Genome-scale Models: Lessons Learned

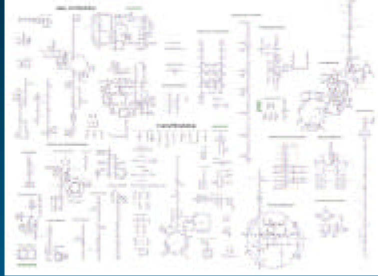
Bernhard Palsson
Lecture #7

September 15, 2003

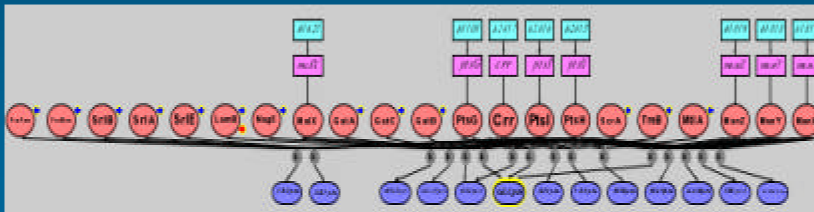
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Outline



1) Status of reconstruction
2) Genome-scale *in silico* models of *E. coli*
3) A Genome-Scale *in silico* Model of *S. cerevisiae*
4) Model applications



The Intracellular Environment is Very Complex:

How do we construct models at a genome-scale?

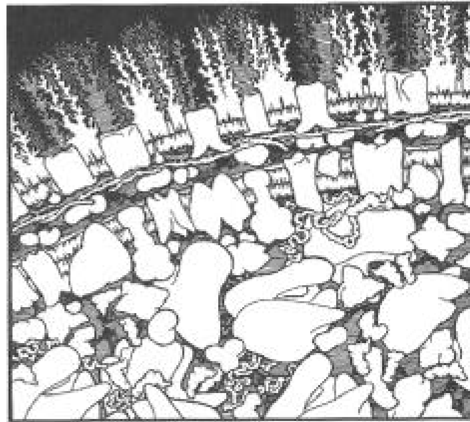


Figure 4.3 Cell Wall

The *Escherichia coli* cell wall, seen in cross section, is composed of several concentric layers. The water outside the cell is at top, in black, and the densely packed cytoplasm is at bottom. The layers of the cell wall from top to bottom (outside to inside) are: the outer membrane, with its glory polysaccharides extending outward and porin pores spanning the bilayer; the thin layer of crosslinked peptidoglycan strands, connected to the outer membrane by small lipoproteins; the periplasmic space, containing a few small proteins; and the complex inner membrane, studded with many different proteins. (1,000,000 ×)

From The Machinery of Life, David S. Goodsell, Springer-Verlag, New York, 1993.

The Intracellular Environment

The above figure depicts the crowded environment that results from having 30% biomass. The depiction of the intracellular environment shown here is very helpful when trying to formulate a strategy on how to go about mathematically modeling and simulating this very complex environment.

What is in a reconstruction?

Genome:
- Annotated genes
- Gene location
- Regulatory regions
- Wobble base pairs

Biochemistry:
- Stereochemistry
- pH and pKa (charge)
- Elemental balance
- Charge balance
- Multiple reactions/enzyme
- Multiple enzymes/reaction

Transcription/translation:
- Gene to transcript to protein to reaction association
- Transcript half-lives
- tRNA abundances
- Ribosomal capacities

Physiology:
- Flux data
- Knock-outs
- Balanced functions
- Overall phenotypic behavior
- Location of gene product compartmentalization

WHAT IS IN A RECONSTRUCTION?

There are numerous sources we use in a metabolic reconstruction. The first type of data we use relates to an organism's genome. For example, genome annotations provide useful insight into potential functional assignments of open reading frames. Gene location may be provide insight into the way a gene is expressed. It is also important for understanding the regulatory structure that controls this expression. By studying the known effects of repressors and activators, we can implement rules to describe a gene's regulation. Our lab is also interested in wobble base pairs, which may lead to clues about an organism's protein composition.

Another source of information is biochemical data. Stereochemistry, pH, and pKa are all important in determining the exact state of compounds that are involved in the reactions within the organism. We also must consider the elemental and charge balance of these reactions, which are known to hold true in vivo. Finally, we are also concerned with the association between enzymes and reactions, such as identifying isozymes and protein complexes.

We also consider transcription and translation in a metabolic reconstruction. It is important to understand the associations between genes, proteins, and reactions in order to properly characterize these processes. The half-lives of transcripts, tRNA abundancies, and ribosomal capacities all impact an organism's protein composition.

Finally, physiological data is required to understand the network as a whole. Flux data and knock-outs are useful for testing our models in their prediction of phenotypic behaviors. We can also use physiological data for the model development, such as identifying the compartmental location of a particular gene product.

Status of Genome-Scale Models

Reconstruction of metabolic models is routine

- Have several predictive bacterial models
- First eukaryotic model built
- First draft to human model emerging

Organism	Year	Genes	Reactions	Metabolites
Escherichia coli (core)	1990	24	14	17
Escherichia coli	1993	250	146	118
Escherichia coli	1998	306	317	305
Haemophilus influenzae	1999	362	488	343
Escherichia coli (v1.0)	2000	695	720	436
Helicobacter pylori	2002	291	388	339
Escherichia coli (regulated)	2002	149	113	63
Saccharomyces cerevisiae	2002	957	1294	801
Escherichia coli (v1.1)	2003	906	1043	594

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Gathering biological data and using it to **reconstruct metabolic models is routine**. We have build several predictive bacterial models such as E. coli, H. influenzae, and H. pylori [point out in table] and our first eukaryotic model, S. cerevisiae [point out in table].

