Cell Lessons That “We” Learned the Hard Way (but we can explain & deal better with now)

Michael M. Domach
Carnegie Mellon University
Pittsburgh PA 15232
md0q@andrew.cmu.edu

Take Aways

- Defining what makes a cell “good” more precisely can have advantages (additional screening tools, management, etc).
- One way to obtain the definition is familiar in concept to managing power-load distribution.
- The above has actually worked in some biosystems.

Bulk Bioprocesses

Problem →Less Problem →$ Value

-$ +$

Less waste
disposal drain

Multi-product
streams/markets
w/Σ $ value

Questions

To find, adapt, or engineer algae to

$CO_2 + h\nu \rightarrow$ Oil + Σ Useful Products/Markets

What are good cells?

Best culture/medium management?

What are good intracellular determining-attributes?

→More screening tools, patterning, base-lining…

[Attribute = transformative $step_i$ +/-]

Things that Can Confound

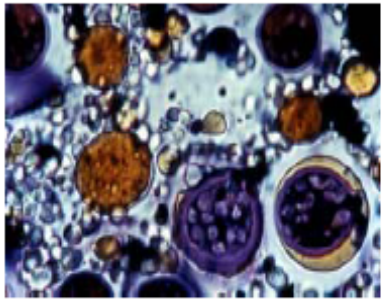
- Metabolic systems are redundant & state-switching can occur.
- Equal “phenotypes” are not equally robust.
- Not all metabolic politics are local.
- System validation produces different mechanistic explanations (poorly designed ^{13}C tracer analysis).

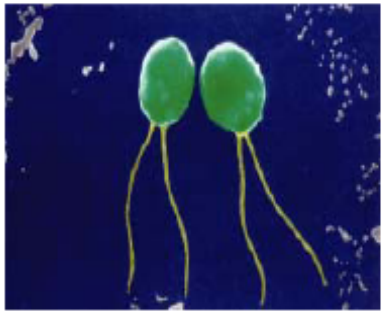
1. Local Politics

National Renewable Energy Laboratory

NREL/TP-580-24190

A Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae





Close-Out Report

1. Amplifying ACCase

National Renewable Energy Laboratory

Breakthroughs in molecular biology and genetic engineering.

Plant biotechnology is a field that is only now coming into its own. Within the field of plant biotechnology, algae research is one of the least trodden territories. The slower rate of advance in this field makes each step forward in our research all the more remarkable. Our work on the molecular biology and genetics of algae is thus marked with significant scientific discoveries. The program was the first to isolate the enzyme Acetyl CoA Carboxylase (ACCase) from a diatom. This enzyme was found to catalyze a key metabolic step in the synthesis of oils in algae. The gene that encodes for the production of ACCase was eventually isolated and cloned. This was the *first* report of the cloning of the full sequence of the ACCase gene in *any* photosynthetic organism. With this gene in hand, researchers went on to develop the first successful transformation system for diatoms—the tools and genetic components for expressing a foreign gene. The ACCase gene and the transformation system for diatoms have both been patented. In the closing days of the program, researchers initiated the first experiments in metabolic engineering as a means of increasing oil production. Researchers demonstrated an ability to make algae over-express the ACCase gene, a major milestone for the research, with the hope that increasing the level of ACCase activity in the cells would lead to higher oil production. These early experiments did not, however, demonstrate increased oil production in the cells.

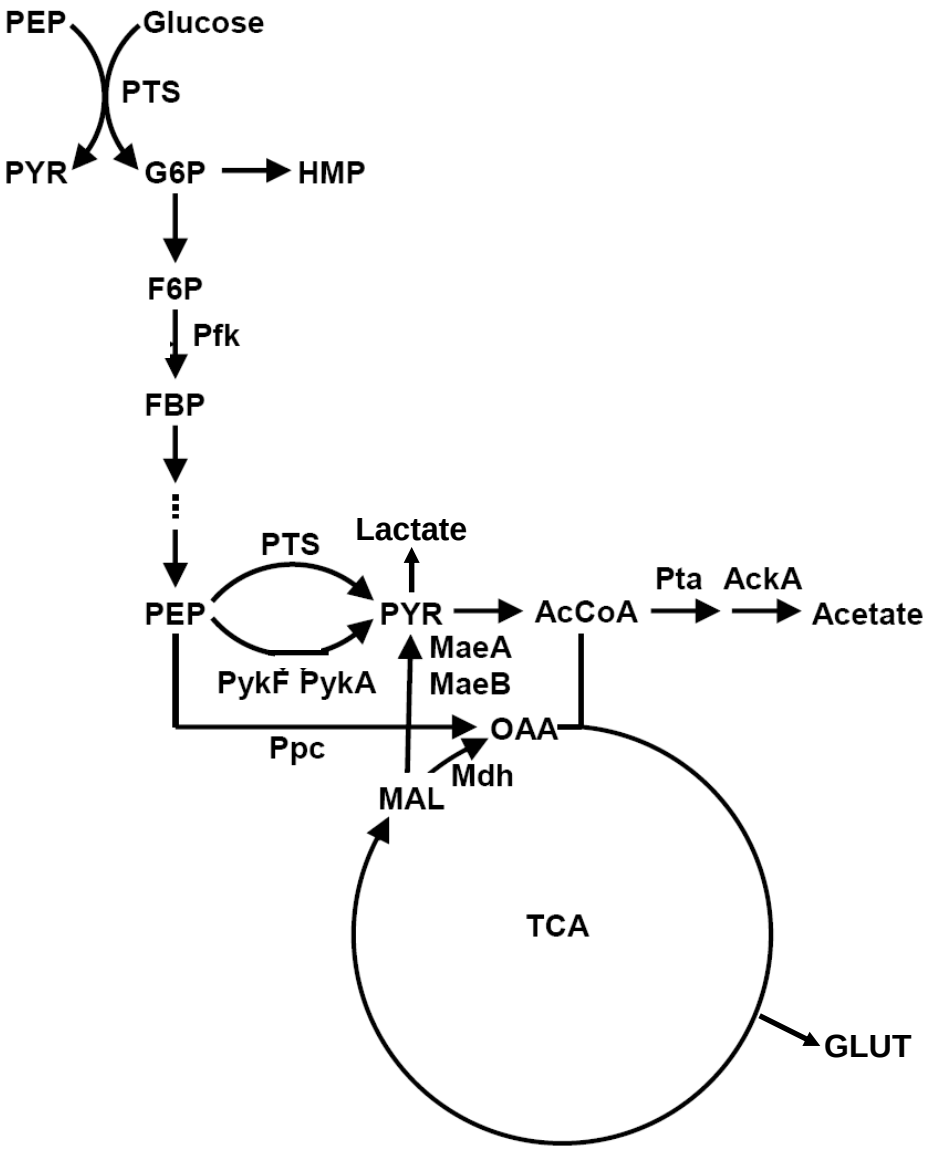
Ex. 2 Stop *E. coli* Acetate-Production

Best trait to look for or endow?

Minimal

***Pta capacity*?**

NO-Creates a new problem.



System Analysis-Informed Help

A model enables sorting &/or design because

(i) Inter & intra species variants/mutants have altered overall stoichiometries.

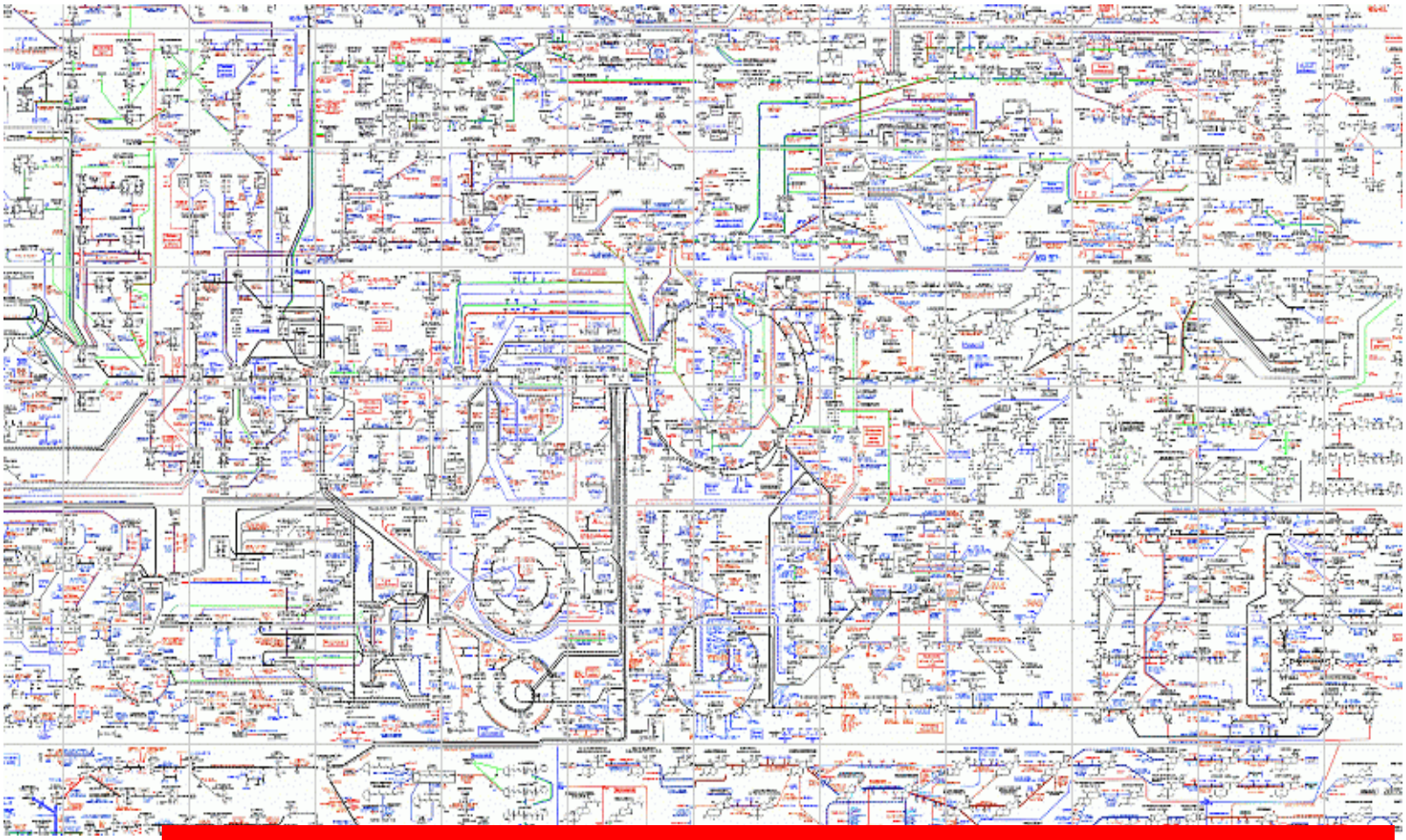
(ii) Nice to know the range before starting (yield horizon prediction).

