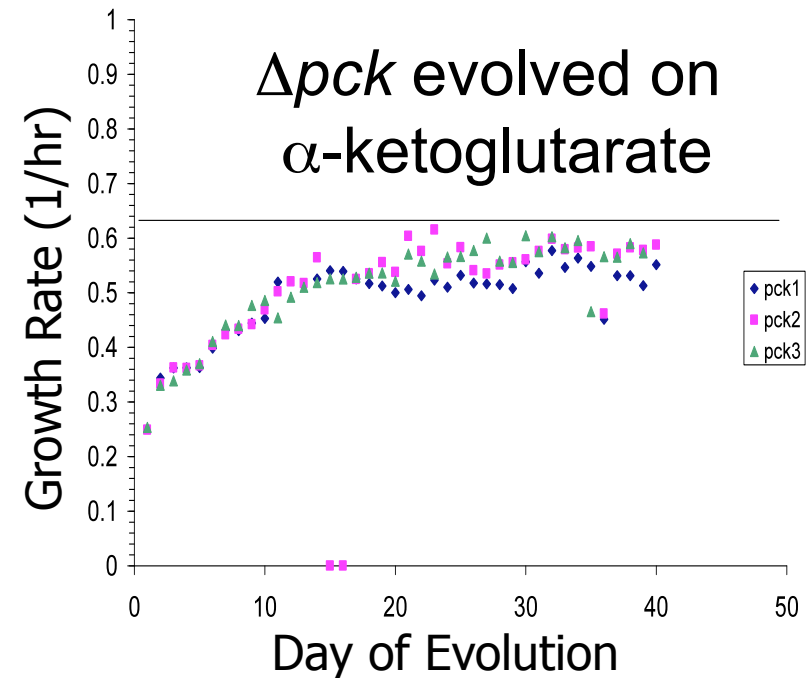
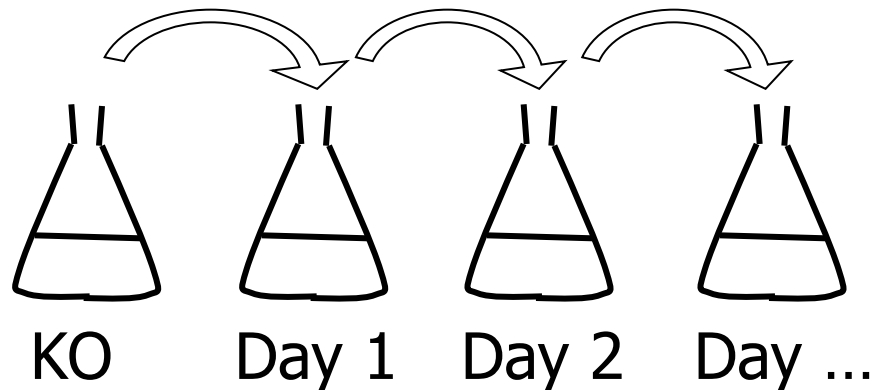
Alper & Stephanopoulos
Nat Rev Microbiol (2009)



Mutant Prediction Methods



What Happens if Cells Evolve?