We have already laid the groundwork for reconstructing **regulatory networks** and incorporating them into our metabolic models. In fact, we have begun to test a regulated E coli network and have started to implement regulatory rules in yeast.

Finally, we have just begun to reconstruct signaling networks, such as the JAK-STAT pathway in yeast.

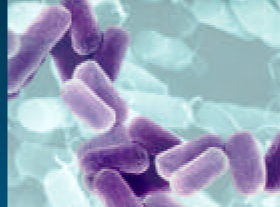
Genome-scale *in silico* **models of** *Escherichia coli*

University of California, San Diego
Department of Bioengineering

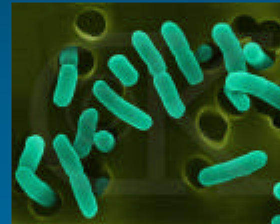
Systems Biology Research Group
http://systemsbiology.ucsd.edu

E. Coli at a Glance

- *E. coli* is a Gram-negative bacterium
- It is a prokaryote and belongs to the family *Enterobacteria*.
- It is a well-studied organism in terms of its metabolism and regulation.
- There are currently 3 DNA sequenced strains of *E. coli*:
  - K12: MG1655
  - O157:H7 EDL933
  - O157:H7 VT2-Sakai



http://www.chromosome.com/bacteria_wallpaper.html

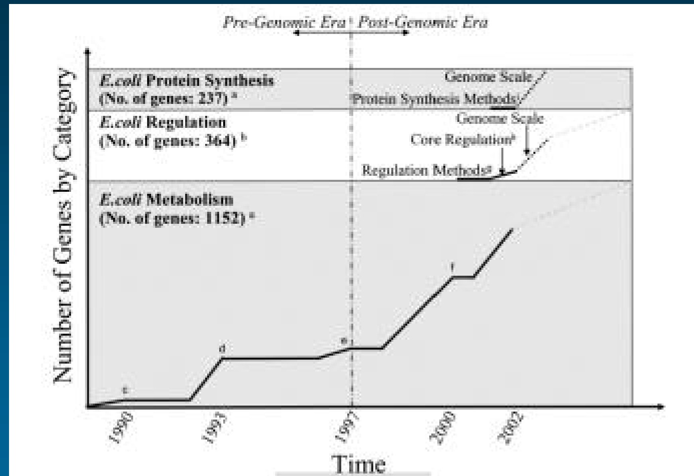


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

E. coli is a Gram-negative bacteria. It is a prokaryote belonging to the family Enterobacteria. *E. coli* has been studied physiologically and biochemically for a long time. Its metabolism and regulation are more characterized than other bacteria. There are many different strains of *E. coli*. Three strains have been fully sequenced and annotated, K12:MG1655 in 1997, and O157:H7 in 2001.

Development of the E. coli Model

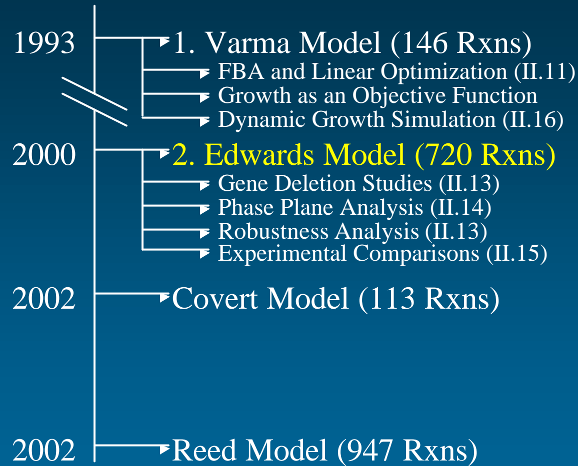


"Thirteen years of constraint-based model building of E. coli" J Bact, May 2003

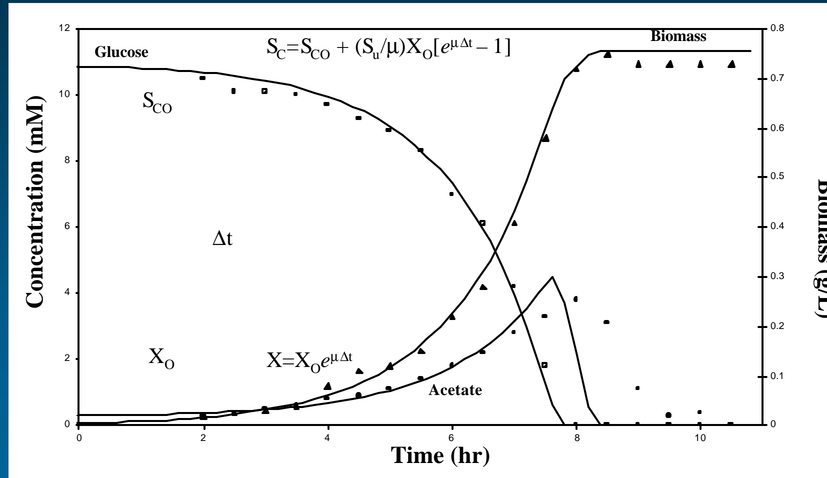
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

The development of successive constraint-based FBA models of E. coli. Constraint-based models of E. coli first focused on metabolism. By the time the complete genome was sequenced (1997), only 26% of metabolic genes were accounted for in the FBA models. Over the next 5 years the number grew to include nearly 80% of the metabolic genes. Methods for incorporating transcriptional regulation have been developed and implemented in the core metabolic model of E. coli, as have methods for including protein synthesis. Expanding the regulatory and protein synthesis models to the genome scale can be accomplished by using the information that is known today (indicated in dotted lines). Further gene annotations should increase the size of models (dashed lines). These three components can be combined to form an integrated model that accounts for nearly 2,000 genes.