(iii) The most obvious endowments can yield the least bang.

(iv) Other common markers likely exist.

(v) Know more-claim more; strengthens IP/claims.

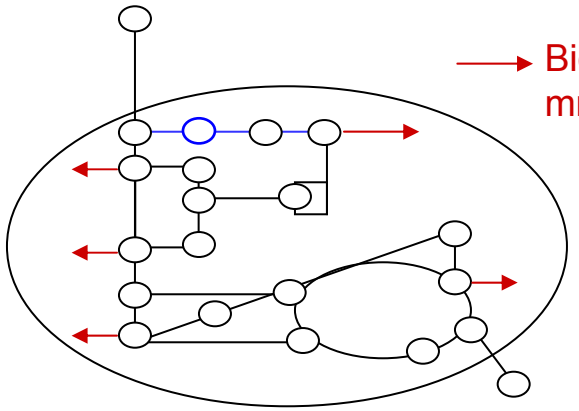
Analysis Means…



Thinking like a utility/power distribution company

Metabolic Net Analysis

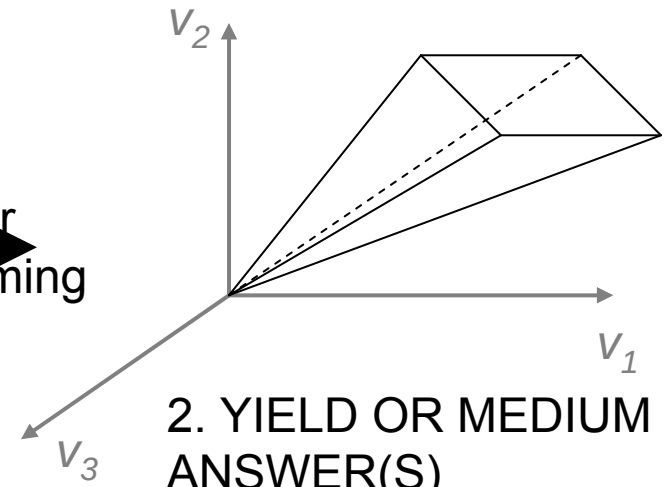
NETWORK RECONSTRUCTION



Mixed Integer
Linear Programming

1. SOLUTION SPACE

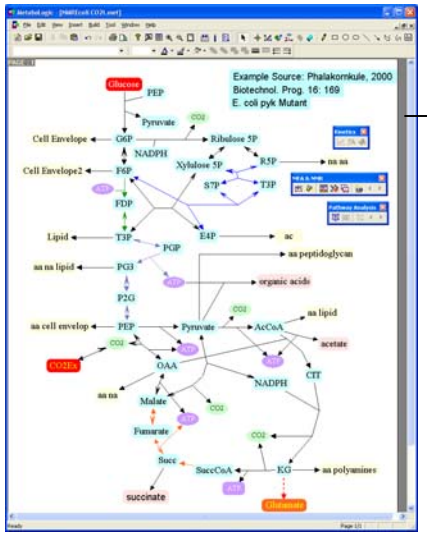
ALL MASS-ENERGY INTEGRATED
EXTREME POINT SOLNS THAT ATTAIN GOAL
OR UNDERLIE PERFORMANCE



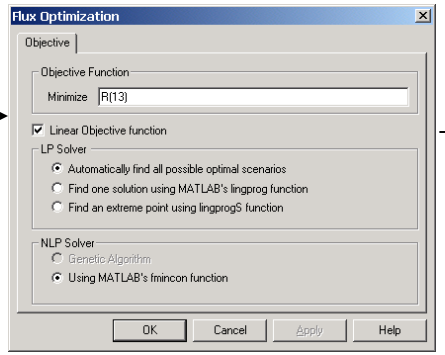
2. YIELD OR MEDIUM ANSWER(S)

min/max f(v) or [0]
Subject to:
[S] v = 0 (flux in = flux out)
biosyn flux loads met
Σ ATP formation $\geq$ Sufficient
Σ NADPH formation = Sufficient
Selected flux measurements to constrain further (optional)
Etc

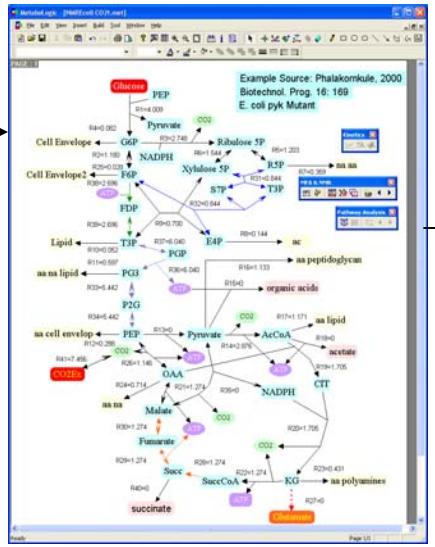
Modeling via METABOLGICA



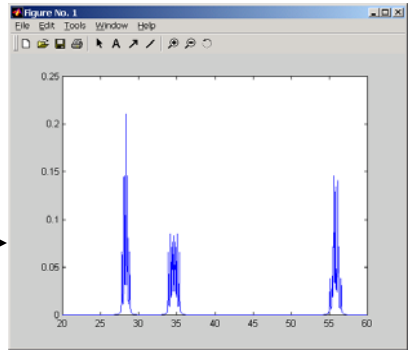
Draw Model Picture



State Objective, Mutation, &/or Input Extracellular Measurements: Compute Alternate Flux Patterns



Display Alternate Pathway Solutions

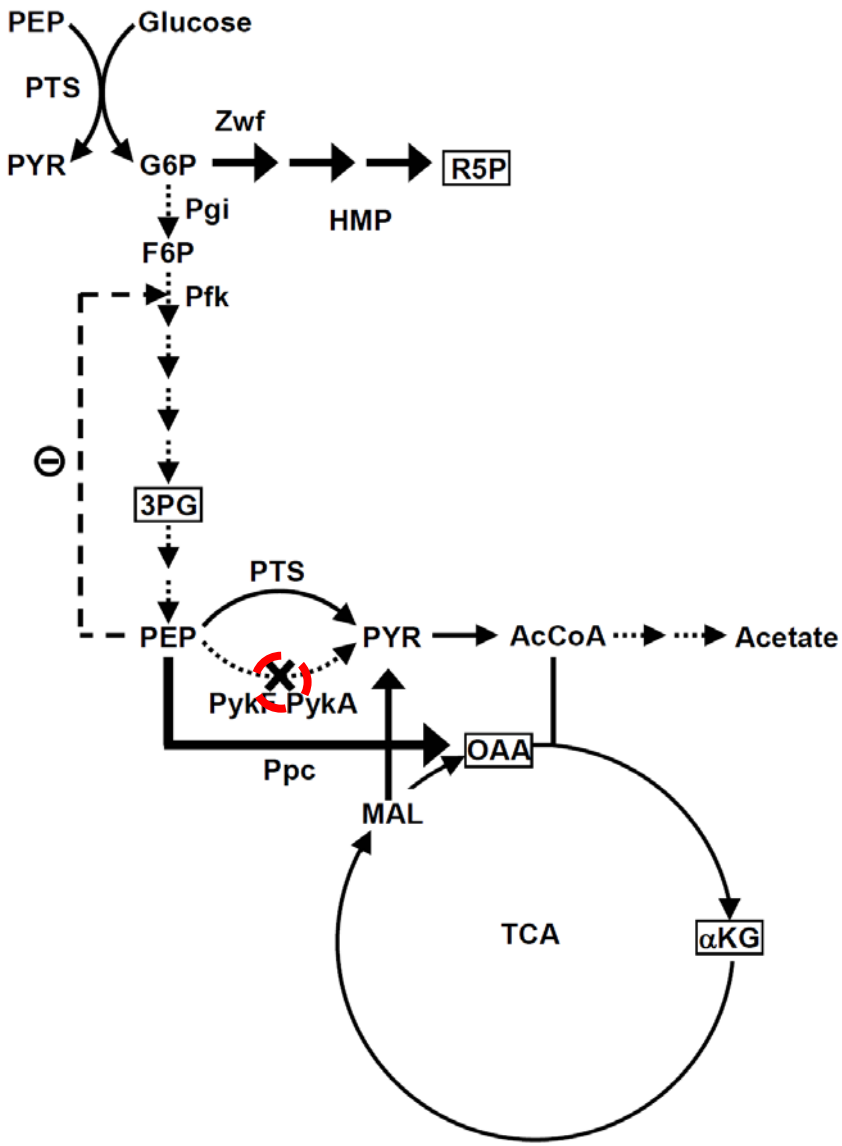


NMR Design & Prediction

Also Perform Inverse Problem

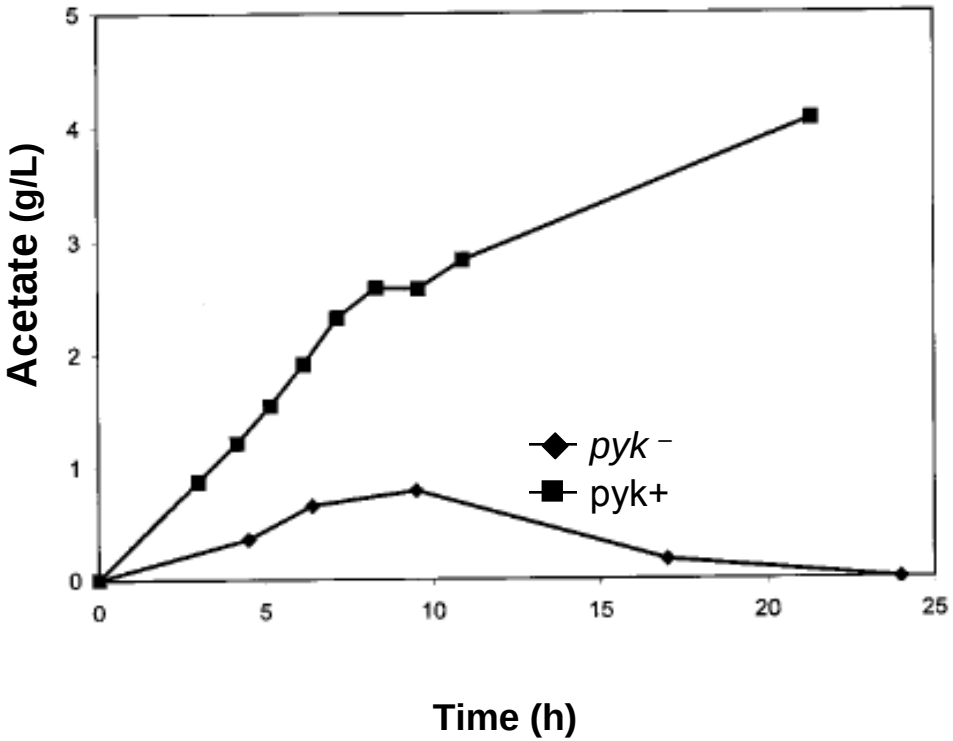
(Zhu *et al*, *Metabolic Engin*, 2003)