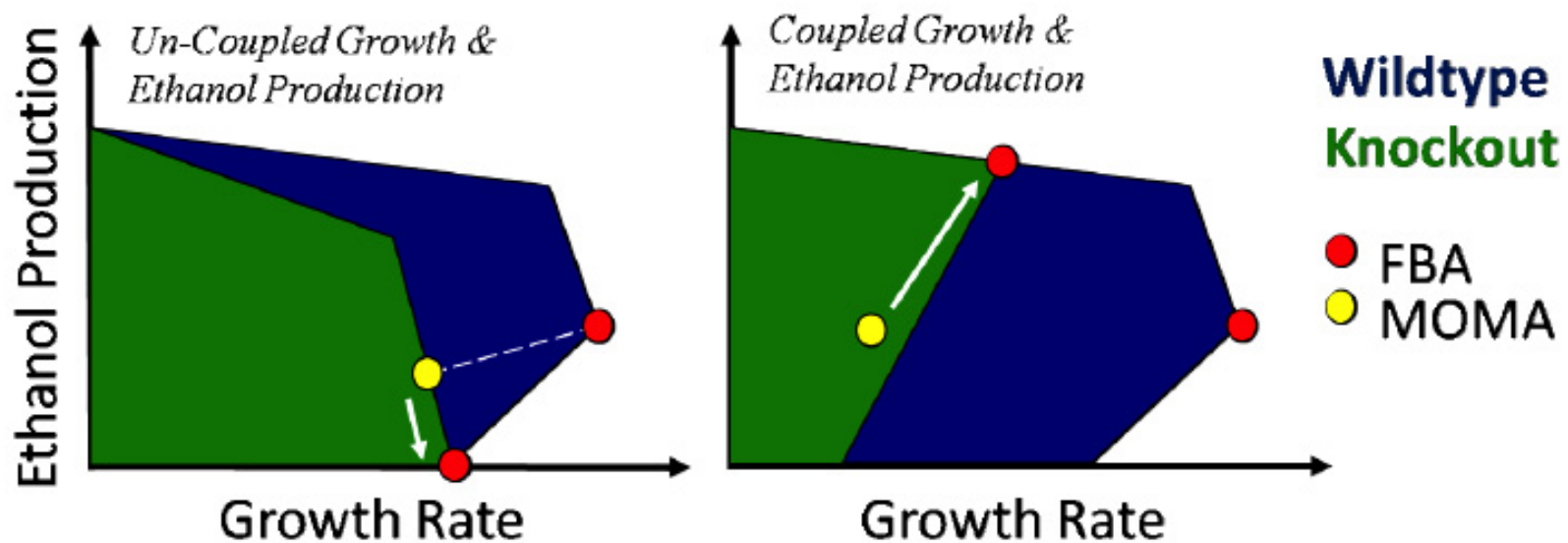


Faster growing cells outcompete others and select for cells with higher growth rates



OptKnock: Identifies Mutants with Coupled Biomass & Metabolite Production

Knockout Production Capabilities



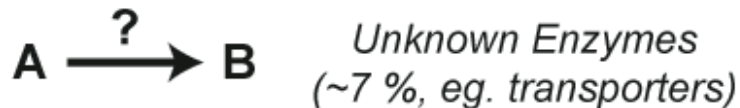
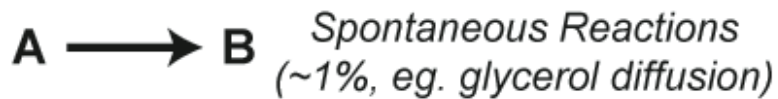
Finds reactions, that if removed, **couple biomass production to metabolite production** (ie. higher growth = higher production)

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

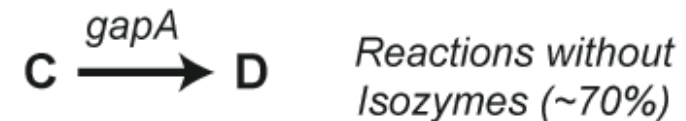
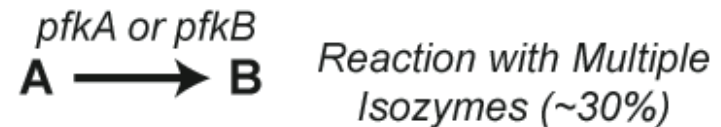


Benefit of Considering Genes and Regulation

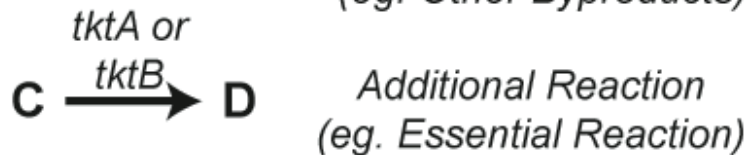
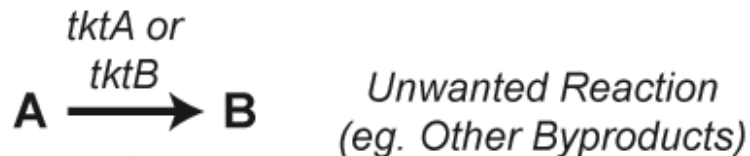
I. Reactions without Genes