The first genome-scale model of *E. coli* was reconstructed in 2000 from genomic, biochemical, and physiological data. The model contained 720 reactions, associated with metabolism and transport. Again most of the results from this model have already been presented throughout the course. Gene deletion studies were performed; the model was found to correctly predict the *in vivo* results in 86% of the cases studied. The model has been extensively used in calculating phase planes for various carbon sources. Experiments have been performed validating the predictive capabilities of the model, see topic II.15. An example will be shown on the next slide.



Shown here is a time course for an aerobic batch culture of E. coli grown on glucose minimal media. Glucose is being consumed to make biomass and acetate is being secreted as a by-product. When glucose is depleted the acetate is take up and utilized. The curves represent the model predictions based on the equations shown. Experimental data points are also shown. There is a good agreement between the model and experimental data. The model does not predict the time lag associated with reabsorption of acetate since gene expression and protein synthesis are not incorporated into the model.

Phase Plane Analysis

- Calculated the following Phase Planes: Glucose, AKG, Glycerol, Acetate, Pyruvate, Malate, Succinate, and Lactate.
- Results: Glucose, Pyruvate, Succinate, Malate, Acetate, Glycerol* and Lactate phase planes (2D and 3D) appear to be the same. Biomass yields do not change. AKG was the only one that was different.

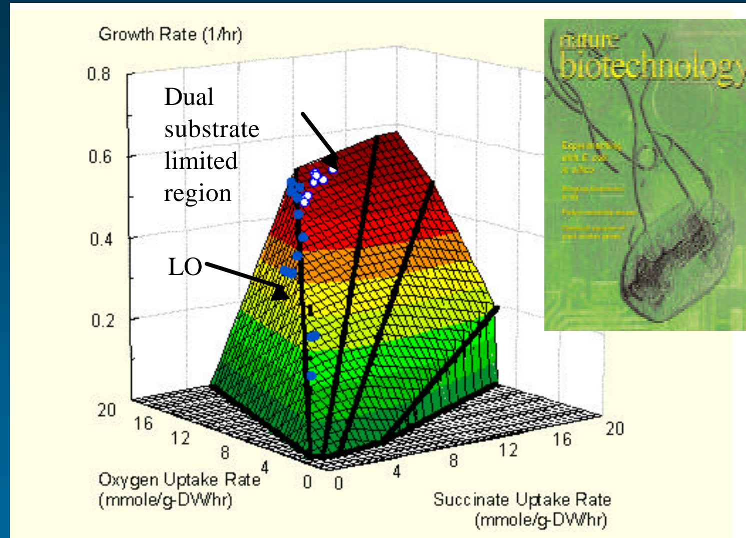
*Glycerol was the same only after deletion of katG and katE genes

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

A comparison of the phenotypic phase planes between models yielded minor differences. Originally the new phase plane for glycerol was different, but after looking at calculated flux distributions it was decided that *katG* and *katE* (that were added to the old model) , which encode hydroperoxidase (2 H_2O_2 ? O_2 +H_2O) should be removed. The model predicted an high flux through some of the vitamin B6 pathways creating a cycle that produced H_2O_2. With *katG* and *katE* present, the model predicted that this cycle would operate along with the hydroperoxidase enzyme to generate or consume O_2. The final version of the new model (v1.1) didn't have the *katG* and *katE* enzymes.

After removal of the hydroperoxidase from the new model, the phase planes for glucose, pyruvate, succinate, malate, acetate, glycerol and lactate the phase planes were identical to those generated by v1.01. a-ketoglutarate (AKG) was the only carbon source tested whose phase plane changed.



This figure, shows the Succinate-oxygen PhPP in three dimensions.

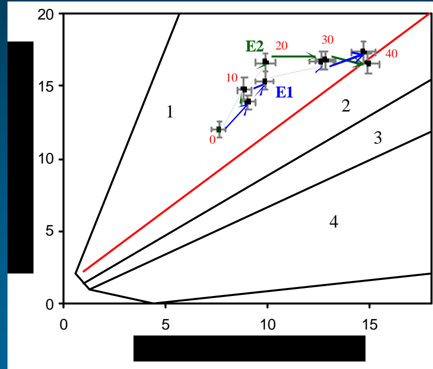
•The formalism is similar to the 3-D acetate PhPP

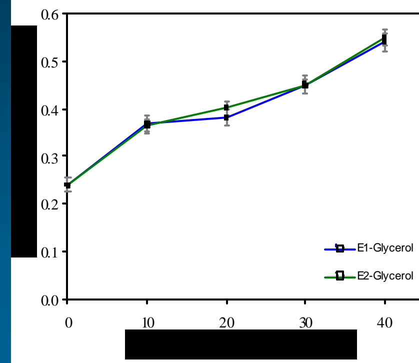
•Here the effect of the carbon source on the structure of the PhPP can be seen.

•The LO is shown here, and the data points with reduced succinate uptake rates all lie on (or near) the LO,

•However, when the succinate uptake rate was increased, the experimental data followed the LO until the oxygen mass transfer constraint was reached. At this point, the growth rate and the succinate uptake were increased by moving into region 2 of the phase plane.