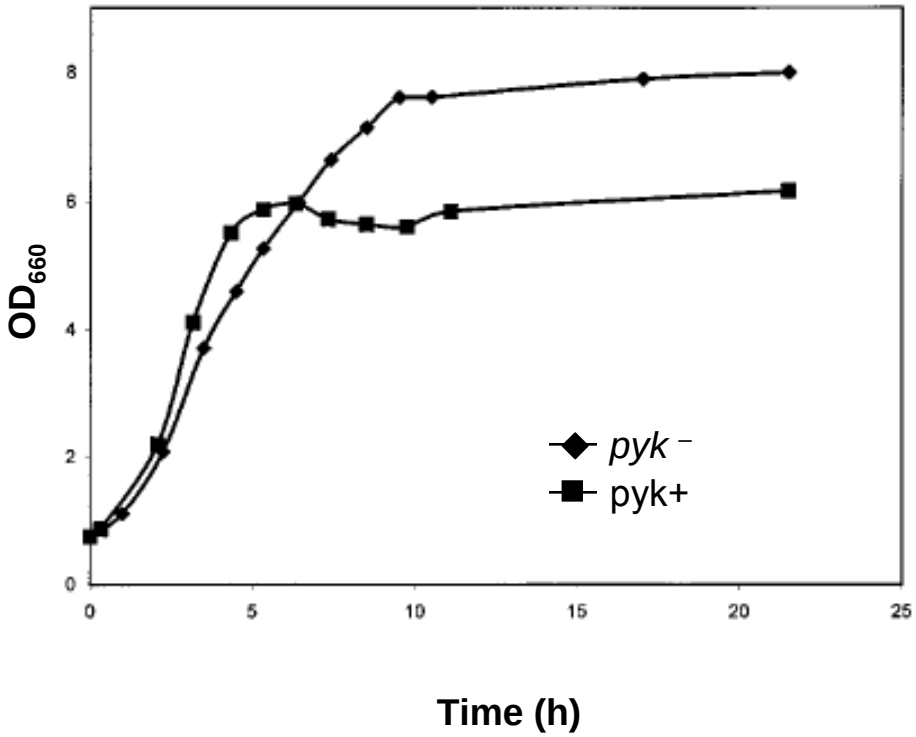
For *E. coli*, Not Local-Low *Pyk* = *Good*



Acetate, Growth, & Cell Density

Medium: M9 + 10 g/L glucose

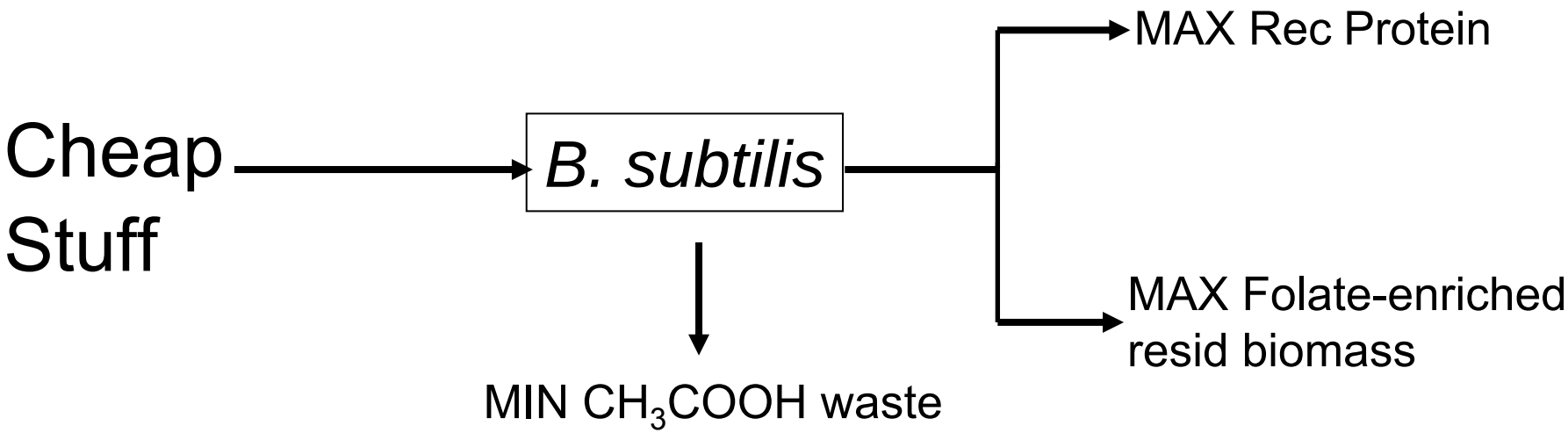




(Maps also to strain variances)

Zhu et al. 2001. *Biotechnol. Prog*. **17**:624-628.

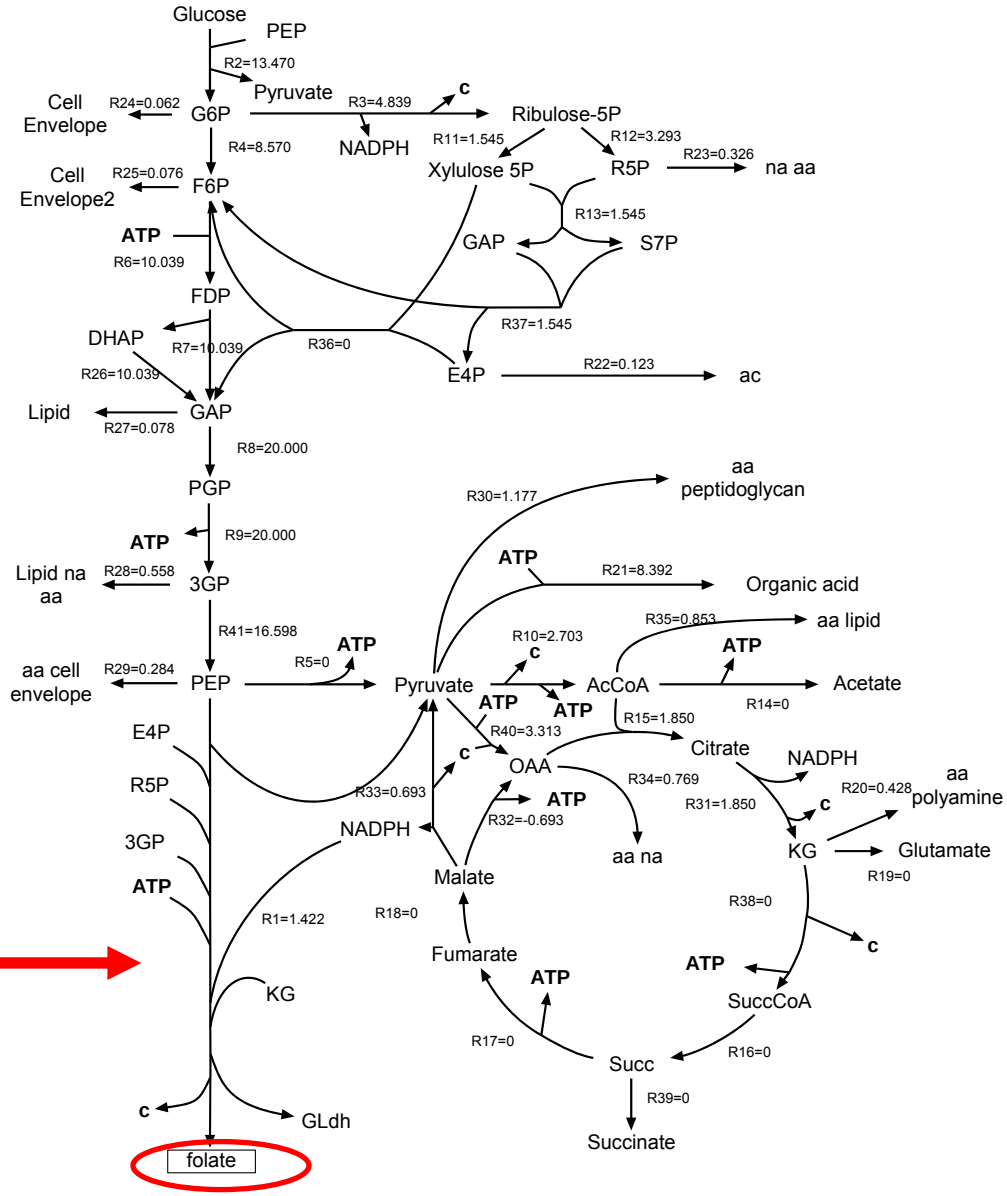
Problem 2



Three screens + growth medium variations; one screen is a lengthy multi-dilution bioassay.