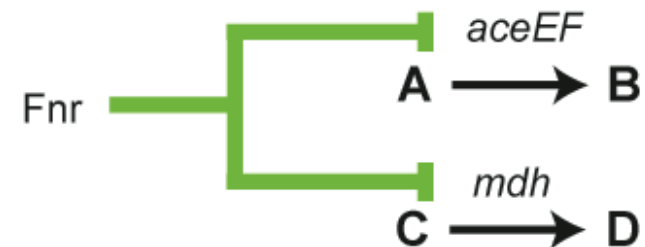
II. Reactions with Isozymes



III. Different Phenotypic Behavior



IV. Transcription Factor Prediction

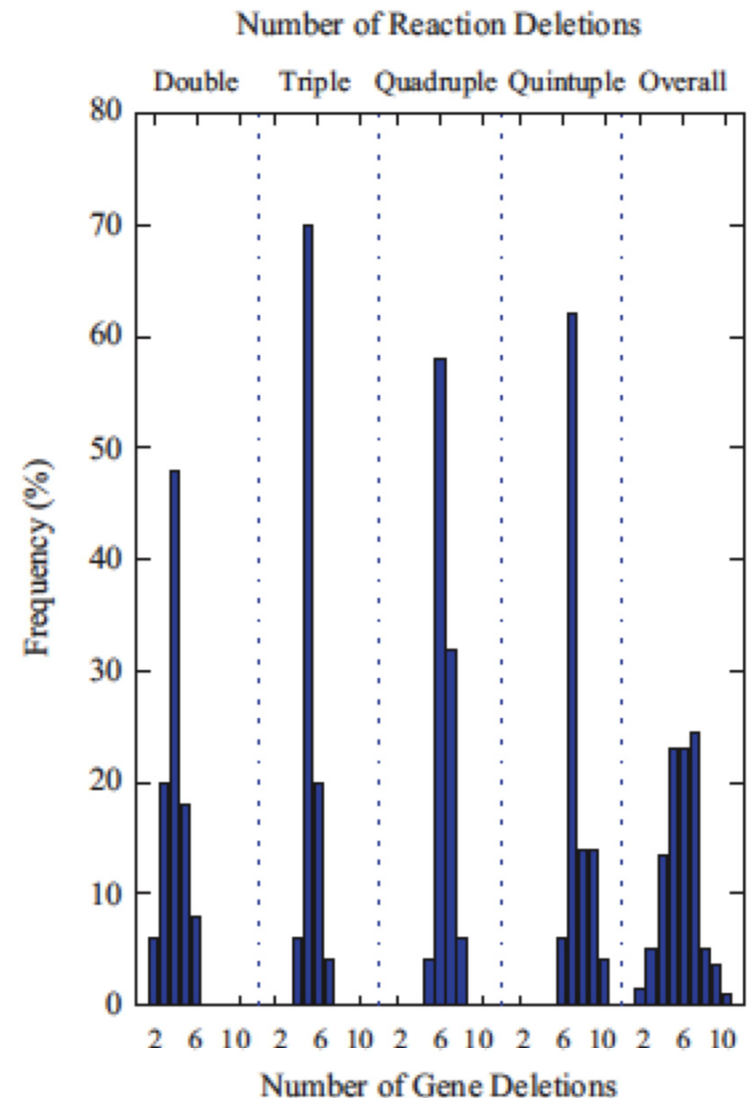


Deletion of Single Transcription Factor
Affects Multiple Genes & Reactions

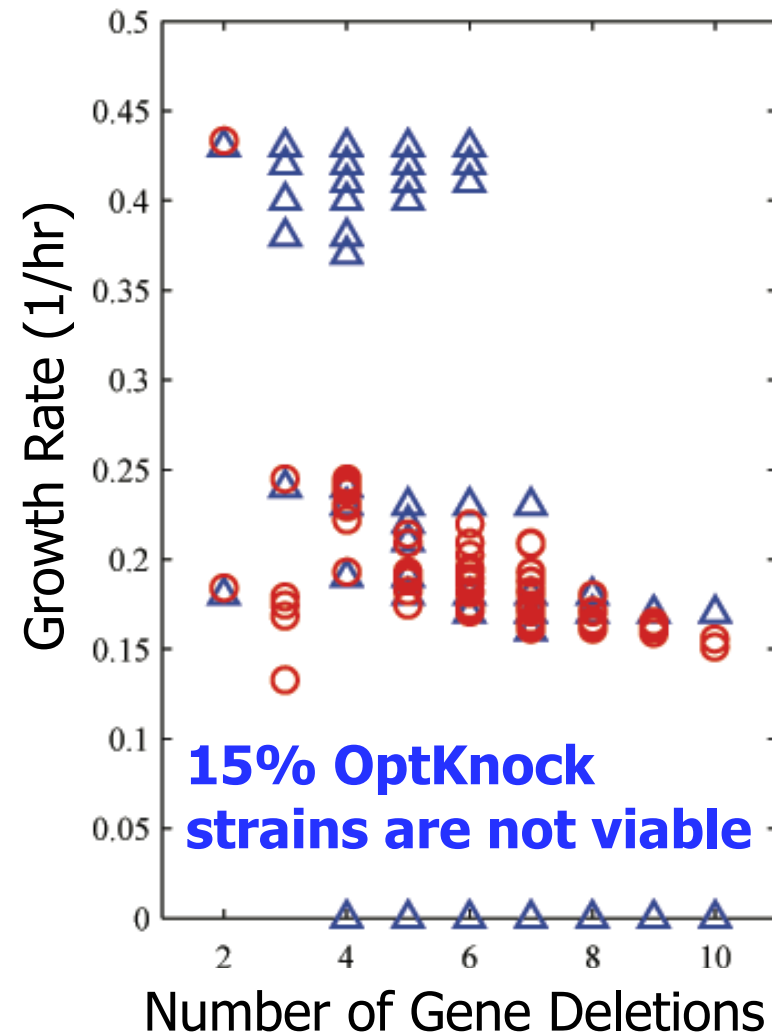
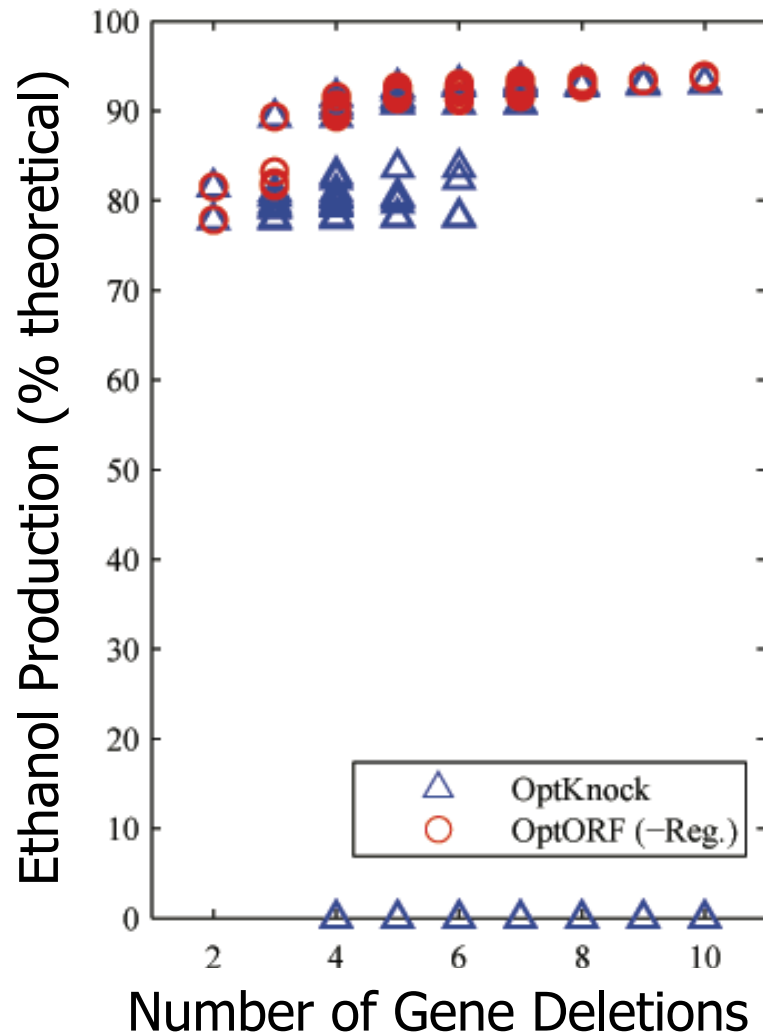