E. coli does evolve towards stoichiometric optimality



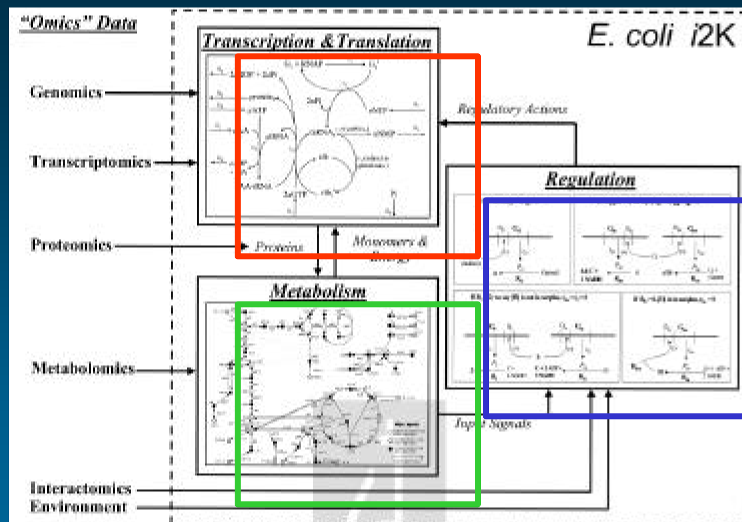


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

When *E. coli* was grown on glycerol, the cells initially exhibited sub-optimal growth behavior. They operated in region 1 of the phase plane where futile cycles are being used. Serial batch cultures were performed and the cells evolved towards the line of optimality that was predicted *a priori in silico*.

Integrated Model of E. coli

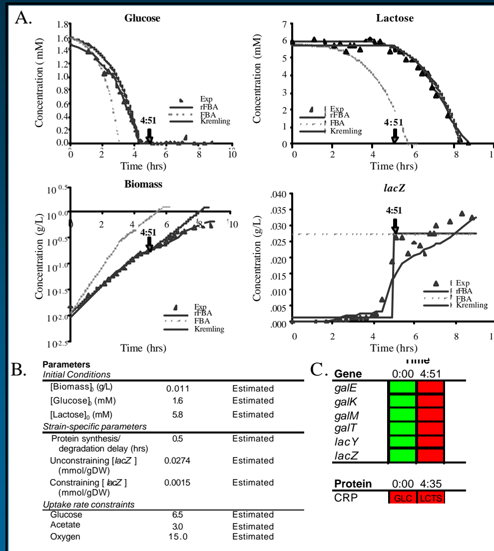


"Thirteen years of constraint-based model building of E. coli" J Bact, May 2003

Integrated constraint-based model of E. coli: the E. coli i2K model. Constraint-based modeling frameworks have been developed for metabolism, regulation, transcription, and translation. The connectivity among the three modeling components is shown here. Integration of these three modeling components should produce an integrated model of E. coli that accounts for nearly 2,000 genes, referred to as the E. coli i2K model. This model can be used to reconcile diverse "-omics" data and utilize the data to more accurately predict a cellular phenotype.

Combined Regulatory/Metabolic Modeling

- Physiological time courses (growth, uptake/secretion)
- Microarray simulation
- Effects of gene deletions on cellular behavior
  - More genes may be evaluated
  - More accurate overall



University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

COMBINED REGULATORY AND METABOLIC MODELING

This slide shows what kind of calculations are possible using the regulated flux balance approach and the regulated *E. coli* metabolic network in a simulation of the glucose-lactose diauxic shift. Using this approach it is possible to generate time courses of growth as well as glucose and lactose uptake. It is also possible to infer concentrations of proteins and even to simulate, qualitatively, gene expression data. We can also simulate the effects of gene deletions on cellular behavior with more accuracy and broader scope.

A Genome-Scale *in silico* **Model of** *Saccharomyces cerevisiae*

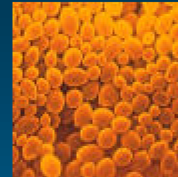
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

S. cerevisiae at a Glance

Genome Characteristic

Fully sequenced in 1996
12,052 Kbp genome length
6,259 ORF, ~4,300 functionally known genes



Applications

Medicine: production of insulin
Industry: proteins for feedstock enrichment, glycerol, food additives, etc.
Domestics: winery, brewery, and bakery

The Model Eukaryotic Organism

Well studied genetic, metabolic, and regulatory biology
Easy to grow and manipulate
Basis for studying more complex eukaryotes

University of California, San Diego
Department of Bioengineering

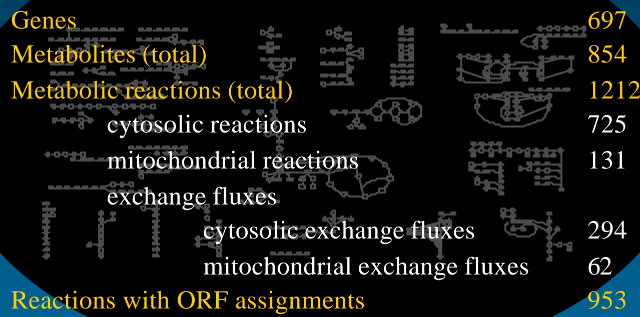
Systems Biology Research Group
http://systemsbiology.ucsd.edu

The genome of Saccharomyces cerevisiae was fully sequenced, annotated, and became available to the public in 1996. The length of the genome is about 12 mega base-pair and it contains about 6300 open reading frames of which about 4300 have known annotation based on genetics, biochemistry, or homology.