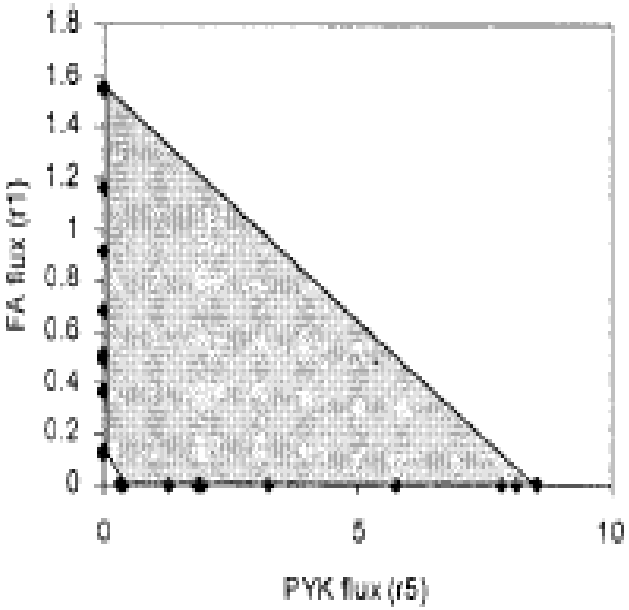
Combinatorics or focus effort?

Ex 2: Attribute that MAX Folate Synthesis?

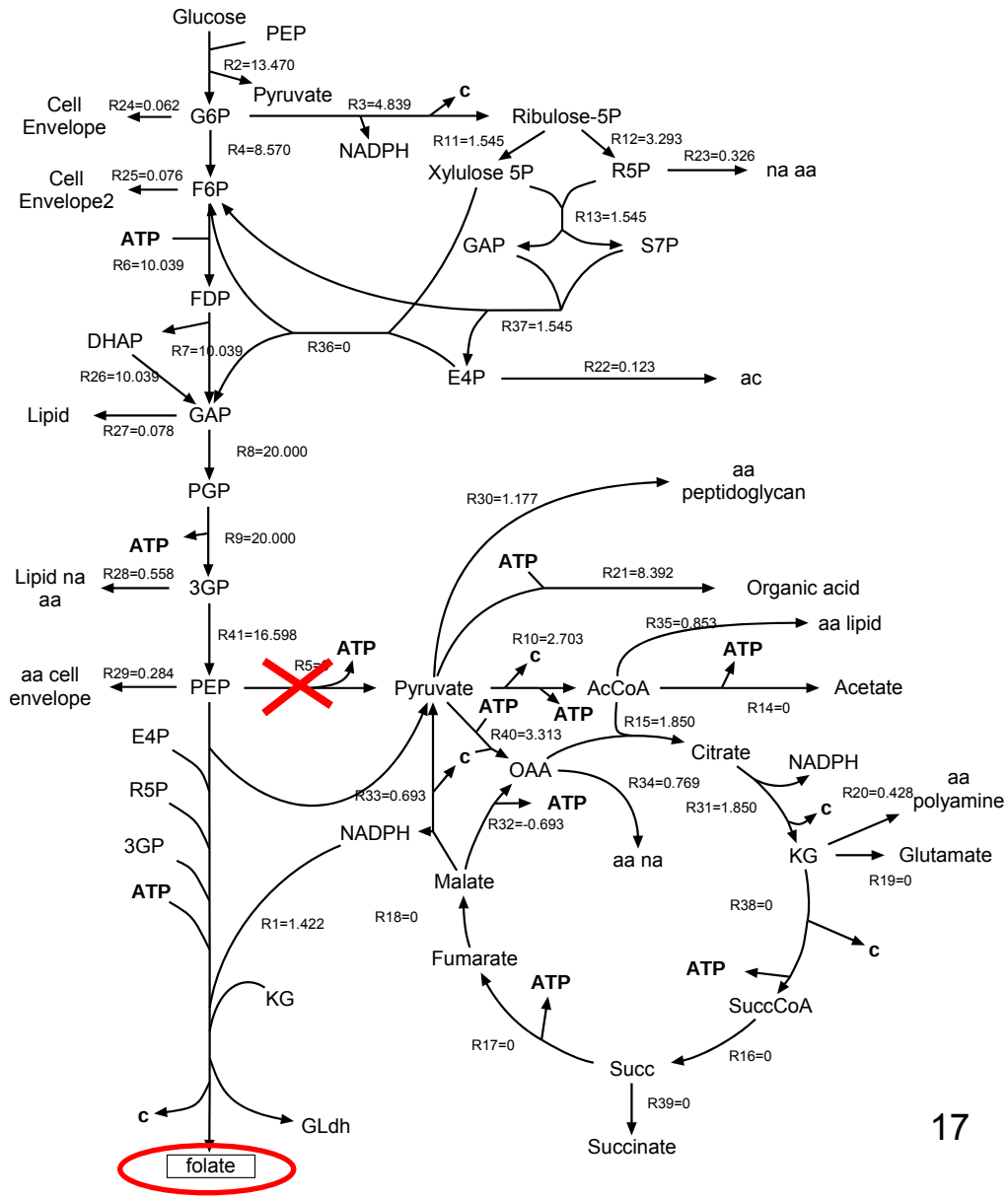


High folate operon expression?

No…Low Pyk = Good



(a)



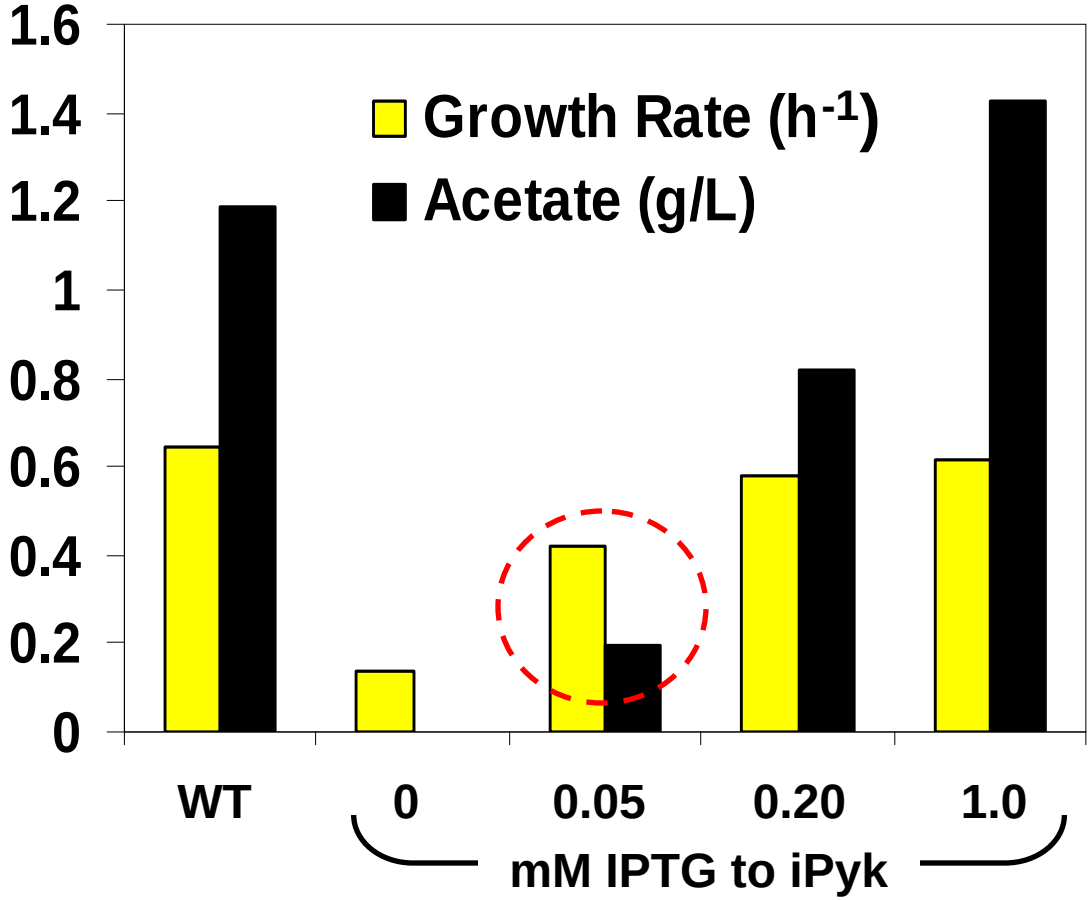
Folate Results *B. subtilis*

Table 3. Folic Acid Yields of Wild-Type & Engineered *B. subtilis* in Glucose Minimum Medium (5g/l).

Strain Name	Description	Doubling Time (hour)	Folic Acid Mass Yield (μg /g cell)	Dry Cell Mass Concentration (g/L)	Volumetric Yield Folic (μg/L)
168	*trp2C*	1.08 ± 0.08	14.7 ± 0.7	1.95 ± 0.05	28.7
1A96	*trp2C, pheA*	1.08 ± 0.08	30.2 ± 2.0	1.98 ± 0.04	59.8
BSZT0410	*trpC2, aroH*^{over}	1.15 ± 0.08	26.8 ± 1.3	1.85 ± 0.05	49.6
BSZT0412	*trpC2, aroH*^{over}*, tktA*^{over}	1.40 ± 0.1	14.3 ± 0.7	1.60 ± 0.05	22.9
BSZT0408	*trpC2, mtrB::erm*	1.20 ± 0.1	26.3 ± 1.5	1.78 ± 0.05	46.8
BSZT0419	*trpC2, mtrB::cm, pyk*^{ind}	1.50 ± 0.1*	56.0 ± 4.0	1.75 ± 0.05	98
BSZT0425	*trpC2, aroH*^{over}*, pyk*^{ind}	1.25 ± 0.1*	60.0 ± 4.0	1.90 ± 0.06	114
BSZT0437	*trpC2, aroH*^{over}*, mtrB::cm, pyk*^{ind}	1.60 ± 0.2*	113 ± 11.0	1.45 ± 0.05	163

Zhu et al, *Appld. Env. Microbiol.* **71**: 7122-7129, 2005.

iPyk or Medium Design for *B. subtilis*



Pan et al. 2006. *Biotechnol. Prog.* **22**:1451-1455.

One key attribute, low pyk, yields all three positive outcomes. Achieve with Ca^{2+}/citrate in medium or engineer strain directly.

Other By-Products to Think About

CO_2 → Bio-Oil + Algal Cell Mass Resid

Bio-Oil → MW tuning for input into HC processes

Bio-Oil → Diesel + Glycerol → Insulin + CO_2

Glycerol is superior to glucose as a substrate.

CO_2 & Algae can be linked to high value things too.