Deleting by Gene versus Reaction

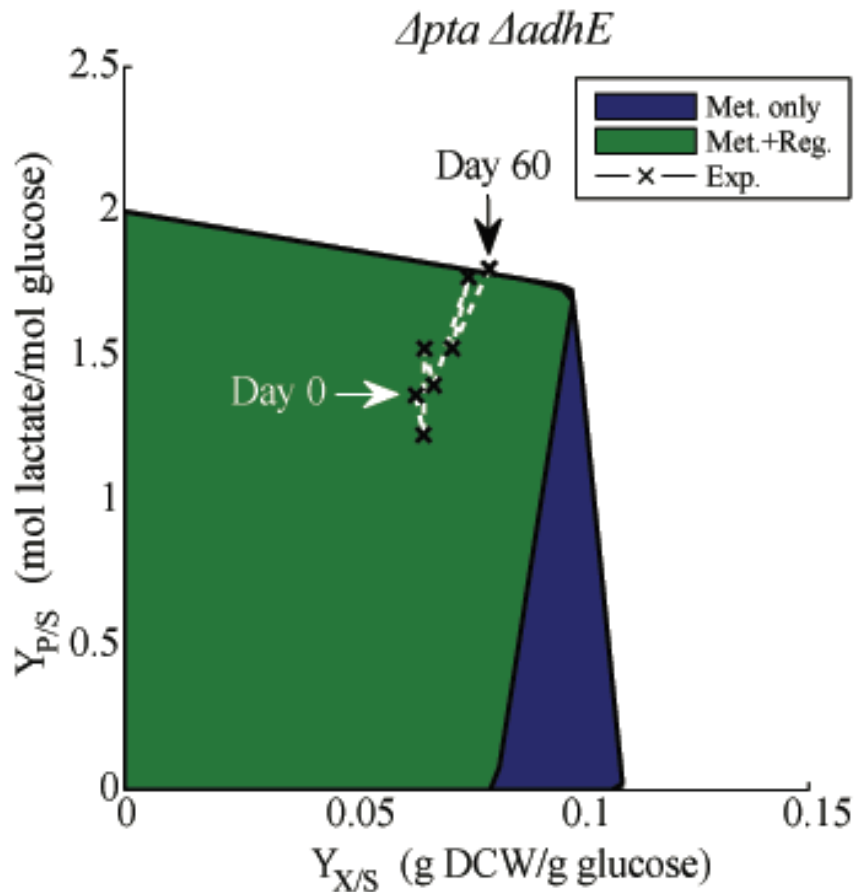
1. 200 Total OptKnock Strategies
 - 50 Double Reaction Deletions
 - 50 Triple Reaction Deletions
 - 50 Quadruple Reaction Deletions
 - 50 Quintuple Reaction Deletions
2. Mapped reaction deletions to gene deletions
 - OptKnock Strategies had between 2 and 10 genes
3. Found OptORF strategies with the same number of gene deletions