S. cerevisiae has numerous applications in medicine, for instance in production of pharmacological proteins such as insulin; in industry in mass-production of various proteins and food additives; and it has a long history in domestics in production of alcoholic beverages and bakery, which make it the oldest microorganism studied by humans.

S. cerevisiae has been chosen by the scientific community as the model eukaryotic organism for several reasons. Its genetic, metabolic, and regulatory biology is well studied and well understood. Its relatively simplistic biology and the easiness in manipulating its genetic composition makes it attractive to work with and convenient to grow in laboratories. Studying and understanding the biology and physiological behavior of this organism can therefore be the basis for understanding more complex eukaryotic organisms such as human cells.

Network Characteristics



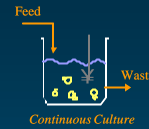
***S. Cerevisiae* Metabolic Network**	
Genes	697
Metabolites (total)	854
Metabolic reactions (total)	1212
cytosolic reactions	725
mitochondrial reactions	131
exchange fluxes	
cytosolic exchange fluxes	294
mitochondrial exchange fluxes	62
Reactions with ORF assignments	953

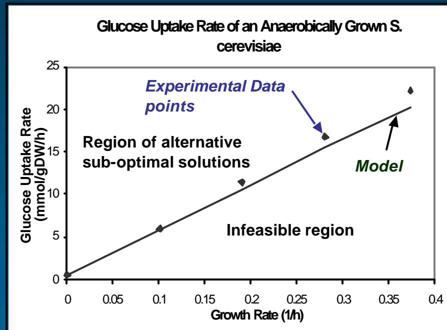
University of California, San Diego
Department of Bioengineering

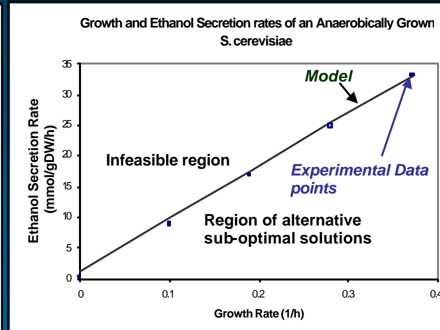
Systems Biology Research Group
http://systemsbiology.ucsd.edu

The yeast *in silico* model is composed of 697 genes, 868 metabolites, and 1212 reactions, of which 131 occur in mitochondria, 725 in cytosol, and 356 are exchange fluxes across the mitochondrial and cytocolic membrane. The genes included in the model constitute about 16% of the known genes in the genome and 58.7% of the known metabolic genes.

Anaerobic Glucose-Limited Continuous Culture

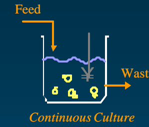


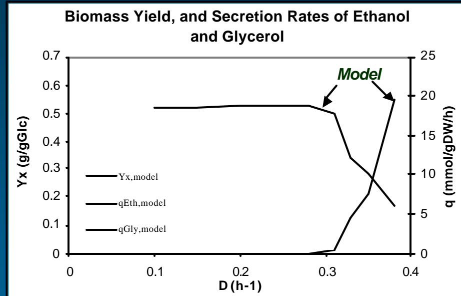


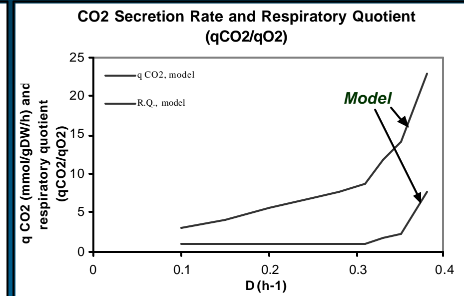


Experimental data are adopted from: Nissen, T.L., et al., Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. Microbiology, 1997. 143(Pt 1): p. 203-18

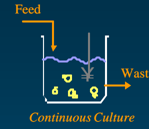
Aerobic Glucose-Limited Continuous Culture

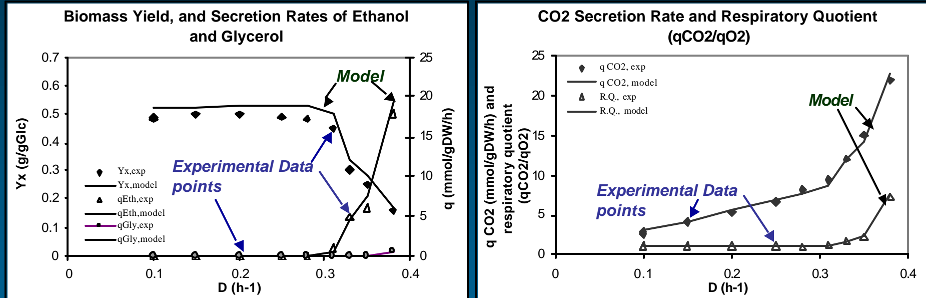






Aerobic Glucose-Limited Continuous Culture





Experimental data are adopted from: Overkamp, K.M., et. al., In vivo analysis of the mechanisms for oxidation of cytosolic NADH by Saccharomyces cerevisiae mitochondria. Journal of Bacteriology 2000 May, 182(10):2823-30.