Summary

1. The positive attributes of promising strains may not be due to improved local politics.
2. Load/constrained resource allocation analyses can provide guidance for what defines "good," & results concur

   Nil by-product strains of *E. coli*
   Nil by-product, high folate, high rec prot *B. subtilis*
   High tDNA-yielding *E. coli* strains

3. Results are being linked to fitness/competitive results.
4. Helps to demystify & baseline (e.g. pyk common denom in problem subset).
5. May be additional value-added facets to hν-derived biofuels.

Applications to Algae

- Delineate differences between high & low producers.
- Fitness projection by MOMA or other methods.
- Susceptibility to population successions examined when viewed dynamically with cyclic forcing.
- Map other value-channels for resid.

Special Thanks

Funding provided by:

- National Science Foundation
- Pennsylvania Infrastructure Technology Alliance
- National Institutes of Health

Some Strategies

+

- Mine & Try

  Find a strain that works (Fairly easy)

- Inquiring minds want to know why, & other uses

  Characterize cell internal works to build knowledge (How interpret info?)

- Deliberate Engineering

  Pinch & enlarge capacities in the metabolic network (Really know where to start?).

How Know Loads? Know Paths & Stoich

Biosynthetic Load Profile for *E. coli*

Glycolysis HMP Path Krebs Cycle

Table 3-1. Metabolic cost of monomer synthesis from amphibolic precursors in *Escherichia coli* B/r*.

Monomer	content	G6P	F6P	TP	R5P	E4P	3PG	PEP	PYR	AcCoA	OAA	KG	CO_2
		Precursors used for biosynthesis											
Ala	0.488								0.488				
Arg	0.281											0.281	0.281
Asn	0.229										0.229		
Asp	0.229										0.229		
Cys	0.087						0.087						
Gln	0.250											0.250	
Glu	0.250											0.250	
Gly	0.582						0.582						
His	0.090				0.090								
Ile	0.276								0.276		0.276		-0.276
Leu	0.428								0.856	0.428			-0.856
Lys	0.326								0.326		0.326		-0.326
Met	0.146										0.146		
Phe	0.176					0.176		0.352					-0.176
Pro	0.210											0.210	
Ser	0.205						0.205						
Thr	0.241										0.241		
Trp	0.054				0.054	0.054		0.054					-0.054
Tyr	0.131					0.131		0.262					-0.131
Val	0.402								0.804				-0.402
ATP	0.165				0.165		0.165						0.165
GTP	0.203				0.203		0.203						0.203
CTP	0.126				0.126						0.126		
UTP	0.136				0.136						0.136		
dATP	0.0247				0.0247		0.0247						0.0247
dGTP	0.0254				0.0254		0.0254						0.0254
dCTP	0.0254				0.0254						0.0254		
dTTP	0.0244				0.0244						0.0244		
Lipid A	0.129			0.129			0.129						
Lipid B	0.258									2.1156			
Lipo A	0.0157	0.0157	0.0157										
Lipo B	0.0235	0.0353		0.0235		0.0235	0.0235	0.329					
Pep A	0.0276		0.0552					0.0276	0.0276	0.0552	0.0276	0.0276	
Pep B	0.0552								0.0552				
Glyc	0.154	0.154											
1 C	0.0485						0.0485						
Orn	0.0593											0.0593	
Sum	6.582	0.205	0.0709	0.129	0.897	0.361	1.493	0.7191	2.833	2.928	1.786	1.078	-1.522

* All values obtained from Ingraham, *et al.*, 1983; units are mmole/g dried cells. The costs of precursors for amino acid synthesis have been examined and corrected according to Mandelstam, *et al.*, 1982.

Precursor abbreviations: see Figure 2-1.

Monomer abbreviations: Lipid A, glycerol phosphate and serine in lipid components; Lipid B, average fatty acid ($C_{16:0}$ fatty acid : $C_{16:1}$ fatty acid : $C_{18:1}$ fatty acid = 43 : 33 : 24) in lipid components; Lipo A, uridine-diphosphate-glucose and thymidine-

PYR + OAA → LYSINE

Total mM R5P used in all major pathways to produce amino acids & other things: Load = μ * 0.897 [mmol R5P/g cell h]

Another Key Step: ^{13}C NMR

- Send in a tracer & spectra “output” allows one to “measure’ individual/key rates.
- Spend a lot of $ & supposed to get THE answer.
- Before-answer sets baseline, & after answer supports or refutes strategy.
- If refutes, provides important feedback.

Problems

- Experimental design (labeling of input materials, analyte selection) provides low contrast between feasible/viable outcomes.
- Data → rates numerics quits after finding close or alternate solution, but not the real deal. Compounded by the 1st.

JOURNAL OF BACTERIOLOGY, Jan. 2002, p. 152–164 Vol. 184, No. 1
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.1.152–164.2002
Copyright © 2002, American Society for Microbiology. All Rights Reserved.

Metabolic Flux Responses to Pyruvate Kinase Knockout in *Escherichia coli*

Marcel Emmerling,[1]† Michael Dauner,[1] Aaron Ponti,[1] Jocelyne Fiaux,[2] Michel Hochuli,[2] Thomas Szyperski,[3] Kurt Wüthrich,[2] J. E. Bailey,[1] and Uwe Sauer[1]*

Institute of Biotechnology[1] and Institute of Molecular Biology and Biophysics,[2] ETH Zürich, CH-8093 Zürich, Switzerland, and Department of Chemistry, University at Buffalo, State University of New York, Buffalo, New York 14260[3]

Both looked at same mutant, yet found different metabolic fluxes from ^{13}C NMR data !!

ELSEVIER

FEMS Microbiology Letters 235 (2004) 25–33

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*

Khandaker Al Zaid Siddiquee [a], Marcos J. Arauzo-Bravo [a], Kazuyuki Shimizu [a,b,*]

[a] *Department of Biochemical, Engineering and Science, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan*
[b] *Institute for Advanced Biosciences, Keio University, Tsuroka, Yamagata 997-0017, Japan*

Received 20 January 2004; received in revised form 12 March 2004; accepted 5 April 2004

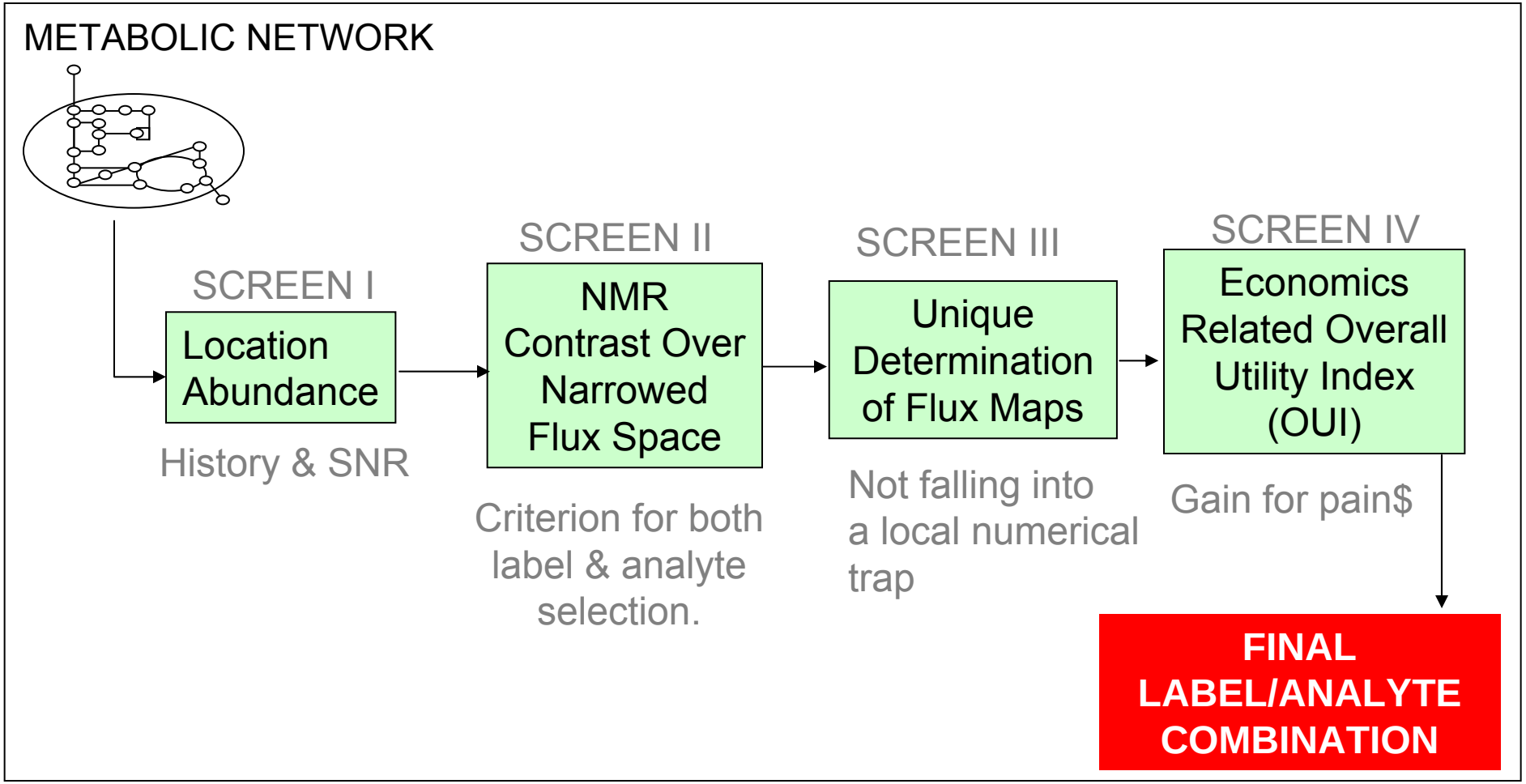
First published online 16 April 2004

General Solution Outline

- Predict alternatives in advance.
- Design ^{13}C expt to distinguish alternatives.
- Rank contrast, & pick cheapest with sufficient contrast.