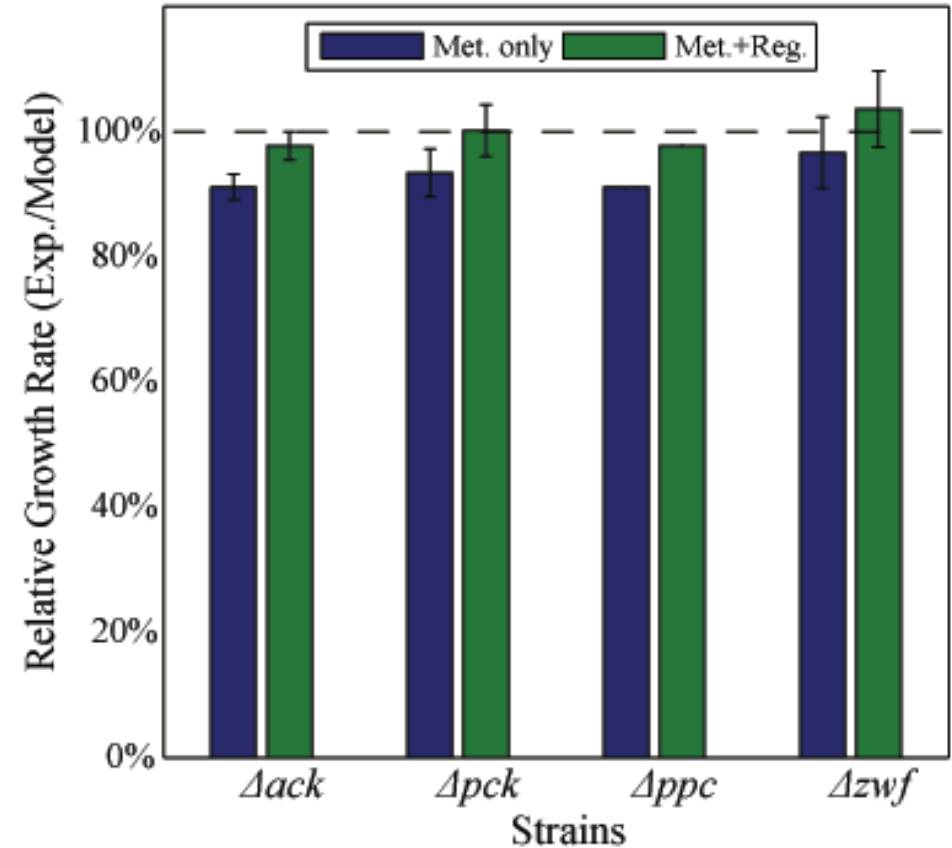
Deleting by Gene versus Reaction



Adaptive Evolutionary Outcomes are Consistent with Regulatory Predictions



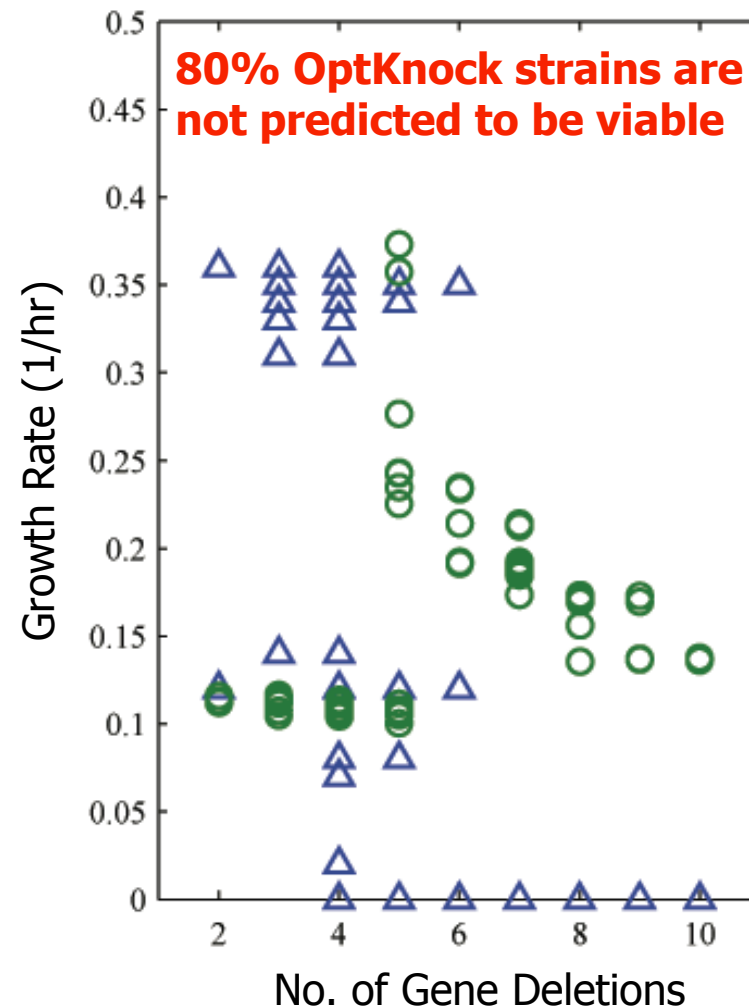
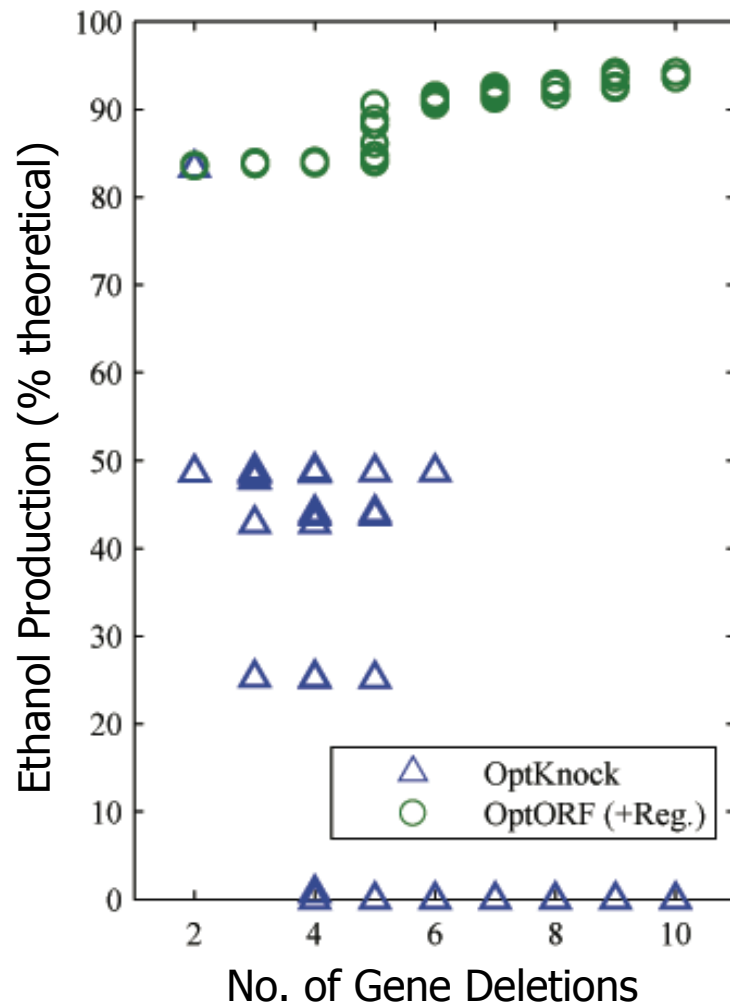
Data from S.S. Fong et al. *Biotech & Bioeng* (2005)