Gene Deletion Study

Central Metabolism:

81.5% agreement in knockout viability (93 out of 114 cases)

Genome-Wide Gene Deletion Studies:

85.6% agreement in knockout viability (499 out of 583 cases)

Gene name	defined complete Glc (*in silico*/*in vivo*)	defined minimal Glc (*in silico*/*in vivo*)	defined minimal Ace (*in silico*/ *in vivo*)	defined minimal Eth (in silico/ in vivo)	References: (for minimal media)
ACO1	(+/+)	(-/-)			Gangloff (1990)
CDC19#	(+/-)	(+/-)			Boles (1998)
CIT1	(+/+)	(+/+)			Kim (1986)
CIT2	(+/+)	(+/+)			Kim (1986)
CIT3	(+/+)				
DAL7	(+/+)	(+/+)	(+/+)	(+/+)	Hartig (1992)
ENO1	(+/+)				
ENO2[$$]	(+/-)	(+/-)			
FBA1*	(+/-)	(+/-)			
FBP1	(+/+)	(+/+)		(-/-)	Sedivy (1985), Gancedo (1984)
FUM1	(+/+)				
GLK1	(+/+)				
GND1[##]	(+/-)	(+/-)			
GND2	(+/+)				
GPM1[¦]	(+/-)	(+/-)			
GPM2	(+/+)				
GPM3	(+/+)				
HXK1	(+/+)				
HXK2	(+/+)				
ICL1	(+/+)	(+/+)			Smith (1996)
IDH1	(+/+)	(+/+)			Cupp (1992)
IDH2	(+/+)	(+/+)			Cupp (1992)
IDP1	(+/+)	(+/+)			Loftus (1994)
IDP2	(+/+)	(+/+)			Loftus (1994)
IDP3	(+/+)				
KGD1	(+/+)	(+/+)			Repetto (1991)
KGD2	(+/+)	(+/+)			Repetto (1991)
LPD1	(+/+)				
LSC1	(+/+)		(+/+)	(+/+)	Przybyla (1998)
LSC2	(+/+)		(+/+)	(+/+)	Przybyla (1998)
MAE1	(+/+)	(+/+)		(+/+)	Boles (1998)

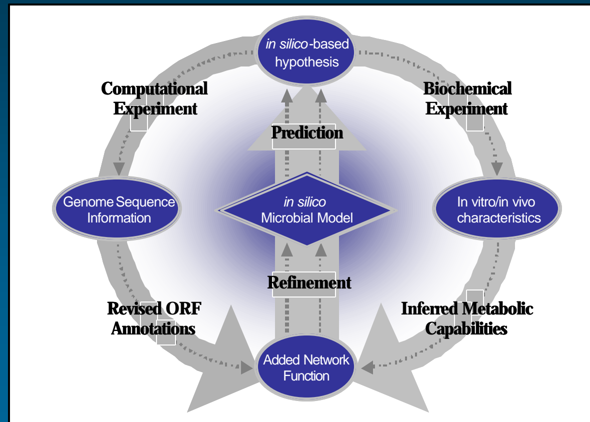
(+/-) (growth/ no growth)

Model Applications

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Two Facets: Computational and Experimental



University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

We now elaborate that this iterative model building can occur via two routes: experimental and computational. Here is a diagram of a model-centric discovery process. The model is used to make predictions which are basically *in silico*-based hypotheses which must be tested. If the predictions are tested experimentally, the result is the determination of new metabolic capabilities; if the predictions are tested computationally, the result is a revised ORF annotation. In either case, network functions are added and the model is improved, leading to new, more accurate predictions, and so on.

We can illustrate both arms of this process using our experience from the *H. pylori* model.

Computational: H. Pylori Sequence Analysis (BLAST)

Enzymes included in the *in silico H. pylori* strain without direct evidence, with locus numbers of ORFs with significant similarity to genes encoding these enzymes in other organisms.

HP Locus	Organism	Gene Product Name	Similarity	Identity
HP0086	*Corynebacterium glutamicum*	Malate dehydrogenase	36.81%	25.93%
HP0104	*Escherichia coli*	5'Nucleotidase	36.71%	25.76%
HP0133	*Escherichia coli*	Threonine transporter	50.00%	33.33%
HP0192	*Synechocystis sp.*	Aspartate oxidase	42.08%	30.94%
HP0328	*Francisella novicida*	Tetraacyldisaccharide 4'kinase	42.34%	29.20%
HP0474	*Synechococcus sp. (strain PCC 7942)*	Sulfate transporter	38.81%	26.48%
HP0561	*Leishmania tarentolae*	Dihydrofolate reductase	39.59%	30.20%
HP0618	*Sus scrofa (Pig)*	Cytidylate kinase	41.40%	30.65%
HP0672	*Schizosaccharomyces pombe*	Alanine transaminase	35.54%	25.73%
HP0723	*Pseudomonas sp. (strain 7A)*	Glutaminase	54.57%	44.51%
HP0940	*Campylobacter jejuni*	Histidine transporter	40.41%	29.80%
HP0976	*Escherichia coli*	Ornithine transaminase	39.17%	27.74%
HP1017	*Salmonella typhimurium*	asparagine transport protein	43.86%	32.63%
HP1017	*Escherichia coli*	Lysine transporter/permease	49.25%	37.10%
HP1017	*Escherichia coli*	Phenylalanine transporter	44.20%	30.64%
HP1017	*Saccharomyces cerevisiae*	Tryptophan transporter	40.68%	31.94%
HP1045	*Staphylococcus aureus*	O-Succinylbenzoate-CoA ligase	33.95%	23.66%
HP1232	*Pneumocystis carinii*	dihydroneopterin aldolase	41.02%	28.15%
HP1282	*Pseudomonas aeruginosa*	Isochorismate synthase 1	32.58%	21.80%

- Many of the above HP loci have already been assigned functions which are similar to those included in the *in silico* strain (*rocE*, transaminase).
- Construction of *in silico* strains can direct bioinformaticists and experimentalists in more complete genome annotation (malate dehydrogenase)

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

First, computationally. For each of the reaction predictions which were low-confidence, based on either physiology alone or inference, we assumed that the genes might be found in the *H. pylori* genome. Accordingly, we searched for these hypothetical genes using a BLAST search and found that a potential locus could be found which had significant similarity to corresponding genes in other organisms. Some of the results are shown here. These results represent a possibility which can be explored further.