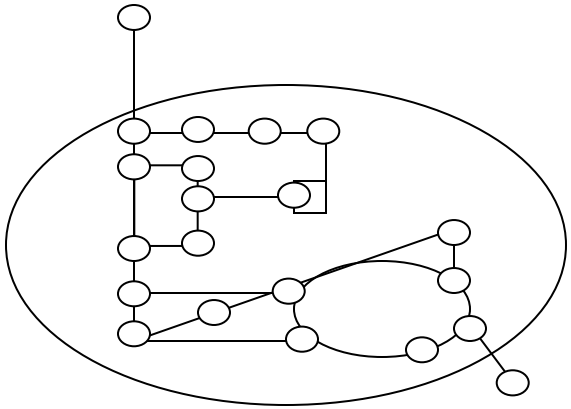
Ghosh et al, “A three-level problem-centric strategy for selecting NMR precursor labeling & analytes,” *Metabolic Engineering*, 2006.

Experiment Design Flowsheet



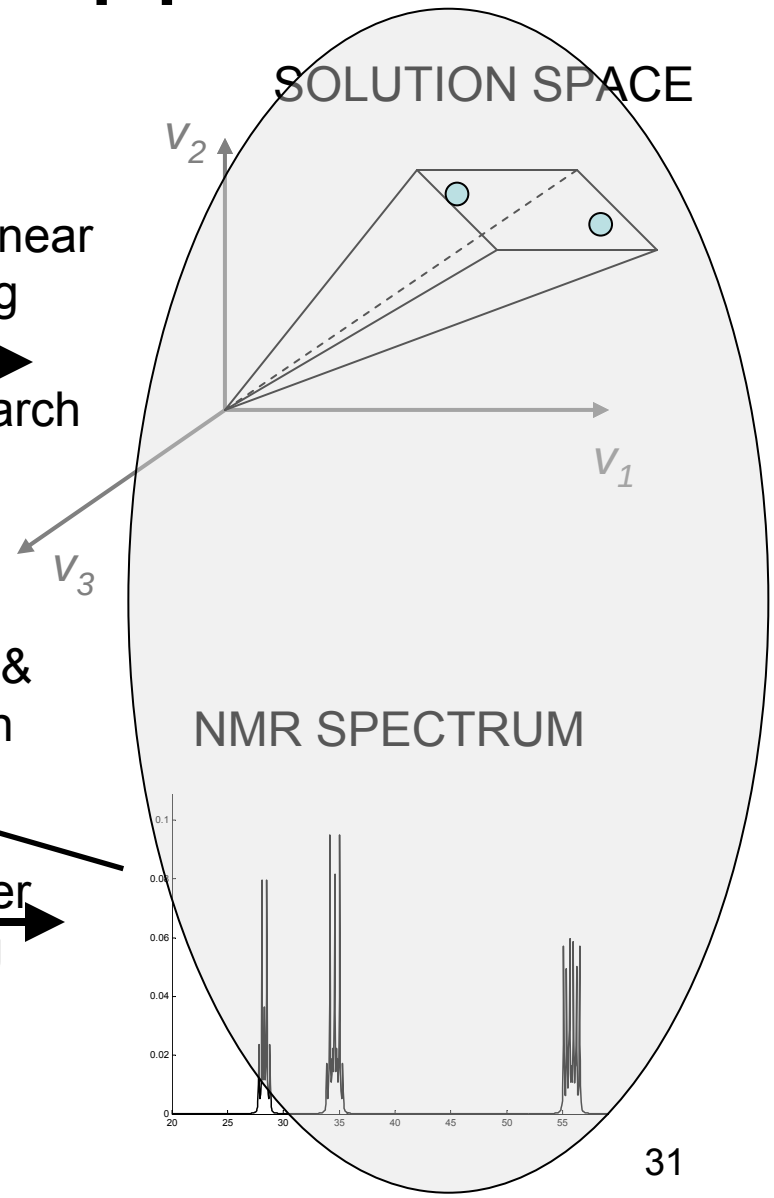
Ghosh, Grossmann, Ataai, & Domach. 2006. A three-level problem-centric strategy for selecting NMR precursors & analytes. *Metabolic Engineering,* **8**: 491-507.

Our 'Integrated' Approach



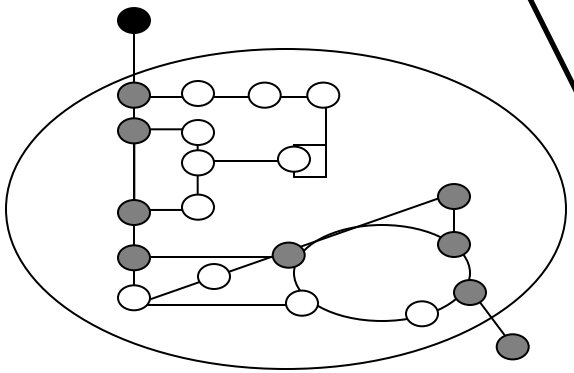
Mixed Integer Linear Programming

Depth First Search



Optimal Label & Analyte Design

$^{13}C_6$ ISOTOPIC LABELING



Isotopomer Mapping

Correct Fluxes

(Ghosh *et al*, 2005, 2006)

Redundancy?

Based on cellular macromolecule composition & known energetics

MIN 1/30 mol ATP → 1 g cells

Many ATP-generating rxns to Σ to fulfill total

18 mmol NADPH → 1 g cell

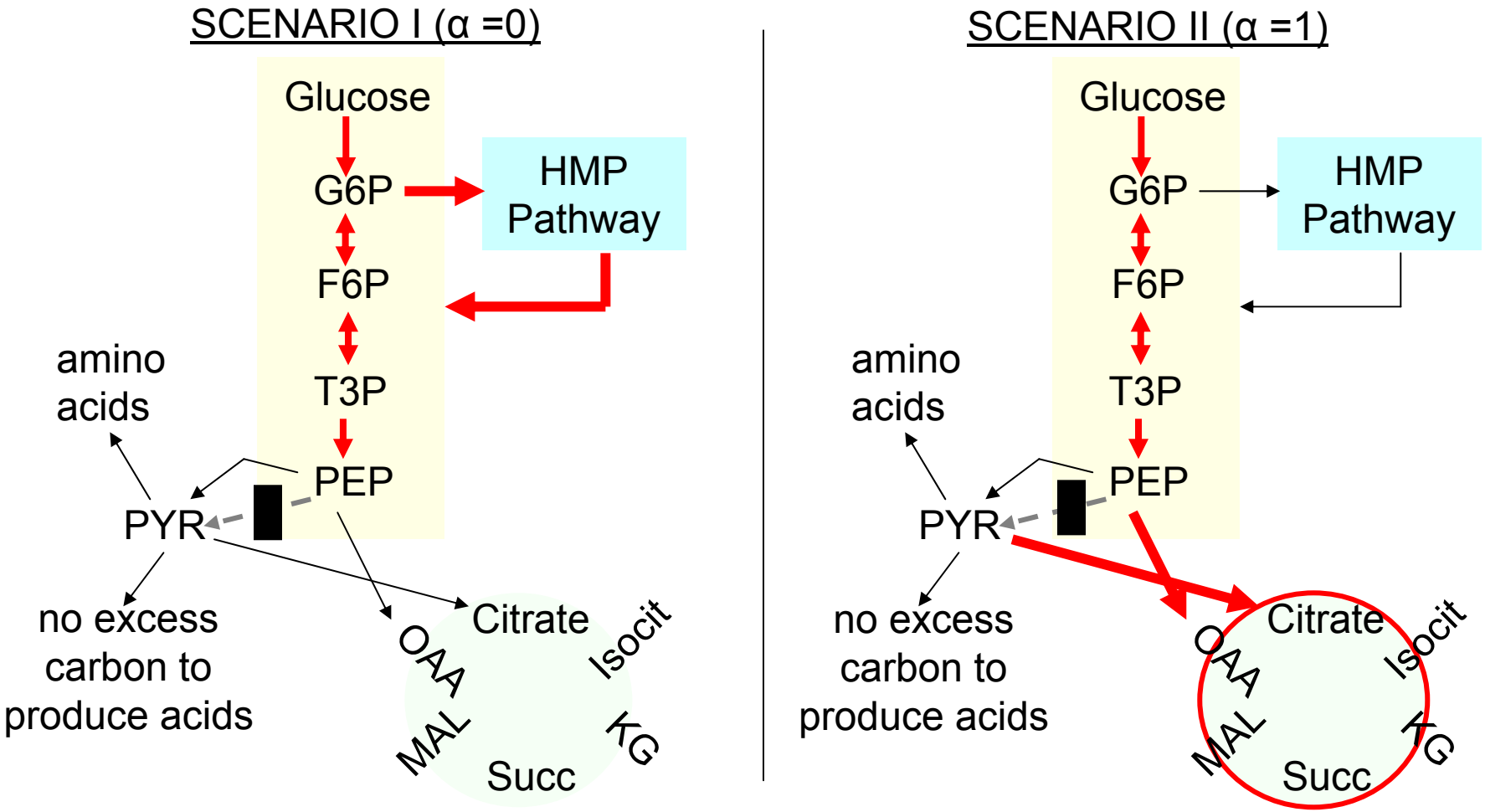
2 or more NADPH dehydrogenases & transhydrogenase to enable reductive chemistry

Many ways to traffic resources to do above & satisfy fixed loads such as

For μ = 0.4 h^{-1}, 0.431 mmol/ g h α-KG from TCA → Glut

Why? Redundancy in Pyk$^-$ *E. coli*

Two Extreme Flux Patterns



Actual α lies between these 2 cases – supposed to be resolved by NMR.

Redundancy Example for *B. subtilis*

Two extreme ways to get same thing done—produce 0 by-products & cometabolize glucose + citrate to same (+/- 10%) overall carbon yield.