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

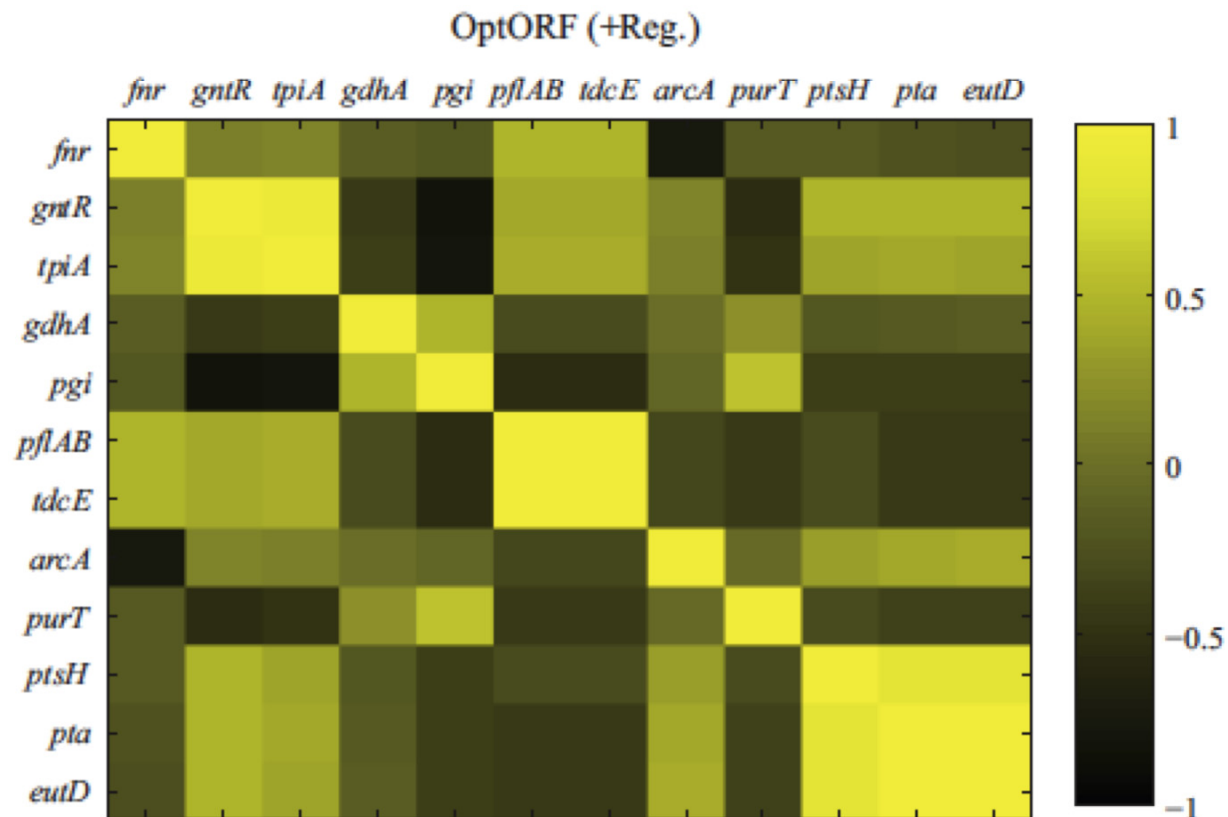


Transcriptional Regulation Restricts Growth and Ethanol Production



Patterns of Mutations for Improving Ethanol Production

32



Anti-Correlated Mutations:

- *arcA* and *fnr*
- *pgi* and *tpi* (and *gntR*)

Correlated Mutations:

- *gntR* and *tpiA*
- *pflAB* and *tdcE*
- *ptsH*, *pta*, and *eutD*



Strains for Ethanol and Isobutanol (via BCAA pathways) Production

| ETHANOL: Gene Deletions | Gene Over- Expression | Growth Rate | Ethanol Production (% max yield) |
|---|--------------------------|----------------|--|
| $\Delta fnr \Delta pflB \Delta tdcE \Delta pgi$ | <i>edd</i> | 0.225 | 86.2% |
| $\Delta fnr \Delta pflB \Delta tdcE \Delta tpi$ | <i>edd</i> | 0.235 | 90.5% |