The first thing that we noticed is that many of the above HP loci have already been assigned functions which are similar to those included in the in silico strain. For example, many amino acid transport proteins showed similarity to the *rocE* gene on HP1017, which encodes an "amino acid permease". More research could indicate which amino acids in fact use this permease.

Another important finding has to do with malate dehydrogenase and will be discussed on the next slide.

Network Reconstruction as a Predictive Science: Malate dehydrogenase

Enzymes included in the *in silico H. pylori* strain without direct evidence, with locus numbers of ORFs with significant similarity to genes encoding these enzymes in other organisms.

HP Locus	Organism	Gene Product Name	Similarity	Identity
HP0086	*Corynebacterium glutamicum*	Malate dehydrogenase	36.81%	25.93%
HP0104	*Escherichia coli*	5'-Nucleotidase	36.71%	25.76%
HP0133	*Escherichia coli*	Threonine transporter	50.00%	33.33%
HP0192	*Synechocystis sp.*	Aspartate oxidase	42.08%	30.94%
			42.34%	29.20%
			38.81%	26.48%
			39.59%	30.20%
			41.40%	30.65%
			35.54%	25.73%
			54.57%	44.51%
			40.41%	29.80%
			39.17%	27.74%
			43.86%	32.63%
			49.25%	37.10%
			44.20%	30.64%
			40.68%	31.94%
			33.95%	23.66%
			41.02%	28.15%
			32.58%	21.80%

in silico Prediction:

The *H. pylori* Network includes a malate dehydrogenase function

L-Malate + NAD^+ V Oxaloacetate + NADH + H^+

Computational Investigation:

BLAST search indicates the presence of a Malate:Quinone Oxidoreductase (MQO) in *C. glutamicum* with significant similarity (36.81%) and identity (25.93%) to locus HP0086 in *H. pylori.*

Biochemical Verification:

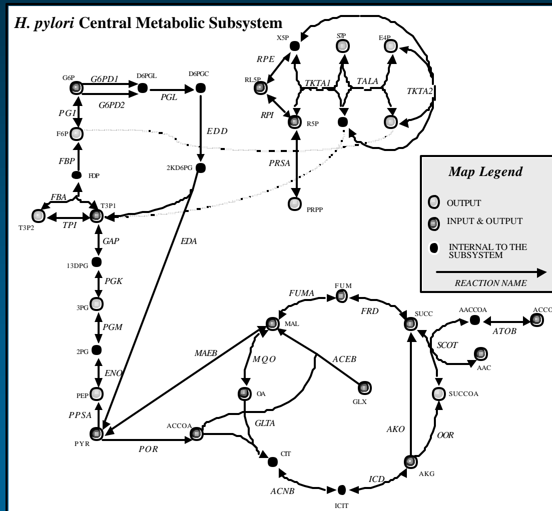
Kather et.al. (*J Bact*, June 2000) demonstrate MQO activity of locus HP0086 in *H. pylori*

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

One case where the locus was explored further and produced a significant hit was the case of malate dehydrogenase in *H. pylori.* Construction of the model indicated the presence of this enzyme although no evidence was found in the genome or biochemically. Our BLAST search indicated the presence of a Malate:Quinone Oxidoreductase (MQO) in *C. glutamicum* with significant homology to locus HP0086 in *H. pylori.* The activity of the HP0086 gene product was later verified to be that of malate:quinone oxidoreductase in an independent study. This shows that network reconstruction can be a predictive tool in its own right.

Experimental: a more palatable *H. pylori* and the peer-review process



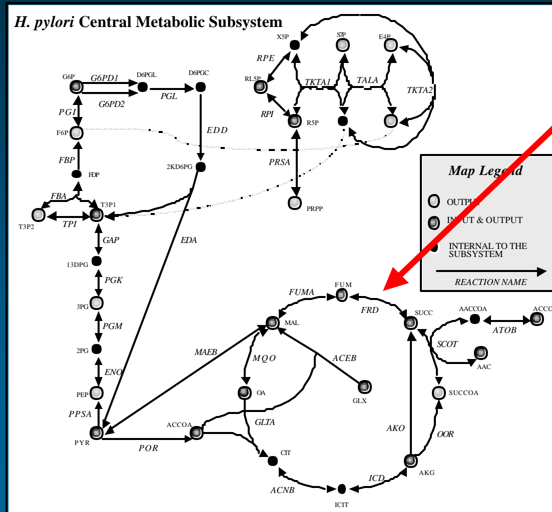
Our *H. pylori* manuscript received two detailed biologically-oriented peer reviews

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Our *H. pylori* model also underwent experiment-based model-building via the peer-review process. Our model was the beneficiary of two very detailed and biologically-oriented peer reviews – each one several pages worth!