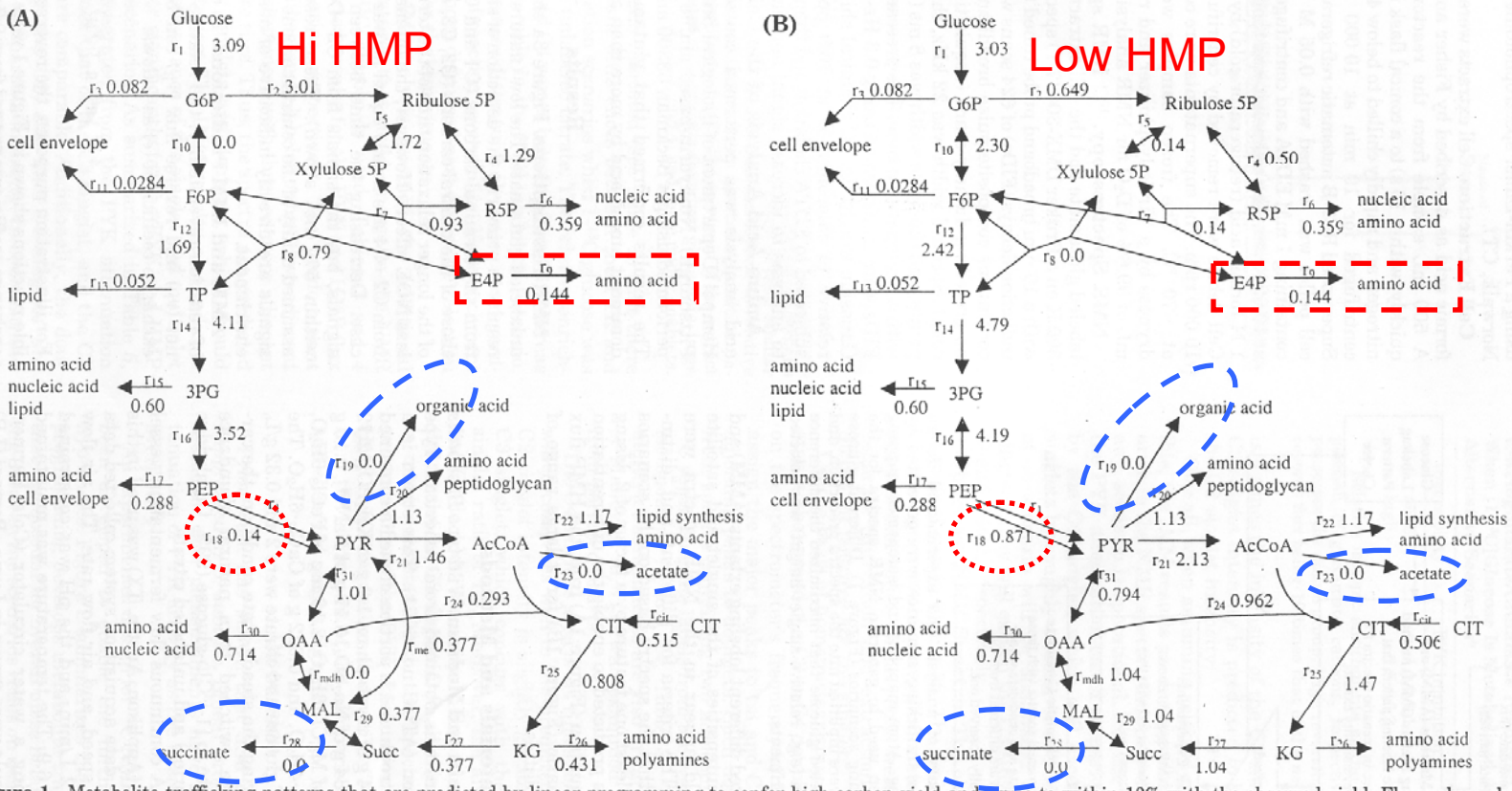


Figure 1. Metabolite trafficking patterns that are predicted by linear programming to confer high carbon yield and agree to within 10% with the observed yield. Flux values shown correspond to mmol/h g cell (basis growth rate is 0.4 h^{-1}). The maximized yield solutions also correspond to minimized (a) pyruvate kinase (r_{18}) or (b) hexose monophosphate pathway (r_2) flux. Potentially reversible steps are shown. Abbreviation: AcCoA, acetyl-CoA; CIT, citrate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; α-KG, α-ketoglutarate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; Succ, succinyl-CoA; TP, triose phosphate; Xu5P, xylulose-5-phosphate; 3PG, 3-phosphoglycerate; MAL, malate; PYK, pyruvate kinase; and PFK, phosphofructose kinase.

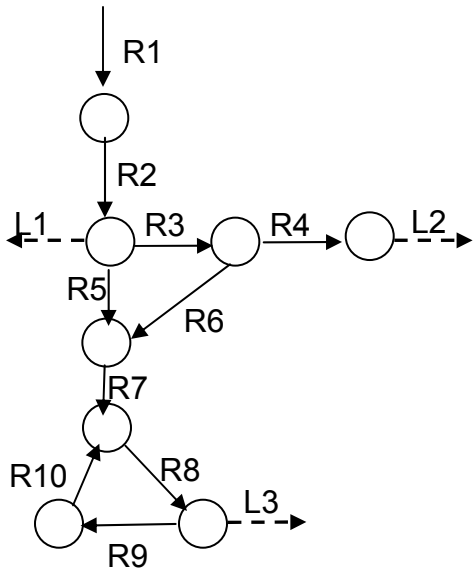
Biotechnol. Prog., 2000, Vol. 16, No. 2

Basic Computation Idea: Simultaneously Optimize While Insuring Viability

Look at cell from Linear Programming standpoint

R = metabolic flux (mmol/g cell h)

L = biosynthetic load on a pathway (e.g. pyr → ala)



Given a growth rate μ

MIN [R1] = resource use (Objective)

Subject to (Constraints):

R1 = R2

R2 = L1 + R3 + R5

R5 + R6 = R7

And other IN = OUT "balances"

ΣATP produced from R_i's > 30.3μ

ΣNADPH produced from R_is =18μ

*L3 = constant * μ (e.g. α-ketoglutarate drained off TCA = 1.078μ)*

*L1 = constant' * μ*

*L2 = constant'' * μ*

Etc

Classic Thinking

ATP production tightly controlled to match use.

Example: Acceptor control mechanism in mitochondria

Precise partitioning of resources between energy production & biosynthesis.

Genes are expressed only when needed.

Unique solutions to problems.

A Model of Efficiency?

Theoretical ATP Yield $Y_{X/ATP}$ = 30 g cell/mol ATP

Reality: E. coli or *B. subtilis* growth on glucose w/O_2

- Observed ATP Yield ~ 10 g cell/mol ATP
- ~40% input carbon lost as acetate

Explanations

- Not fully documented consumptive processes (e.g. so-called maintenance energy).
- Complex system where not simultaneously optimize the integration of everything.

Waste = acetate?

- Inefficiency is actually good.
- Significant redundancy in system exists & somehow related.

Inefficiency Derived from Imbedded Regulation Actually Serves a Purpose?

Greedy Cancer Cells Resemble Many Bacteria Metabolically
From *Business Week Online*
CANCER WATCH
Giving A Sugar Shock To Cancer Cells

The drug, glufosfamide, consists of a traditional chemotherapy treatment, ifosfamide, with a glucose molecule attached. Normal cells take in little of the drug, **but cancer cells, greedy for glucose**, suck in large amounts and die. In Phase 2 studies of glufosfamide used against advanced pancreatic cancer, patients survived a median of 5.6 months, two months longer than the average for late-stage disease. Threshold President George Tidmarsh says 9% of patients in the trial are still alive after two years. Glufosfamide's side effects included nausea, vomiting, and kidney impairment, which are common to chemotherapy.

It may also prove useful for lymphomas and breast and colon cancer, Tidmarsh says. Threshold is recruiting pancreatic-cancer patients for a Phase 3 trial, and results are expected in a year.

Inefficiency Derived from Imbedded Regulation Actually Serves a Purpose?

Cooperation and Competition in the Evolution of ATP-Producing Pathways

Thomas Pfeiffer,[1*] Stefan Schuster,[2] Sebastian Bonhoeffer[1*†]

Heterotrophic organisms generally face a trade-off between rate and yield of adenosine triphosphate (ATP) production. This trade-off may result in an evolutionary dilemma, because cells with a higher rate but lower yield of ATP production may gain a selective advantage when competing for shared energy resources. Using an analysis of model simulations and biochemical observations, we show that ATP production with a low rate and high yield can be viewed as a form of cooperative resource use and may evolve in spatially structured environments. Furthermore, we argue that the high ATP yield of respiration may have facilitated the evolutionary transition from unicellular to undifferentiated multicellular organisms.

20 APRIL 2001 VOL 292 SCIENCE www.sciencemag.org

Editorial

- Most analysts include maintenance energy as a hard constraint to reduce degrees of freedom. Math-driven & concerned about multiplicity.

- If apparent maintenance is, in part, a manifestation of structural conflicts & regulation, & engineering is intended to “fix” the cell, why constrain the possibility space in the design phase?

  ↓

  Worrying about your future mother in-law when only on the first date!

The best *E. coli* can do is ½ theoretical w/o assuming maintenance. Or is the conflict part of maintenance along with imbedded regulation? Zero acetate is possible for Pyk = 0.

B. subtilis has potential closer to theoretical & no conflicts. Imbedded regulation makes it behave suboptimally from *engineering* viewpoint? Zero acetate is possible for Pyk ~ 0.

Why do people use *E. coli* vs *B. subtilis*?