| $\Delta fnr \Delta pflB \Delta tdcE \Delta tpi \Delta gdhA$ | <i>edd</i> | 0.214 | 91.4% |
| $\Delta arcA \Delta pta \Delta eutD \Delta tpi \Delta ptsH$ | <i>edd</i> | 0.192 | 91.6% |

| ISOBUTANOL: Gene Deletions | Gene Over- Expression | Growth Rate | Isobutanol Production (% max yield) |
|---------------------------------------|--------------------------|----------------|---|
| $\Delta adhE \Delta gdhA$ | | 0.223 | 89.5% |
| $\Delta gntR \Delta adhE \Delta pgi$ | | 0.128 | 93.8% |
| $\Delta adhE \Delta tpi$ | <i>edd+fbp</i> | 0.128 | 94.3% |
| $\Delta adhE \Delta pntA \Delta nuo$ | <i>edd+fbp</i> | 0.110 | 95.1% |
| $\Delta adhE \Delta pntA \Delta gdhA$ | <i>edd+fbp</i> | 0.102 | 95.5% |



Strategy with Gene Deletions & Gene Over-expression

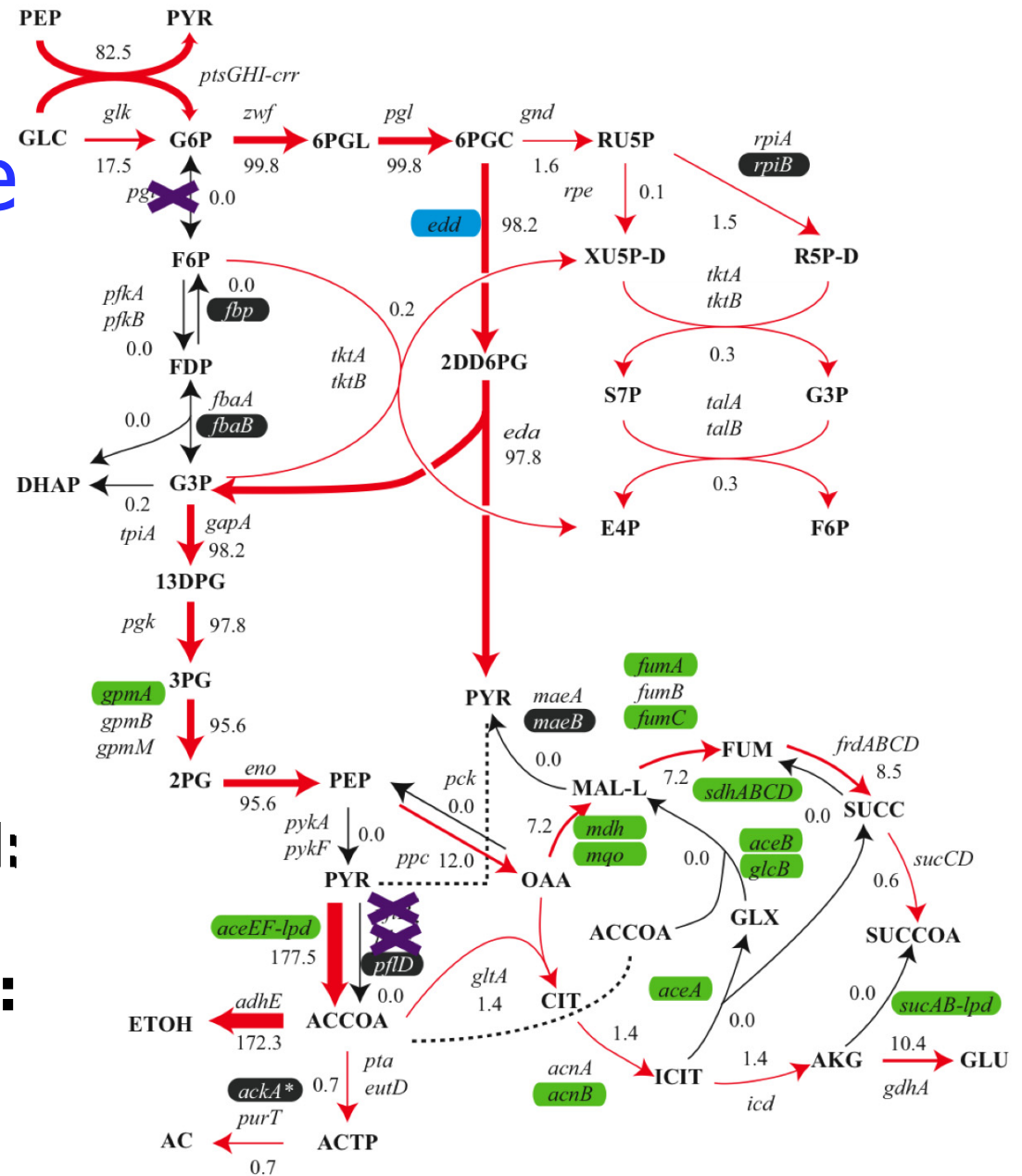
Mutations
 Δfnr $\Delta pflAB$
 $\Delta tdcE$ Δpgi
+edd