Reviewer comments which helped



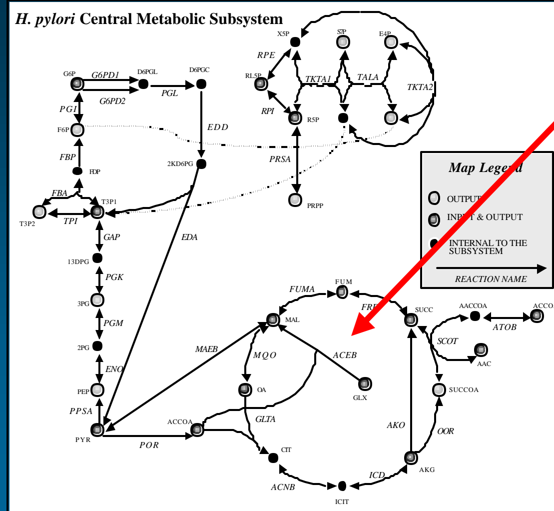
- *frd* – should be reversible
- Adenine biosynthesis pathway should be included (based on physiology)

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Some of the comments were very helpful evaluations. For example, one reviewer commented that the *frd* gene product (fumarate reductase, as was mentioned earlier) is thought to be reversible in *H. pylori*, noting that the organism could utilize succinate. Another encouraged us to include adenine biosynthesis enzymes in the model although they had not been characterized or found in the genome, due to the ability of *H. pylori* to grow in the absence of purines. The incorporation of these (and other) “reconstructive criticisms” led to more accurate model predictions.

Reviewer comments which opened new questions: malate synthase



- "aceB is probably incorrect annotation" (no *aceA*)
- But *aceB* is essential *in silico* for degrade folate biosynthesis products!
- Further experimentation necessary

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

In other cases, the reviewer comments opened new questions which should be examined further. One of these was the case of malate synthase. Malate synthase, encoded for by the *aceB* gene, is the second half of the glyoxylate shunt. Because the first half is thought not to be present in the system and because *aceB* has only been found in the genome, one reviewer said that the *aceB* annotation was probably incorrect. However, in our calculations we found that *aceB* was *essential under all conditions* for growth – but this was due to the production of glyoxylate as a by-product of folate biosynthesis. The *aceB* was required to keep glyoxylate from accumulating. This is a non-intuitive use of the malate synthase enzyme which was only identified by systems analysis and shows the potential of these models to drive discovery!

Model Applications – an Overview

Extreme Pathway Analysis

- Identify reaction subsets (i.e., reactions which are always used together)
  - Possible operon/regulon structure
- Identify unused reactions (i.e., “dead-ends”)
  - Places where further discovery or evaluation needs to take place
- Give insight into network structure

Flux-Balance Analysis

- Gene knockout/mutant studies and robustness
  - Possible antibiotic targets
- Prediction of growth experiments, evolutionary trajectories, etc.

Recent Developments in Analysis

- Singular Value Decomposition (SVD) analysis
  - Overall network properties
- Regulated Flux Balance Analysis (rFBA)
  - *in silico* Array studies, time courses of growth

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

These models have been used in many clever ways, which will be covered in detail later on. For example, Extreme Pathway Analysis (EPA) may be used to identify reaction subsets which are always found active together in simulations. These may point to genes which are regulated together, whether in operons or regulons. EPA may also be used to identify reactions in the network which are never active. These indicate either that such genes would be expected to be lost in evolution or where further discovery or evaluation needs to take place. EPA may also give insight into network structure, as will be explained later.

Flux-Balance Analysis (FBA) is used to analyze gene knockout/mutant studies, which is useful for metabolic engineering, or for pathogens, identifying possible antibiotic targets. FBA may also be used to predict growth experiments, evolutionary trajectories, and the like.

Finally, there have been several recent developments in analysis approaches, such as Singular Value Decomposition and the construction and incorporation of transcriptional regulatory networks, which bring further insight into microbial behavior and systems properties.

Summary

- Metabolic networks can be incorporated into genome-scale models which simulate cellular behavior
- *In silico* metabolic models exhibit the same network characteristics as *in vivo*
- Growth, metabolite uptake rates, and byproduct secretion rates may be quantitatively predicted for aerobic and anaerobic growth by imposing environmental and regulatory constraints
- *In silico* models can be used to analyze, interpret, and predict knockout viability of metabolic genes

References

- Brown, J.C. “What the heck is an *E. coli*?” http://people.ku.edu/%7Ejbrown/ecoli.html
- Varma, A. and Palsson, B.O., “ Stoichiometric flux balance models quantatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110,” *Appl. & Environ. Microbiol*, 60(10): 3724-3731 (1994).
- Edwards, J.S. and Palsson, B.O., “The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities,” *PNAS*, 97(10):5528-5533 (2000).
- Edwards J.S., Ibarra, R.U., and Palsson, B.O., “*In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data,” *Nature Biotech*., 19: 125-130 (2001).
- Edwards J.S., and Palsson, B.O.. “Robustness analysis of the *Escherichia coli* metabolic network,” *Biotechnol. Prog*. 16: 927-939 (2000).
- Covert M.C. and Palsson B.O., “Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*,” *JBC*, Papers in Press (2002).
- Nissen, T.L., et al., Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. Microbiology, 1997. 143(Pt 1): p. 203-18
- Overkamp, K.M., et. al., In vivo analysis of the mechanisms for oxidation of cytosolic NADH by Saccharomyces cerevisiae mitochondria. Journal of Bacteriology 2000 May, 182(10):2823-30.
- Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO. Metabolic modeling of microbial strains *in silico*.Trends Biochem Sci. 2001 Mar;26(3):179-86.