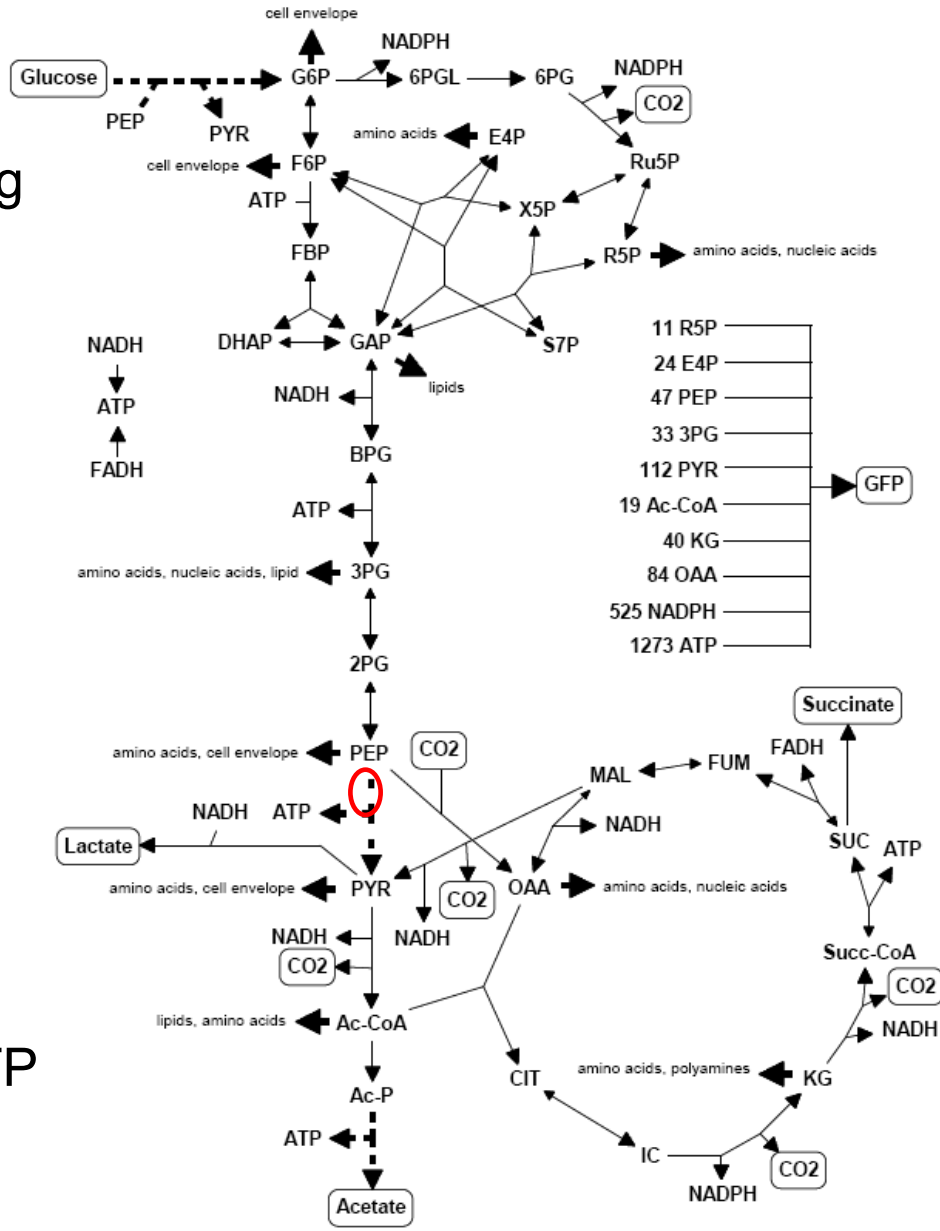
Let optimization guide improvements in also recombinant protein production → →

Model of Recombinant Protein Production

Goal:
MAX r-protein synthesis assuming normal biomass demand is met

Note: aa comp GFP = cell

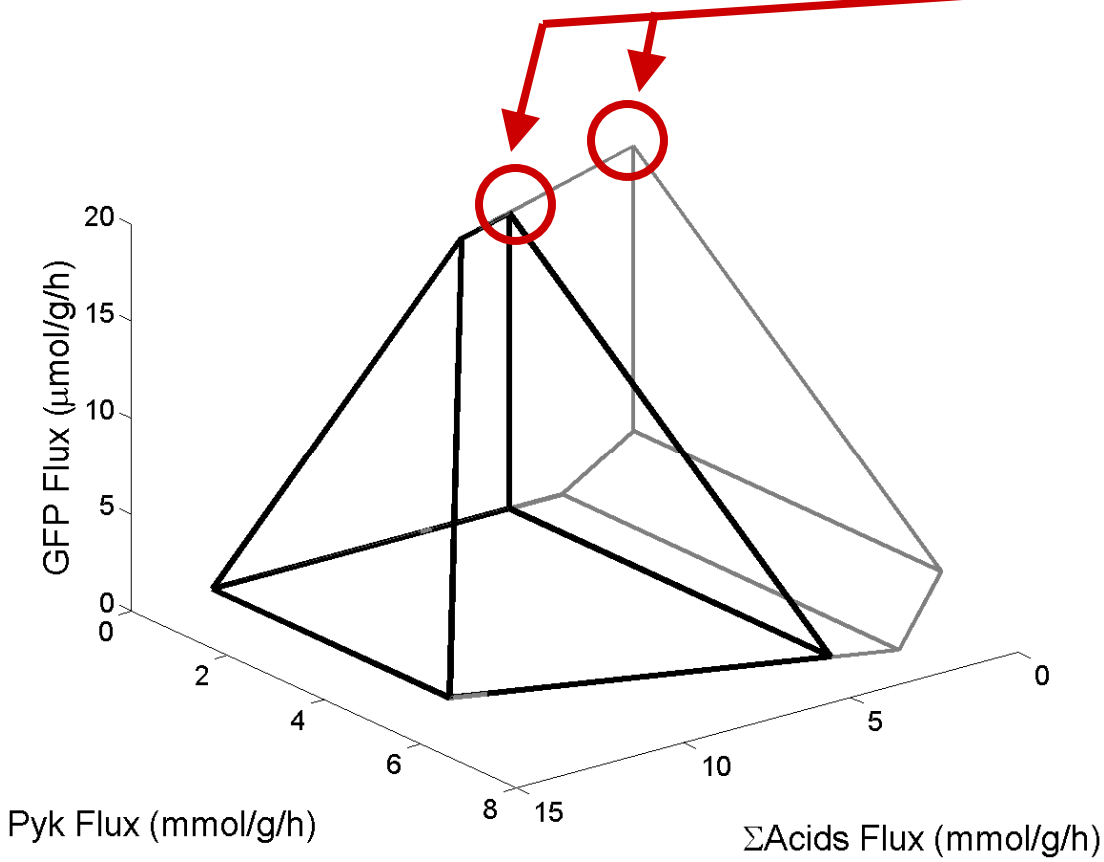
- **Objective**
  - Maximize GFP production
  - s.t. constraints, parameter
- **Solver = METABOLOGICA**
  - Linear optimization via 2-phase simplex method + depth first search
- **System Parameter:** μ
- **Constrained Fluxes**
  - Biomass loads, NADPH, min ATP
- **Metabolite Pools Constrained**
  (steady state => d[ ]/dt = 0)



E. coli Modeling Results Further

Acetate by-production:
constrained / *not* constrained

Max GFP production solution



GFP↑ as acids↓

Intuitive because acids are byproducts that waste carbon. Divert raw materials away from what you don't want to what you do.

More interestingly,

GFP↑ as Pyk↓

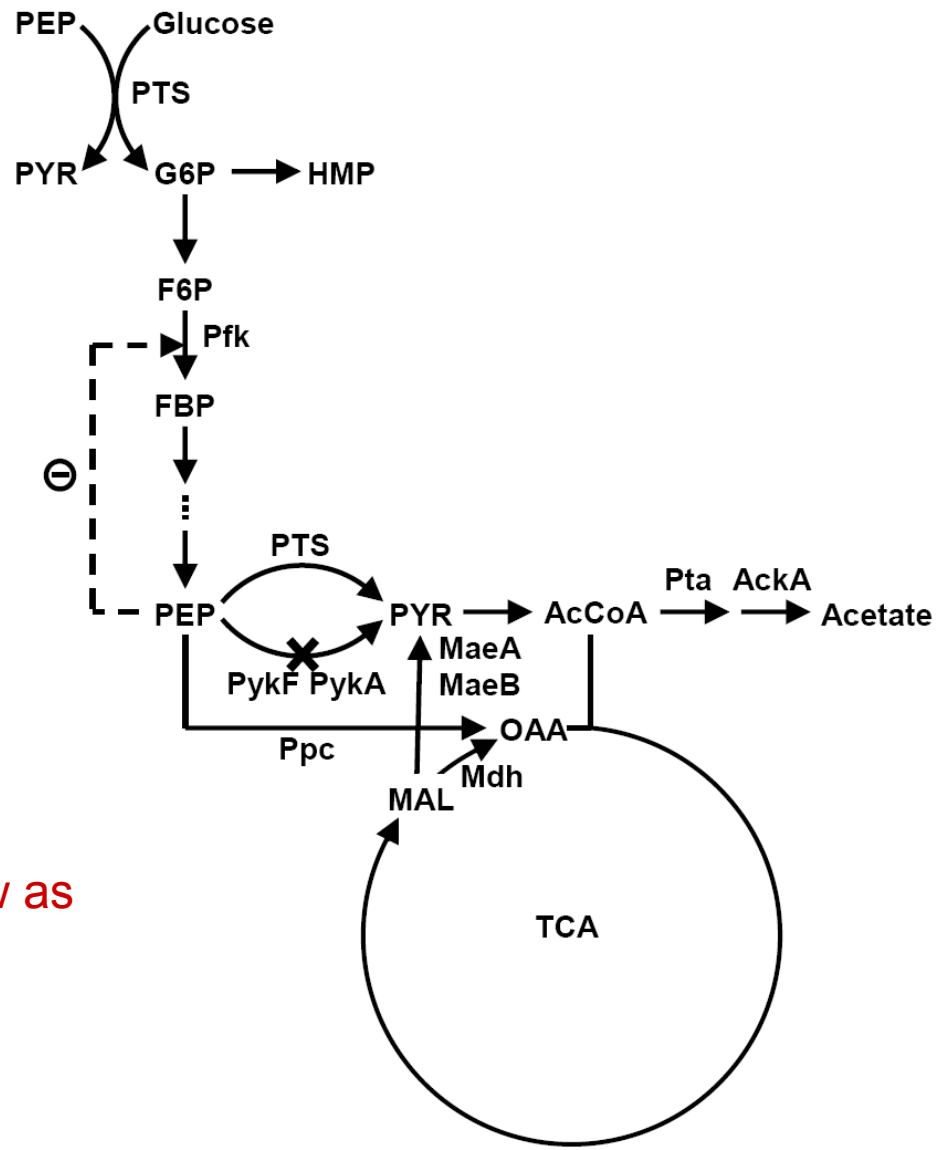
E. coli produces the most recombinant protein when fluxes converge to *pyk*-deficient *E. coli*

Note: Feasible flux space generated from all solutions

Regulatory Check: Pyk$^-$ *E. coli*

- **KO *pykF pykA***
  - Causes PEP accumulation
- **2 Ways to Accommodate ↑ PEP**
  - PEP allosteric inhibition of Pfk
    - ↓ glycolytic carbon use
    - ↑ HMP pathway carbon use
  - "Trickle" effect via **Ppc**
    - PEP provides OAA precursor
    - PTS-derived PYR provides sufficient AcCoA precursor
    - Keeps TCA cycle fed, *but*…

      little leftover carbon to overflow as acetate



Editorial

- Natural regulation helps in this case.
- Find a way to develop MILP & other network models with & without regulation?