Predicted Ethanol Yield:

86%

Predicted Growth Rate:

0.225 hr⁻¹



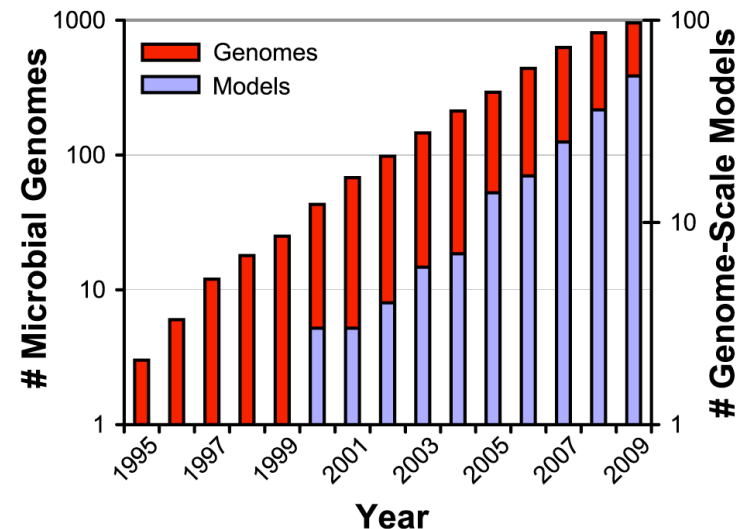
Strain Design Summary

- OptORF is general and can be applied to **other microbes** or for production of other **products** (eg. isobutanol).
- Modeling accounts for the **local and global** affect of mutations to predict behaviors.
- Relatively easy to couple growth and ethanol production under anaerobic conditions, **OptORF provides simplest genetic strategies.**
- Can identify **novel** metabolic engineering **strategies.**



Concluding Comments

- Genome sequencing has enabled the **rapid development** of genome-scale metabolic models.
- Models can be to **predict** or **describe** cellular behavior
- Models can **provide context** for experimental data. New methods for using 'omics' data to further **constrain models** are appearing (e.g. gene expression data).
- Model-data **inconsistencies can be interesting**, they can indicate problems with models, data, and/or our understanding of biological networks.



Acknowledgements

Research Group:

Dipak Barua (LANL)

Camo Cotten
Josh Hamilton

Joonhoon Kim

Bastian Niebel
Wai Kit Ong
Chris Tervo
Trang Vu

Xiaolin Zhang

Funding Sources:

Metabolic Engineering

Christos Maravelias

Salmonella Collaborators:

Anu Raghunathan (MSSM)

Simon Daefler (MSSM)
Sookil Shin (MSSM)
Bernhard Palsson (UCSD)

Shewanella Collaborators:

Alex Beliaev (PNNL)

Jim Fredrickson (PNNL)

Oleg Geydebrekht (PNNL)

Allan Konopka (PNNL)

Grigoriy Pinchuk (PNNL)

Samantha Reed (PNNL)

Margie Romine (PNNL)

Jessica De Ingeniis (Burnham)

Andrei Osterman (Burnham)

James Scott (Dartmouth)



UW-Madison, Chemical & Biological Engineering